IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 7, 2020, accepted September 10, 2020, date of publication September 18, 2020, date of current version October 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3024926

MuLViS: Multi-Level Encryption Based Security System for Surveillance Videos

AMNA SHIFA[1], MAMOONA N. ASGHAR[1,2], MARTIN FLEURY[3], (Member, IEEE), NADIA KANWAL[2,4], (Senior Member, IEEE), MOHAMMAD S. ANSARI[2,5], (Senior Member, IEEE), BRIAN LEE[2], MARCO HERBST[6], AND YUANSONG QIAO[2], (Member, IEEE)

[1]Department of Computer Science and IT, The Islamia University of Bahawalpur, Punjab 63100, Pakistan
[2]Software Research Institute, Athlone Institute of Technology, Athlone, N37 HD68 Ireland
[3]School of EAST, University of Suffolk, Ipswich IP4 1QJ, U.K.
[4]Department of Computer Science, Lahore College for Women University, Lahore 54000, Pakistan
[5]Department of Electronics Engineering, Aligarh Muslim University, Aligarh 202002, India
[6]Evercam Pvt., Ltd., Dublin 15, D01 FW20 Ireland

Corresponding author: Mamoona N. Asghar (masghar@ait.ie)

This work was supported in part by the Marie Sklodowska-Curie (MSC) Career-FIT Postdoctoral Fellowship under Project MF 2018-0179, in part by the European Union's Horizon2020 Research and Innovation Programme through the MSC under Grant 713654, in part by the Science Foundation Ireland (SFI) under Grant SFI 16/RC/3918, and in part by the European Regional Development Fund.

ABSTRACT Video Surveillance (VS) systems are commonly deployed for real-time abnormal event detection and autonomous video analytics. Video captured by surveillance cameras in real-time often contains identifiable personal information, which must be privacy protected, sometimes along with the locations of the surveillance and other sensitive information. Within the Surveillance System, these videos are processed and stored on a variety of devices. The processing and storage heterogeneity of those devices, together with their network requirements, make real-time surveillance systems complex and challenging. This paper proposes a surveillance system, named as Multi-Level Video Security (MuLViS) for privacy-protected cameras. Firstly, a Smart Surveillance Security Ontology (SSSO) is integrated within the MuLViS, with the aim of autonomously selecting the privacy level matching the operating device's hardware specifications and network capabilities. Overall, along with its device-specific security, the system leads to relatively fast indexing and retrieval of surveillance video. Secondly, information within the videos are protected at the times of capturing, streaming, and storage by means of differing encryption levels. An extensive evaluation of the system, through visual inspection and statistical analysis of experimental video results, such as by the Encryption Space Ratio (ESR), has demonstrated the aptness of the security level assignments. The system is suitable for surveillance footage protection, which can be made General Data Protection Regulation (GDPR) compliant, ensuring that lawful data access respects individuals' privacy rights.

INDEX TERMS GDPR, ontology, partial encryption, privacy protection, video surveillance, surveillance cameras, encryption, visual surveillance data.

I. INTRODUCTION

Video surveillance (VS) systems using fixed cameras have many applications, which range from the monitoring of threatened locations by the security and defense forces to checking up on children at play, monitoring tourist attractions, and keeping an eye on critical infrastructures, to name a few. They are now being networked, individually or collectively, by means of the Internet of Things (IoT) [1]. In particular, the IoT has been applied to emerging Smart Cities [2]. However, IoT devices are susceptible to attack [3] because of their constrained resources. Yet they provide potential access points to the traditional Internet, where hitherto some measure of security has been carefully built up. Thus, [4] provided a layered framework for IoT security within a Smart City. The IoT devices were lightweight microcontrollers. However, the security management software structure appears to be conventional, based upon software managers and, as a result, may be insufficient. Besides, traditional VS systems represent a threat to the privacy of personnel working within

The associate editor coordinating the review of this manuscript and approving it for publication was Alessia Saggese.

environments under surveillance, such as a city airport, as well as a threat to the privacy of members of the public passing through that environment, who do not otherwise pose any threat. In that regard, the system builds upon prior work by the authors [5] on achieving compliance with the European Union's (EU's) recent legislation for privacy protection, namely General Data Protection Regulation (GDPR) [6].

Thus, the motivation of this paper is to protect the video content captured by a surveillance system. A surveillance system [7] consists of different devices with different processing and storage capacities. The surveillance devices may have some embedded intelligence but may also be constrained in terms of processing and storage capability [8]. The smart security cameras, including those based around the popular Raspberry Pi embedded processor, are capable of sending text and with the Multimedia Messaging Service (MMS) can send message notifications, images, and video clips [9].

In response to data and privacy protection of video contents, which is the focus of the current paper, surveillance video can be fully encrypted or selectively encrypted [10] during communication, display, and storage. Though, cryptographic protection of video is possible, the large amounts of video data created and, possibly, stored, along with the heterogeneity of surveillance devices, pose a problem as to what forms of encryption are technologically appropriate. Moreover, conventional security measures cannot necessarily be applied to all of the surveillance data and, consequently, this presents a challenge within the resource constraints and dynamics of surveillance environment. In addition, the heterogeneous nature of devices in operation in advanced surveillance systems requires scalable and adaptable management frameworks, alongside streamed video confidentiality, and secure storage of such 'Big Data'.

This paper newly reformulates the design choices that contribute to a system-level design. Several device-related characteristics/parameters could be considered for the security adaptation. Overall, storage remains a relatively high expense and, therefore, making the right choice continues to be crucial. For that reason, in this study, the following device-related characteristics/parameters are considered to formulate the different security levels to achieve data confidentiality of surveillance videos:

1) Storage Capacity/Memory: the memory/storage capability of the surveillance applications is considered to determine which security level may be adopted to achieve data confidentiality without generating significant encryption overhead and energy consumption.

2) Energy/Power: Surveillance devices within surveillance system may provide finite energy, as they are frequently battery powered, particularly in the case of smart devices. Encryption on these devices should be implemented in a way that it cannot consume too much energy. Thus, this parameter is selected to direct which security measure required to be taken in the surveillance video.

3) Resolutions: Display resolution is another important consideration that plays a vital role in surveillance videos. Higher resolution provides a better-quality bitstream, which escalates the possibility of identifying people and objects within surveillance videos. However, the higher the resolution puts a greater demand on network bandwidth, storage space and power consumption.

4) Bandwidth: The videos streams and images captured by smart cameras and camera-enabled sensors require different bandwidths, depending on network technology and capacity. Greater bandwidth can transmit higher resolution, smoother videos at higher quality, even for high-motion scenes.

5) Throughput: In a surveillance system, data packets can be transmitted over various communication technologies. Throughput is one of the most important concerns for efficient power management when dealing with the real-time interconnection between heterogeneous and ubiquitous devices. Recently, the authors of [11] implemented fuzzy logic to determine the sleeping time of the devices according to the battery level and to the ratio of throughput to workload in the smart home. Thus, in this study, throughput is considered an important input parameter.

6) Frame Rate: - Each video stream has a different frame rate, depending on storage and bandwidth. Higher frame rate provides smoother video in high-motion scenes. However, the higher frame rate increases the demands of storage and bandwidth requirements, which are unreasonable for constrained resources environment. Therefore, that should be taken into consideration.

The proposed system is comprised of three main components; (1) Features of Interest (FOI) (i.e. motion, face, human and background) detection, (2) security level selection according to device specifications and (3) encryption on the videos stream according to the security level output from the second component. Notice that GDPR allows data controllers/processors to retain an individual's personal data if they are in the form of pseudonymised information and/or encryption (see Article 6(4) (e) and Article 32(1) (a)). That is GDPR encourages data protection-by-design (see Article 25) [5], according to the sensitivity level of the data. For video, video redaction through encryption is the normal data safeguard provided by GDPR and indeed that safeguard is adopted in the proposed security system if the data warrants that protection. Accordingly, the current research innovates with MuLViS, which is a data protection-by-design solution in the GDPR sense. This solution, by using an SSSO, can protect the sensitive video content by extracting contextual information from real-time, surveillance videos. A preliminary, conference version describing the SSSO appeared as [12]. The solution also recognizes critical storage-device capabilities, such as storage capacity, energy consumption, bandwidth utilization, and privacy protection is then achieved by means of suitable FOI encryption. In short, a solution is provided through MuLViS that is GDPR compliant.

The paper makes both system-level and technical innovations. In summary of the effective original system-level contributions of this paper, the security framework in the paper works as follows:

(i) FOI extraction performed on real-time surveillance video by using different State-of-the-Art (SoA) computer vision techniques. FOI's are extracted on the bases of privacy based use-cases given in Section II.A.

(ii) **Mu**lti-**L**evel **Vi**deo **S**ecurity (MuLViS) designed and implemented, utilizing the ontology. Five different security levels are defined in the MuLViS. One of these levels can be adopted according to security needs and device capabilities.

(iii) Provision of automatic security level selection concerning device resources by means of ontological reasoner. The rules of the reasoner are defined using Semantic Query-Enhanced Web Rule Language (SQWRL).

(iv) Lightweight, partial-encryption implemented on a FOI, according to the selected security level recommended by the ontological reasoner. Notice that though a lightweight cipher, i.e. computationally less intense cipher, is not used, the effect of partial encryption is to reduce the total amount of data encrypted.

In addition, the paper's main effective technical contributions and novelty are as follows:

(i) It presents a GDPR-compliant data protection-by-design solution for surveillance videos by combining some state-of-the-art computer vision algorithms with suitable cipher. It does this, along with using ontology to select devices in a video surveillance.

(ii) Multi-level security is achieved by partially encrypting the specified FOI as per GDPR requirements, i.e. (1) motion or texture within a video footage to conceal activities, (2) human facial features or the full bodies of people to protect the identities of individuals, and (3) background features to conceal locations, all according to each different security level. (Detailed use-cases of feature selection are given in Section II.A in support of the GDPR requirements).

(iii) The proposed solution encrypts the video in such a way that the video will be partially viewable but will not allow any individual to directly access the original video contents. Even a 'hacker' will only be able to access the encrypted form of video footage, thus meeting GDPR requirements.

To demonstrate the quality of these contributions, the paper contains:

(a) Visual and statistical experimental results are discussed in terms of their performance, as a way of evaluating MuLViS with its integrated SSSO, and

(b) A comparative analysis demonstrates the significance of the security framework for smart surveillance devices.

The rest of the paper is arranged as follows: In Section II, the context to this research is outlined. For those unfamiliar with that background, this Section is recommended. Then, MuLViS and its modules are described in Section III. Section IV presents extensive experimental results, along with the performance of the system and its contributions. Finally, Section V makes some concluding remarks and as well as considering possible future research and development.

II. CONTEXT

It is vital to consider the various threats to video surveillance security. Here is an illustrative list of the concerns that arise:

- Attackers may simply intercept surveillance video at intermediate networked devices for the purpose of identifying the individuals under surveillance and the purpose of surveillance.
- Attackers may use surveillance video interception as a tool to better threaten monitored individuals, owing to awareness of the protection measures in place.
- Attackers may modify the transmitted data and recipients may receive critically wrong information.
- Attackers may also record video previously and put this recorded video back into a network so that surveillance operators may regard this video as being sent in real-time.

More generally, surveillance video contains sensitive and personal information and these videos often need to be streamed to screens. The video may subsequently be stored on dedicated storage repositories, such as Digital Video Recorders (DVRs), Network Video Recorders (NVRs), or a cloud. Alternatively, the video may be stored on the end devices themselves (such as on smart cameras) for a period of time, so that if a significant event occurs the video can be further processed for a detailed analysis. Because these videos contain information about the subjects (places and people) and activities around those subjects, during transmission these videos are vulnerable to interception by malicious individuals or groups. Likewise, videos stored on discs, cloud, and end devices are vulnerable to inspection by hackers, who can exploit a system's security weaknesses. All these events may result in data disclosure to unauthorized parties.

Indeed, in a surveillance system, the confidentiality of the data is very important. For example, the location of places under surveillance should remain confidential. Every multimedia message or video stream, captured from surveillance cameras, is sent to storage servers. In doing so, the video passes through several nodes/access points to reach these repositories using heterogeneous communication technologies. However, these access points and intermediate networks can be extremely vulnerable to attacks. In fact, streamed data can be exploited by the terrorist and by malicious users. Therefore, it is necessary to adopt secure methods to protect both live and stored surveillance data.

There are many examples of privacy and information leakages to be found on news channels and on the world-wide-web in general. Hence, the security of the surveillance data and particularly the privacy protection of individuals shown in the videos is a challenging requirement, given that privacy has risen higher on the public agenda [13]. In addition, laws now exist that require that the privacy of individuals should be preserved during surveillance. To this end, the EU has recently adopted GDPR [6], which regulates the privacy protection required for processing and storing personal data within the EU. GDPR applies globally to data protection,

if that data is used with the EU and even if it originates from outside the EU but is used within the EU. It also harmonizes relevant laws within the EU and ensures the rights and protections to both European citizens and visitors to the EU, by providing data safeguards through a reversible process of encryption.

Besides, real-time applications differ from existing conventional surveillance approaches, due to the need for low-latency communication and processing, despite resource limitations and the huge volumes of surveillance multimedia data [14]. In fact, currently, pervasive video-surveillance systems are a significant source of video traffic over the public Internet. According to CISCO statistics [15], video traffic is expected to grow to 82 percent of all internet traffic by 2022 and, of that, 3 percent of all internet video traffic will be from surveillance cameras (such as the well-known, lower-cost Dropcams). In addition, as indicated in Section I, video surveillance operating within Smart Cities [16] are likely to be a major source of surveillance traffic, especially as that surveillance video data, within the architecture of the IoT [17], after a perception layer (or similar) is likely, in a network layer (or similar) to be passed on to cloud data centers [18] across the conventional Internet, before video analytics, in an application layer, takes place. Because the security needs of the IoT and the conventional Internet are not the same, because of the types of devices and the organization of the networks within each, there is a need to integrate the security provision within each [19].

Thus, video streams generated by surveillance systems have become one of the significant contributors to a massive amount of multimedia data moving across computer-based systems. Consequently, a suitable approach to tackle these issues is required and that approach could be through suitable abstraction technologies, such as semantic content representation or context-aware perceptual modeling or through ontologies. In recent years, context-aware, ontological-based, perceptual modeling approaches have been adopted in the video surveillance application [20]–[24] (see Section II.B) and information security [25]–[27] domains. Ontologies enable the content description of basic categories within the domain and relations among them to make them machine understandable [28]. Indeed, the description of the associated concepts of domain by context-aware, perceptual modelling for intelligent systems increases the accuracy of the indexing process.

A. FEATURE OF INTEREST (FOI) DETECTION

Feature of Interest (FOI) or the well-known Region of Interest (ROI) detection is the fundamental process of any Intelligent Video Surveillance (IVS). The FOI is a sensitive area within the surveillance video, which needs to be protected to attain the desired level of security and confidentiality. A FOI could be any real-world instances within the video such as humans, faces, animals, motion, all kinds of vehicles, license plates, background and so on. FOI is often first identified, before employing different methods to protect the FOI, coupled with real-time, context-aware processing; including efficient event recognition, detection, and notification. Researchers have proposed a plethora of FOI detection techniques using computer vision and machine learning algorithms. However, the performance and accuracy of detection algorithms differ according to environmental conditions, the surveillance devices utilized, and their positions. The position of the device affects the Field of View (FOV), detection accuracy and quality of the captured video footage. Some key challenges experienced in FOI detection are shadows, illumination changes, dynamic backgrounds, foreground aperture, noise, camera jitter, image blurring, slow motion of moving objects, low-quality footage and low resolution [29]. For static installations of surveillance cameras, generally, each frame of a video is categorised into two parts: (1) The stationary or static part, called the background, and (2) The moving part, called the foreground. For GDPR compliance, it is not necessary to encrypt complete video frames if there is no sensitive data within the non-encrypted parts. So, in MuLViS, face, human/people (full body), motion, and background are considered as FOIs at each level. For further clarity, use-case scenarios are described below:

Use case 1 (Activity protection): In video surveillance motion is considered to be the most important part of the video. Motion holds information about spatio-temporal relationships among objects in the FOV of the camera. In many cases, fixed video surveillance cameras or visual sensors are installed at indoor and outdoor locations, such as for monitoring children and elderly people or groups of people within a community or in street/parking areas, all with the purpose of monitoring activities for security purposes. In such scenarios, there is often little activity captured by the cameras. It is also impossible to implement advanced algorithms on such limited-resource devices. Moving objects are detected in only a fraction of the captured videos, though they are suitable FOIs to implement security measures upon. In fact, motion can be detected without the implementation of advanced detection algorithms on such resource-constrained devices. Thus, in this study motion is considered as an FOI for encryption of surveillance video data when dealing with simple, low-resolution videos captured by constrained devices.

During the last decade, several motion-detection approaches have been proposed for surveillance cameras [30], [31]. Generally, motion detection can be performed utilizing three methods: (1) Temporal filtering [32], (2) Background subtraction [33], and (3) Optical-flow [34]. In the first method, motion is segmented by calculating the temporal difference between two or more than two consecutive frames. This is the simplest method to implement. The second method, i.e. Background subtraction, in which firstly the background model of the static region is constructed by comparing pixel-by-pixel absolute differences between consecutive frames and then motion is segmented by compared it with live video frames or reference frames. Any pixel of the reference frame that deviates significantly from its previous value is categorized as motion. Alternatively, in the

optical-flow method, intensity changes of frames over time relative to camera motion are compared to find estimates of the motion in a video [34]. The aim of the optical flow method is to distinguish the camera motion patterns from the object motion patterns in a scene. The pixels that have high-intensity differences can be classified as moving objects. Thus, this is a method suitable for dynamic backgrounds. However, performing optical flow is computationally complex, requires special hardware and is sensitive to noise [35]. Therefore, in this work, for motion detection by fixed cameras, temporal difference between the reference frame and the current frame is calculated.

Use case 2 (Individual Secrecy): The face as an individual part of the human body, Human (whole body) and skin the important feature within the surveillance footage. To provide for the privacy of individuals at public and private places such as at parks, airports, bus stations, and shopping malls, the face and human are considered FOIs for encryption. For example, suppose that a person is situated just in front of a camera then their face is a sensitive area that needs to be protected to ensure the privacy of that person. However, if the height of the camera is too high or the pose variations of the person such as side-on poses or poses that are not in the FOV of the camera, then face detection will be unreliable. In such a scenario, the skin can be considered a sensitive area that should be protected. However, in the worst-case, the skin may not be detected, due to a variety of causes such as blurriness, low resolution, illumination variations or diverse skin tones of the person within the video scene. Thus, herein, the face and human (full human body) are considered as FOIs for security and privacy protection.

Face detection is the process of determining the location of faces that are present in a video or an image. A variety of face detection techniques have been presented until now [36]. However, face detection remains a challenging task due to external factors (illumination condition, orientation, scaling, FOV, low resolution and so on) and internal factors (facial expression, pose change, glasses, hair, beard, moustache, and shade). Face variation creates major difficulties in the development of an efficient and accurate face detector. The first face detection algorithm for real-world applications was presented by Viola and Jones [37]. This detector was and is based on Haar-like features and a cascade AdaBoost classifier and is until now widely adopted algorithm [38]. The authors of [39] utilized Local Binary Patterns (LBP) for face detection. Some researchers also adopted regional statistics-based face detection approaches (in the form of histograms), as proposed in [40]. However, existing state-of-the-art face-detection methods are not optimized for a complex real-time environment; thus, they suffer from various problems when deployed in a surveillance system, such as when there are dynamic backgrounds, illumination changes, camera jitter processing and unconstrained conditions (arbitrary variations in pose and occlusion) is required. Therefore, in work herein, a Normalized Pixel Difference (NPD) face detector [41] method has been utilised for face detection because of its efficiency and accuracy in unconstrained scenarios, such as illumination variations, out-of-focus imaging, blurring and low resolutions.

Researchers have presented many different techniques for human detection such as in [29], [40], [42]–[44]. For instance, in [42] Dalal *et al.*, proposed Histogram of Flows (HOF), in which, for human motion encoding, temporal descriptors (features) are defined from optical flows. In [43] the authors proposed human detection in non-controlled environments using Histogram of Oriented Gradient and Gabor filters (HoGG). Zhou and De la Torre [44] proposed a human detection model for videos using spatio-temporal matching, in which the motion of joints was signified by trajectories. In this study, for human (people) detection, a motion-based feature is applied using the Histogram of Flows (HOF) proposed by [42] is employed, due to its robustness in an unconstrained environment.

Other methods may have similar advantages in terms of robustness in an unconstrained environment, including those reviewed in this Section. On the other hand, other recent machine-learning-based methods have a high training cost [45]. Thus, in this work [42] is utilized due to its adaptability to local deformation of the human and to background variations in the video, as well as low computational complexity compared to these other algorithms.

Use case 3 (Location Hiding): In a surveillance environment, individuals may have a strongly-felt objection to losing their location secrecy at certain locations such as military bases, banks, police stations, or automobile tolls. In these scenarios, the location is a sensitive entity that should be protected. Over the years, researchers have proposed various background modeling methods to distinguish the foreground and background in a video. The most extensively used pixel-based parametric background modeling methods are the Gaussian Mixture Model (GMM), and Adaptive GMM (AGMM) [46]. In Mixture of Gaussians (MOG) modeling, each pixel is modeled by more than one (k) Gaussians per pixel (multiple Gaussian distributions) to observe the variations in the color of a pixel in Red-Green-Blue (RGB) color space at any time t. A pixel frequently observed in the recent past that does not fit the k distributions is labeled as foreground. However, in the current study, the MOG2 algorithm is employed, which is based on the work in [47], [48]. In MOG2, k (number of Gaussian distributions) are selected dynamically for every pixel rather than keeping k constant throughout. The model is selected for background detection because it produces robust and efficient results for lower illumination variations compared to other methods [49].

B. ONTOLOGIES IN VIDEO SURVEILLANCE

To minimize the gap between the results interpreted by an intelligent system and those perceived by a human from the multimedia information, ontology-based approaches are adopted [20]–[24]. In [20], Hernandez-Leal *et al.* used ontologies to reduce the semantic gap between low and high-level information in an IVS system. However, the ontology

was not integrated with algorithms for intelligent detection, tracking, and recognition of events. In [21], Tani *et al.* presented a Semantic Web Rule Language (SWRL)-based ontology to bridge the semantic-gap problem and detect a single or multiple objects events in surveillance video. In [22] Calavia *et al.* also proposed an intelligent video surveillance ontology system to analyze object movements and recognize abnormal circumstances. However, the proposed system-domain application was not consistent with the ontology's representation. Pahal *et al.* also presented [23] an ontology-based system for situation tracking in a smart surveillance environment using Dynamic Bayesian Networks (DBNs). In general, the literature shows that a common perspective adopted by researchers in order to design context-aware intelligent systems is focused on the indexing process to support object detection, event detections, traffic monitoring, and abnormal behavior detection and analysis. Notice also that such systems are also employed to provide humans with a way to analyze the reasoning process made by the system that can serve different kinds of analyses.

However, unfortunately, automatic, context-aware representation along with security concepts in the constrained but dynamic Mobile IoT (MIoT) has received nominal attention up to now. In [24], Martínez *et al.* considered the anonymization of categorical datasets through semantic information by employing methods from Statistical Disclosure Control (SDC), such as recoding, micro-aggregation, and resampling. These methods were then adapted to take into account the semantics of the data they were protecting relying on ontologies to model the semantic knowledge associated with the attributes of the dataset. In [50], a multi-layer cloud architectural model was developed to provide better service using an ontology, by enabling secure seamless interactions among heterogeneous devices in smart homes. This research work also demonstrated that ontology-based methods provide better solutions for the heterogeneity problem and that high-security and privacy can be ensured within smart homes. In [51], researchers developed a context platform, KaliSmart, by incorporating an ontology to collect contextual information from sensors for adapting system behavior and semantic event detection and to provide services to clients in a ubiquitous environment. Recently in [45], a knowledge-based modeling of a UAV recorded video scenario was introduced. In the proposed scheme, Semantic Web technologies are utilized to design ontology-based multi-layer knowledge schema for the target detection and description rather than using only classification methods. Through the proposed schema, a high level of abstraction of a scene has been achieved.

Nevertheless, research on integrating the semantic reasoning and security approaches appear to still be in its infancy and existing studies on this topic are probably insufficient. Thus, in the current paper, an ontology-based security system for the surveillance videos has been implemented that can recognize, analyze and store surveillance video content, along with providing video stream confidentiality with respect to device storage and processing capability. In the security system, a context-aware ontology, that is SSSO, has been developed to represent the domain knowledge of videos, heterogeneous devices, and their device-specific security concepts, along with the relationships among them. The SSSO makes the classification of video content easier for the MuLViS system and improves the effectiveness of device-specific security adaptation.

C. SECURITY IN VIDEO SURVEILLANCE SYSTEMS

Related work on protection approaches for video surveillance can be divided into two categories: Non-scrambling based protection methods and Scrambling based protection methods. In the Scrambling based protection methods, the video footage is encrypted using various cryptographic methods, whereas non-scrambling methods the sensitive region is protected without utilising the cryptographic methods. The scrambling based protection methods are considered more efficient and secure; thus, in this paper, the latter method is adopted. More recently, the authors Ciftci *et al.* [52] proposed reversible privacy protection for static images in which the original colour information of entire frames is replaced with some other colour palette information called false colours. It is a reversible technique, and the replaced false colours can be reversed back to the original.

In another work, Hoo *et al.* [53] presented a privacy filter framework in which human skin regions are detected by incorporating various current skin-detection methods and detected skin regions are removed from the video to achieve privacy protection. Wang *et al.* [54] proposed a privacy protection scheme for ubiquitous surveillance systems, in which intra prediction modes (IPM) are encrypted along with the SNC within the privacy region. In the proposed scheme, to avoid the drift error and to reduce the BR overhead, the re-encoding method is integrated along with the spiral binary mask mechanism. Moreover, the encryption is performed using Rabbit stream cipher. Xiaojing *et al.* [55] applied complete encryption on the face region to obliterate it, so that, no one could reveal the identity of the person present within the surveillance video. Though that scheme did not obscure all the information in the video frame, nevertheless the behaviour of a person is no longer perceptible. In contrast, scheme proposed in this paper preserves the structure of a protected sensitive region. Hence, the solution proposed in this work can be used for behaviour analysis, without revealing the identities of people.

Moreover, most prior proposals for privacy in real-time IVS systems proposed by researchers focus on more narrow techniques rather than abstraction technologies. For example, Carillo *et al.* [56] proposed a selective encryption scheme (one in which only visually significant semantic elements are encrypted) in which pixels in the ROIs are permutated before compression. The scheme ensured format compliance at the decoder (allowing intermediate processing of encrypted regions without decryption) but it is known that compression

efficiency significantly declines when encryption is applied before compression due to the loss of exploitable redundancy.

Then, in another example, Ahmad *et al.* [57] proposed a real-time, occupancy-monitoring system with ROIs in which images of people present in the video are encrypted with an advanced encryption method (Tangent Delay Ellipse Reflecting Cavity Map). However, the usability and adaptation of this system and, for that matter, the method of [56], did not take into account current surveillance systems, which exist in an environment where heterogeneous devices are the norm. The design of a self-optimizing, context-driven, and energy-aware IoT wireless video sensor node for surveillance applications is presented by [58]. Another recent work [59], proposed distributed three-layered architectural framework named IoT-guard for the real time crime detection and security management within the smart home surveillance system while conserving the resources usage such as energy, bandwidth, storage and CPU usage. The proposed framework is an event-driven video surveillance system in which edge-and fog-integrated approach employed. For the crime detection and confirmation, Artificial Intelligence (AI) and an event-driven approach are utilized.

Moreover, many organizations have started to incorporate multilevel policies into their systems. Examples of these are the access-control policies implemented in Microsoft Vista and Red Hat Linux. Multi-Level Security (MLS) concepts were started within military systems [60], in which information was classified into confidential, secret, and top secret. Researchers have utilized these concepts in their hardware and software systems in a number of ways. For example, in [61], researchers incorporated MLS into real-time systems so as to integrate system requirements and the implementation environment, thus enhancing the performance of those systems and their associated security measures.

Table 1 is a comparative summary of a variety of existing systems appearing in the text, as well as the proposed system of this paper. It is difficult to make a direct comparison as, for example, each system has a different FOI. However, some things stand out such as only one system has a moving node and only two systems achieve data confidentiality.

III. PROPOSED MULTI-LEVEL SECURITY SYSTEM

The context-aware MuLViS system is shown in Figure 1. The system has eight main modules namely: (1) Data Acquisition; (2) Device virtualization; (3) Real-time preprocessing. (It is in this module that the environmental context in respect, for example, to camera position and ambient lighting, is taken into account.); (4) Ontology modeling and reasoning; (5) Feature adaptation and security encoding; (6) Video chunking and tagging; (7) Lightweight encryption; and (8) Information retrieval modules. The details of the framework and its modules are explained below. The framework is suitable for decision-making for the security of surveillance data, based on storage device attributes. Sensors are used to extract the information about the end-device type, including the capabilities of sensors.

A. DATA ACQUISITION MODULE

In this module, the raw surveillance data (i.e. video frames) are captured from heterogeneous data sources, such as fixed CCTV cameras and smart cameras (Raspberry Pi cameras in our case). Initially, the footage output from the camera is stored in the camera's own registers as frames and some minor pre-processing such as flipping (horizontal and vertical), line skipping, and pixel binning happens within those registers. After that, the frames are transferred to the real-time preprocessing module for further processing and feature extraction, such as extracting features which might be faces, human bodies, or the background).

B. DEVICE VIRTUALIZATION MODULE

It is important to identify the attributes/characteristics of the devices within the surveillance system. Therefore, the identification of devices (surveillance and storage devices) and information about their attributes (storage, processing, and power) are explicitly acquired from sensors, which are installed on the surveillance devices. These device attributes, such as device identification number, device category, storage capacity, and processing power, are provided as an input to the ontology module for context-aware categorization and automatic mapping of device-specific attributes and security-level recommendations.

C. REAL TIME PRE-PROCESSING MODULE

The pre-processing module performs a low-level analysis of the raw data (video frames) acquired from the data acquisition module for Feature of Interest (FOI) selection. The processing unit, which may be within the same device or maybe outside the device connected by a network, first processes the raw surveillance data for FOI extraction. In the work herein, preliminary face detection, motion detection, human detection, and background selection are considered as FOIs. Feature extraction is performed by applying efficient, computer vision and background subtraction techniques. For motion detection by fixed cameras, firstly, the video current frames are divided into non-overlapping MBs of 16×16 pixels at a particular time interval *T-1* for Motion Estimation (ME). After that a comparison is performed of the block position in the current frame F_i at time T with either the previous $F_i - k$ frame or the next frame $F_i + k$, where k is the number of the frame used to compare with the present video frame. When the block matches with the block of the reference frame, motion vectors are generated. Moreover, in this work, the last frame is subtracted from the present frame and the residual frame Motion Vector Difference (MVDs) are calculated. The number of MVDs is dependent on the motion within a video. Background detection is performed by the optimal MOG2 [48], [49] method. The methods implemented in the processing module are chosen because of their robustness and efficiency compared to other background subtraction algorithms.

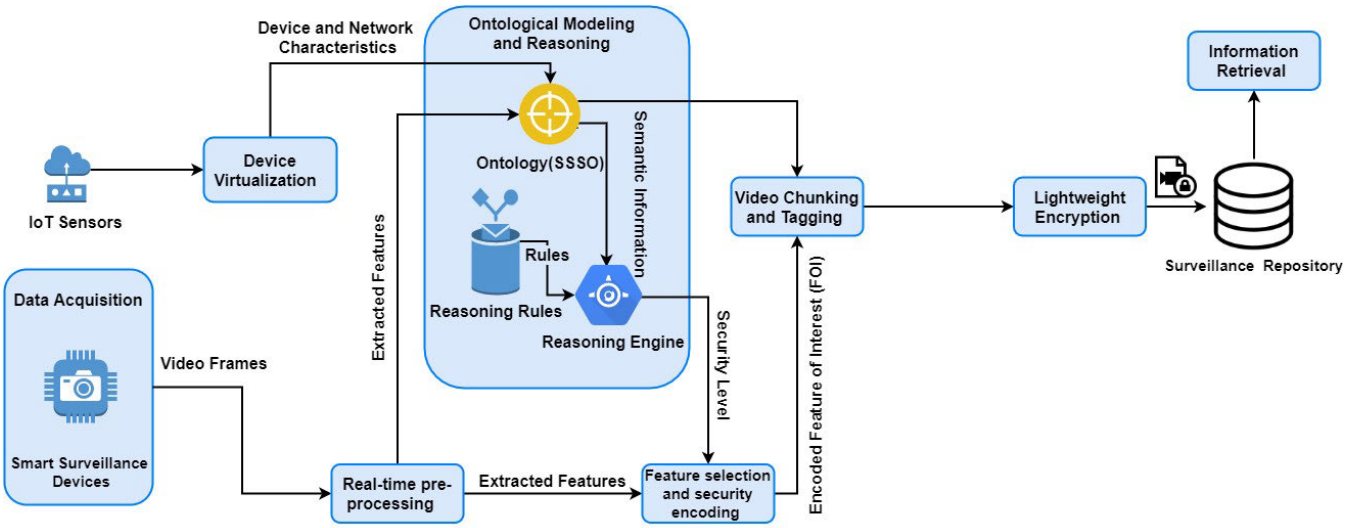


FIGURE 1. Architecture of the Multi-Level Video Security (MuLViS) system.

TABLE 1. Comparison of other existing systems with the proposed model.

System	[22]	[45]	[52]	[58]	[59]	Proposed MuLViS
FOI	Abnormal and alarming situations detection	Object, activity and situation detection	Entire frame	Detection of human presence	Crime detection	Motion, face, human and background detection
Video Capturing Nodes	Fixed	Moving	Fixed	Fixed/Moving	Fixed	Fixed
Context Aware	Yes	Yes	No	Yes	Yes	Yes
Semantic Web Technologies/Ontology	Yes	Yes	No	No	No	Yes
Environment	Constrained	Unconstrained	Static	Dynamic	*	Unconstrained
Data Confidentiality Achieved	No	No	Yes	No	No	Yes
Resources Aware	No	No	No	Yes	Yes	Yes

*Not mentioned in the paper

Besides, in this paper, for human (people) detection, a motion-based feature is applied using the Histogram of Flows (HOF) proposed in [42]. The algorithm is employed due to its adaptability to local deformation of the human and background variations in the video and its low computational complexity compared to other algorithms. In this work, the oriented gradients of both the boundary motion descriptors that describe motion vectors at body edges (boundary vectors) and also internal motion descriptors that describe motion vectors within the internal regions (including the relative movements vectors of different parts of the human body, e.g. left vs. right hand or leg,) are horizontally and vertically extracted from Spatio-temporal derivatives relative to the subsequent frame. A separate histogram is built for each and then combined with the Histogram of Oriented Gradient (HOG) descriptors. The temporal difference is estimated independently at each over a small N × N neighborhood. After that, a linear Support Vector Machine (SVM) classifier, working as a baseline classifier, is used to classify the extracted descriptor into human and non-human descriptors. For a more detailed description, interested readers are encouraged to read [42]. The advantage of the SVM classifier is that it is fast to run compared to the other linear classifiers. Likewise, existing state-of-the-art face-detection methods are not optimized for a real-time complex environment; thus, they suffer from various problems when deployed in a surveillance system, such as when there are dynamic backgrounds, illumination changes, and camera jitter processing is required. Therefore, in the work herein, a Normalized Pixel Difference (NPD) face detector [41] method has been utilized for face detection because of its efficiency and accuracy in unconstrained scenarios, such as illumination variations, out-of-focus imaging, blurring, and low resolutions. In the NPD features method, the relative difference between two pixels is calculated as:

$$f(a, b) = (a - b)/(a + b) \tag{1}$$

where the value of function $f\ (a, b)$ represents the relative difference of intensity values of the two pixels 'a' and 'b' while the sign of the function $f(a, b)$ represents an ordinal relationship between the two pixels. A zero value represents that there is no difference between the two pixels.

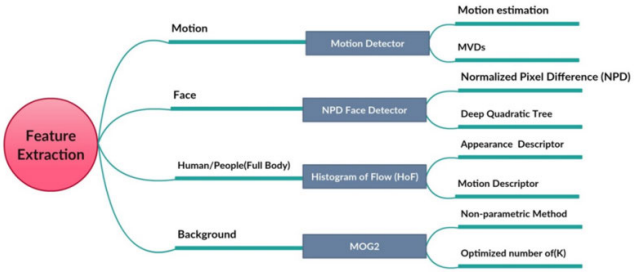


FIGURE 2. Overview of the state-of-art algorithms applied for feature extraction.

A summary of the algorithms applied to feature extraction is illustrated in Figure. 2. Notably, the following methods particularly MOG2 is employed because they are characterized by their high precision and low false positive rates in dynamic background and difficult challenging scenarios such as various illumination conditions, bad weather, and low frame rate [62]. The extracted low-level data is subsequently passed to the ontology module to interpret meaningful semantic knowledge for data fusion and reasoning. Additionally, this module automatically generates a real-time autonomous response to communicate without delay between the end devices.

D. ONTOLOGICAL MODELING AND REASONING MODULE

As previously mentioned, within the Ontological Modeling and Reasoning Module, a Smart Surveillance Security Ontology (SSSO) has the purpose of autonomously selecting the privacy level that matches a device's hardware specifications and network capabilities. To achieve this processing passes through two phases: 1) Ontology modeling and 2) Ontological reasoning. In the ontological modeling phase, the ontology is structured to formalize the basic concepts (specification of objects), attributes of concepts, and the relationships between these concepts. In the ontological reasoning phase, the aforementioned ontologies, their description and the relation among the sub-domain are taken as the reasoning objects in the reasoning engine to achieve automatic security-level selection for heterogeneous surveillance devices. In terms of implementation of these two phases, in particular, video domain concepts and device-specific security concepts are integrated. We now treat separately each, of these two phases, ontological modelling and reasoning, in more detail.

Ontological Modelling: Thus in the SSSO modelling phase, firstly the security ontology is constructed by classifying the devices in operation through the instance similarities and semantic similarities and making them an SSSO instance. After that, the related concepts are associated/mapped to the security level in the security ontology. Subsequently, the device-specific security concepts are mapped to the multimedia domain ontology to finally establish the SSSO. Therefore, through the SSSO, a set of correlated video and device concepts abstracted from the surveillance video scenario are structured.

The concepts are classified into a top-layer and low-layer hierarchical structure. The top-layer structure captures five high-level, generalized concepts, which are defined as *Places, Objects, Motion, Security,* and *Storage_Media*, as shown in Figure 3. Further, these top-level generalized concepts are divided into their associated sub-concepts in the low-level hierarchy.

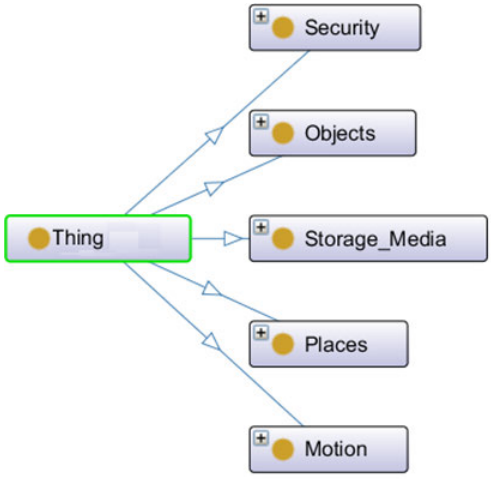


FIGURE 3. The top level structure of the Secure Smart Surveillance Ontology (SSSO).

In the current study, for the ontological development, the Web Ontology Language (OWL) is used in Protégé [63]. Protégé is employed because it is an extensible, and platform-independent. It also supports a variety of formats to construct and edit ontologies. For the interested reader, a code 'snippet' for the ontology of the SSSO can be found in Figure 4.

The details of the general concepts and their features in each sub-domain are defined in the low-level structure of SSSO. *Storage_Media* is the abstraction for device entities of a VS system. This sub-domain can then direct which security measure is required to be taken in the surveillance video. The *Storage_Media* and its low-level concepts are illustrated in Figure 5. Moreover, the basic device information is defined by Device_ID, Device_Name, Device_Type, Storage_Capacity, Processing_Memory, and Power_Battery.

In the *Security* concept, the associated low-level sub-concepts, defined as Level_1, Level_2, Level_3, Level_4, and Level_5, are shown in Figure 6. The attributes of security and their related entities are defined as Level_ID, Level_Name. Moreover, the *Security* concept and its sub-concepts are linked with the Motion, Objects and Places sub-domains defined in the SSSO for FOI selection. Level_1 has a minimum security policy while Level_5 has the highest security policy. Thus, the Security sub-domain corresponds to *Storage_Media* as well as *Motion, Objects*, and *Places* sub-domains for managing and automatic device-specific security-level selection.

Ontological Reasoning: The proper selection of parameters is an important factor for the accurate selection of

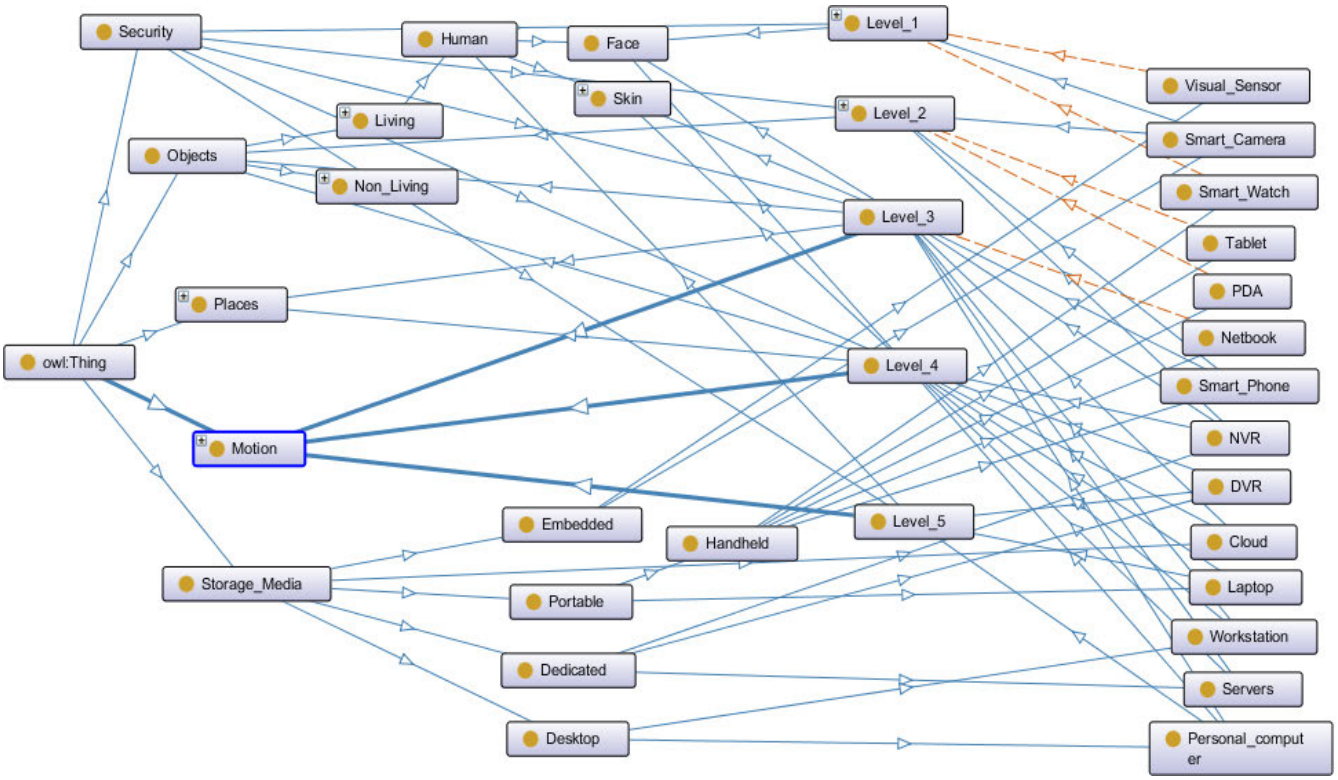


FIGURE 4. Code snippet for the SSSO.

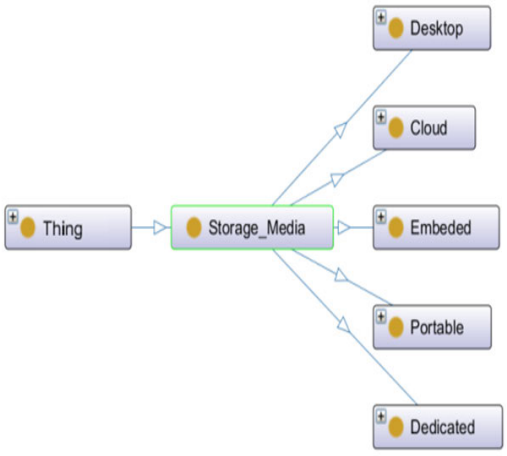


FIGURE 5. *Storage_Media* concept and its low-level concepts.

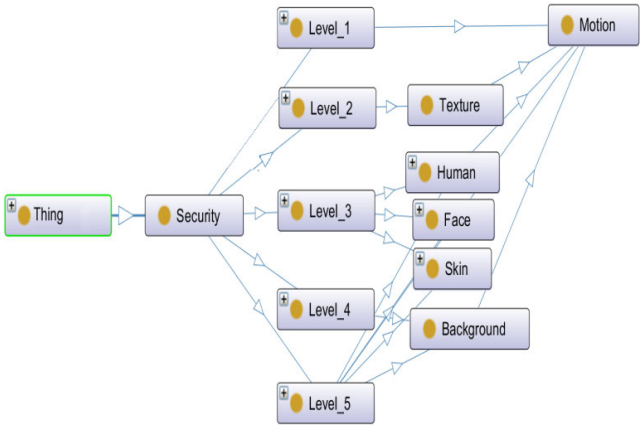


FIGURE 6. *Security* concept and its low-level concepts and their relationship with FOI.

the appropriate security-level in a real-time and dynamic environment. In the work herein, to define the reasoning rules for the automatic selection of the security level by the reasoning engine and extract the relevant knowledge from the ontology, the Semantic Query-Enhanced Web Rule Language (SQWRL) [64] is used. The automatic reasoning engine will take the following steps for security-level selection, by considering the necessary conditions. The Reasoning Rules defined for automatic security level selection are illustrated in Figure 7.

1. The *Storage_Media* is a superclass that has different devices in its sub-classes. Therefore, the Device_ID, Device_Type (t) and Storage_Capacity (s) are captured and then the Storage_ Capacity is compared with the predefined storage threshold and classified as very limited, limited, medium, high, or unlimited.

2. Since devices in *Storage_Media* have different battery power, which is another important attribute, Power_Battery (p) is measured and then the Power_Battery (p) is compared with the predefined Power threshold and classified as very limited, or low.

```
Security(? x)
∧ (Storage_Media (?Device_ID) ∧ hasDevice_Type (?Device_ID, ?t) ∧
has Storage_Capacity
(?Device_ID,?s) swrlb:lessThanOrEqual (?s,Storage_threshold))
-> sqwrl:select(?s,?Device_ID)
∧ (Storage_Media (?Device_ID) ∧ hasDevice_Type (?Device_ID, ?t) ∧
has Power_Bettery (?Device_ID, ? p) ∧
swrlb:lessThanOrEqual(?p,Power_threshold))
-> sqwrl:select(?p, ?Device_ID)
∧ ( Storage_Media (?Device_ID) ∧ hasDevice_Type (?Device_ID, ?t)∧
hasthroughput (?Device_ID, ?r) ∧ has
∧ swrlb:lessThanOrEqual(?r, Bandwidth_threshold))
-> sqwrl:select(?r,?Device_ID)
∧ (Network(?Network_ID) ∧hasDevice_ID (?Network_ID, ?Device_ID)∧
hasBandwidth (?Device_ID, ?b) ∧
has ∧ swrlb:lessThanOrEqual(?b, Bandwidth_threshold))
-> sqwrl:select(?b,?Device_ID)
-> SelectSecurity(?s,?p,?r,?b, ?x, ?Level_Name)
```

FIGURE 7. *Reasoning Rules for Security Level Selection.*

3. The *Storage_Media* superclass supports various data-transfer rates. Therefore, Throughput (r) is measured and then compared with the pre-defined Throughput (r) threshold ranges and classified as very low, low, medium, or high.

4. The Network class has the properties of Network_ID and Network_Bandwith, therefore, the Network_Bandwidth (b) is captured and compared with the predefined Bandwidth threshold and classified as low, medium, or high.

5. Finally, from the obtained resultant classification of Storage_Capacity (s), Processing Memory *(m),* Power_Battery *(p),* Throughput *(r),* and Network_Bandwidth (b) obtained from the above steps, the reasoner in the reasoning engine will identify which Security_Level (x) should be selected.

Moreover, the threshold values of selected parameters/characteristics of devices are explicitly classified (using an expert-derived evaluation) as very small, small, large, or unlimited for device storage [GB], battery power [watts], and throughput [Mbps], and bandwidth [MHz] parameters. The power consumption is considered when the devices are in active states. Thus by expert-derived evaluation, the device-specific parameters values are defined as follows:

Critical (Storage capacity $\leq$ 1 GB)
Low (Storage capacity $>$ 1 GB and $\leq$ 64 GB)
Medium (Storage capacity $>$ 64 GB and $\leq$ 500 GB)
Large (Storage capacity $>$ 500 GB and $\leq$ 10 TB)
Unlimited (Storage capacity $>$ 10 TB)

The values for power consumption are defined on the basis of the following expert-derived characteristics:

Critical (Power consumption $\leq$ 5 Watts)
Low (Power consumption $>$ 5 Watts and $\leq$ 15 Watts)
Medium (Power consumption $>$ 15 Watts and $\leq$ 50 Watts)
High (Power consumption $>$ 50 Watts and $\leq$ 75 Watts)
Very High (Power consumption $>$ 75)

The values for network capabilities are define low, medium and high on the basis of the following expert-derived characteristics:

Low (Bandwidth $\leq$ 2.5 Mbps)
Medium (Bandwidth $>$ 2.5 Mbps $\leq$ 100 Mbps)
High (Bandwidth $>$ 100 Mbps)

Furthermore, the privacy value is defined as ranging between high and low. The security levels defined based on the above-mentioned device and network specific parameters along with the privacy levels are listed in Table 2. For the interested reader, the parameter threshold ranges for the Storage capacity attribute of the *Storage_Media* class in the implementation of the framework are given in Appendix A.

As an example, a use case scenario assumes that there is a device, belonging to the Storage_Media class, reports its storage capacity as 0.5 [GB], battery power 3 [Watts], throughput 2 [Mbps], and bandwidth 2.5 [Mhz]. Following the aforementioned steps, in the first step, storage capacity is classified as critical by comparing the reported value, which is 0.5, with the storage threshold. In the second step, the battery power is classified as critical, comparing the reported value 3 with the threshold. Similarly, the throughput and bandwidth are classified as very low and low respectively in the third and fourth steps. Finally, taking all these findings, the reasoner selects Level_1 for the security class. The selected security level is provided as input in feature selection in the security-encoding module.

E. FEATURE SELECTION AND SECURITY ADAPTATION MODULE

The feature selection and security adaption module is responsible for FOI selection by taking the security level selected by the reasoning engine as input. In MuLViS, FOI (i.e. motion, texture, face, skin, human/people (full body), and background) adaptation concerning device characteristics at each security level is described below and shown in Table 2.

L1 Security: In this study, motion is considered as an FOI at L1 for encryption of surveillance video data when dealing with simple, low-resolution videos captured by constrained devices. One form of selective encryption [65] is to encrypt only certain syntax elements output by the final stage, the entropy coder, of a compression codec, such as H.264/Advanced Video Coding (AVC) [66] or High Efficiency Video Coding (HEVC) [67]. Therefore, additionally, to further reduce the computational complexity of encryption only syntax elements Motion Vector Difference (MVD) at L1 are selected for encryption.

L2 Security: Texture Coefficients (TC) at level 2 are encrypted. Restricting encryption to MVD and/or TC syntax elements also achieves compression decoder format compliance (with the requirements of the standardized codec) together with an on average reduction in the bitrate overhead from encrypting the video.

L3 Security: To protect the privacy of individuals, the face and human are considered FOI for encryption at L3. However, if the height of the camera is too high or the pose variations of the person such as side-on poses or poses that are not in the FOV of the camera or due to a variety of causes such as blurriness, low resolution, illumination variations then face

detection will be unreliable or will not be detected within the video scene. In such a scenario, herein, human (full human body) are considered as FOIs for security and privacy protection at L3.

L4 Security: In the model, the background at L4 is considered as an FOI, to protect the privacy of the location.

L5 Security: At L5 partial-full security is employed for the protection for resources sufficient devices

Moreover, In the model, devices at levels above LI and L2 might be lumped together so that in devices at levels L3, L4 and L5, advanced object detection and background subtraction algorithms can be implemented to achieve sufficient security at the respective level.

F. VIDEO CHUNKING AND TAGGING MODULE

In this module, chunking and tagging are implemented. In chunking, similar FOI is grouped into chunks and then tagging with semantic descriptors defined in the SSSO is performed. This makes it possible for a user to quickly search for the desired video chunk and quickly have access to that chunk.

G. LIGHTWEIGHT PARTIAL ENCRYPTION MODULE

In this module, the tagged video chunk is encrypted by partial encryption, i.e. only FOI encryption, to secure the surveillance information. Thus, in this way, partial encryption will ensure its effectiveness in reducing computational and memory costs because the data encrypted is reduced in size in comparison with encrypting all of the video data. In this work, partial encryption on an FOI is implemented by applying the well-known industry standard symmetric cipher, Advanced Encryption Standard (AES) [70] in Output Feedback (OFB) operating mode, i.e. in a stream cipher mode, as described shortly. Partial encryption only takes place on selected part of a video frame, reducing the impact of AES encryption, which can be computationally expensive. The computational complexity issue of AES relative to lightweight ciphers is addressed by implementing partial encryption with the AES block-based encryption, similar to that previously proposed by the author [69], [70] of the current paper, which is a relatively lightweight form of encryption. Thus, the current paper is a more suitable choice for resource-constrained devices.

Advanced Encryption Standard: AES (also known as Rijndael) has been widely deployed as an encryption standard since 2000. AES is considered to be a secure industry-standard cipher and, hence, is extensively utilized for confidentiality in cyber-physical systems [71]. AES is extensively used because of its security, ease of implementation, defense against threats, flexibility in the case of encryption/decryption and keying material. AES is a symmetric key block cipher, which uses a 128-bit key for 10 rounds, a 192-bit key for 12 rounds, or a 256-bit key for 14 rounds of operation. AES processes data in the form of 4×4 matrix known as states. In AES, every round comprises four stages/phases: (1) Byte-substitution, (2) Mix Columns, (3) Shift Rows and (4) Add Round Key. As AES is a symmetric block cipher, so a single key is used for the encryption and decryption processes. It also considered a robust algorithm that can resist many attacks.

The symmetric encryption keys are generated at run-time for each protected video, by using a pseudo-random function (PRF). Furthermore, 128-bit key is secure enough as in current computing powers a key space greater than 2^{100} is considered resilient to brute-force and key guessing attacks over keys [72]. Key security can be further enhanced by using an established chaos-based key randomization scheme [73] or by standard key management schemes [74] in future. (Presently, key security is not emphasized, as it will also increase the computational cost of using Raspberry Pi based surveillance devices, with current maximum processor speed of only 1.6 GHz.)

Output Feedback (OFB): AES can be implemented from a choice of multiple operational modes [75]. This research implemented the AES with OFB mode of operation as OFB has the same code for both encryption and decryption process, resultantly saves the coding space. Another reason of choosing AES with OFB mode is that it also operates as a stream cipher (rather than a block cipher), in which few bits/bytes can be encrypted rather than a complete block. In OFB, X_{t-1} is an input block from the t-1 stage, which has been AES encrypted, using secure key K_s. Then X_{t-1} is again AES encrypted using key K_s to produce X_t. After that X_i and the next plaintext block P_t are XORed together to output encrypted block C_t. For encryption of the following plaintext block, AES encryption with K_e is again performed on the X_t of the previous stage to produce X_{t+1}, and then XORing is performed with the plaintext P_{t+1} to output C_{t+1} and so on. Moreover, OFB generates different output C_t for the identical input P_t because of the random initialization vector IV. The following equations represent the encryption and decryption processes in OFB mode, respectively.

$$C_t = P_t XOR X_t \tag{2}$$

$$P_t = C_t XOR X_t \tag{3}$$

where t=*1, 2,3, ...n*, for n stages of block encryption, and $X_t = \{\text{Encrypt}\ (K_e(X_{t-1}))\}$

Any modifications to a plaintext block P_i are reflected in the corresponding ciphered block C_t, where $t = 1, 2, 3 \ldots n$ with n the number of plaintext blocks, but other ciphered blocks remain unaffected. The OFB mode is error-resilient in that if any modification/error occurs during the transmission, that error is not propagated. Therefore, AES-OFB is suitable for real-time smart surveillance applications.

In fact, surveillance systems usually transport video in a compressed form. However, compression latency then occurs because of compression computation. Therefore, in surveillance real-time applications, H.264/AVC is now widely adopted for surveillance [76], rather than HEVC, due to its low latency (4 ms to 8 ms on average per video frame) and relatively higher compression ratio (50% or more). Once multi-level FOI encryption has taken place, the encrypted

TABLE 2. FOI adaptation with respect to security level.

Security Level	Storage	Throughput	Bandwidth	Power	Security Rank	Features of Interest (FOI)
L1	Critical	Very Low	Low	Critical	Very Low	Motion
L2	Low	Low	Low	Low	Low	Texture
L3	Medium	Medium	Medium	Medium	Medium	Face or Human (Full body)
L4	Large	High	Medium	High	High	Background
L5	Unlimited	Very High	High	Very High	Very High	Motion, Texture, Face, Human and Background

TABLE 3. Summary of test videos configuration.

Sr. #.	Video sequence	Size (MB) (YUV)	Size (MB) H.264	Resolution (pixels/frame)	Frame Rate (fps)	Frame Count
1.	Miss America	5.43	0.016	QCIF (176×144)	30	150
2.	Crew	43.5	0.37	CIF (352×288)	30	300
3.	Video 1	962	1.25	VGA (640×480)	30	2191
4.	Video 2	131	0.51	EDTV (720×480)	30	2821
5.	PETS09-S2L1	503	1.70	PAL (768×576)	10	795
6.	PETS09-S2L2	275	2.12	PAL (768×576)	15	436
7.	Atrium	3040	2.43	SVGA (800×600)	7.5	4542
8.	AVG-Town Centre	333	1.49	qHD (960×540)	15	450
9.	Cricket	1924	7.00	HD 720 (1280×720)	30	1460
10.	MOT17-09	389	2.34	FHD (1920×1080)	30	525

video stream is then stored in the surveillance database for future use.

H. INFORMATION RETRIEVAL MODULE

In this final module, the required information can be retrieved in an encrypted form. The original information can be viewed after decryption with the cipher key. Consequently, only authorized person(s) can view sensitive and private data.

IV. EXPERIMENTAL RESULTS & DISCUSSION

This Section firstly describes the experimental setup and then the results of implementing the framework.

A. EXPERIMENTAL SETUP

The MuLViS was implemented as an in-house built simulator, using the C++ programming language. To show the feasibility of the proposal, MuLViS was tested on a general-purpose laptop, with Intel Core i5 CPU. The system is evaluated on publicly-available datasets, derf's collection (https://media.xiph.org/video/derf/), Urban Tracker [77] PETS2009 [78], MOT17 [79] and ABODA [80]. Representative video frames for each dataset are shown in Figure 8. Pixel resolution is another important consideration, as VS systems operate across a wide range of resolutions and frame rates. Higher resolution provides a better-quality bitstream, which enhances the ability to identify people and objects within surveillance videos. However, higher resolution video places a greater demand on network bandwidth, storage space, and energy consumption. Thus, in this work, experiments were performed on datasets of various resolutions (QCIF, CIF, VGA, EDTV, SVGA, qHD, HD, FHD) (see Table 3) and frame rates (ranging from 7 to 30 frames/s (fps)). The original MP4, AVI format video containers were processed in the YUV file format. FOI detection and encryption algorithms were implemented in an IPPPBB…frame structure with Group of Pictures (GOP) size of 16 frames and quantization parameter (QP) of 32 (refer to [81] for frame structure terms). For the FOI detection, the implemented detection schemes were discussed in Section III.3. To measure the impact of partial encryption on FOIs, the set of surveillance videos listed in Table 3 were used.

B. EXPERIMENTAL RESULTS

The different hardware and network related attributes of devices in operation are collected through the sensors. The devices in operation are represented by the Device_id (as d1, d2, d3 and so on). After that, for real-time device-specific security level selection data generated by a sensor is used to simulate the test results of the method. Through semantic mapping, an instance of the SSSO is stored into owl files, where the security concepts are mapped/integrated and their corresponding relationships the multimedia domain concepts. (Figure 4 briefly indicates the structure of the SSS ontology, as discussed in Section III.4.) After that, the auto-level selection is simulated and results are illustrated in Table 4. The rules defined for security level selection in the ontological reasoner were described in Section III.4. Column 10 with bold values in Table 4 represents the appropriate level

FIGURE 8. Representative test videos from benchmark datasets. (a-b) Videos from Xiph, (c-d) Videos from ABODA dataset, (e-f) PETS2009 dataset, (g-h) Videos from Urban Tracker, (i-j) Videos from MOT17 dataset and (k) Cricket video.

selected by the proposed method to protect the video with respect to (w.r.t.) device specifications.

After taking the resultant security level as input, encryption with the AES was implemented on the test videos to protect confidentially/privacy within the surveillance video bit-stream. Detailed visual results with FOI encryption at L1, L2, L3, L4 and L5 are given in Figures 9, 10, and 11. Visual results at L1 are presented in Figure 9b, 10b and 11b. L2 is shown in Figures 9c, 10c and 11c. L3 is shown in Figure 9 (d, e), Figure 10 (d, e) and Figure 11 (d, e). L4 is shown in Figures 9f, 10f and 11f and finally L5 appears in Figures 9g, 10g and 11g.

The results illustrate that the visual quality of an encrypted video stream declines significantly relative to the original videos. However, FOI encryption does not completely disrupt the video stream, as, in a real-time surveillance system, it can be beneficial to view content of low-sensitivity. Indeed, a low-quality preview of the actions performed by the objects within a video may allow event recognition to be performed without breaching the privacy of individuals.

To support the visual results illustrated previously and evaluate the quality of encrypted videos, objective quality analysis was performed by calculating the Peak Signal to Noise Ratio (PSNR) [82], Mean Square Error (MSE) and Structural Similarity (SSIM) index [83], [84]. Comparative PSNR, SSIM and MSE results of FOI encrypted bitstream for each security level (L1, L2, L3, L4, and L5) are provided in Table 5. A lower PSNR value indicates a greater distortion within the video. The results demonstrate that the average PSNR of FOI encrypted video streams at all security levels remain below 40 dB which implies sufficient security and protection has been achieved. Moreover, the results show that the average PSNR of motion encrypted FOI for the luma component (Y) is on average 19.04 dB and for the chroma components U and V [41] is on average 33.02 dB and 33.3 dB respectively for the MOT17-09 video. The low luma value of all motion encrypted bitstreams (less than 25 dB) demonstrates that considerable security has been achieved with a negligible bitrate overhead for constrained devices at L1.

Likewise, the average SSIM and MSE of motion-encrypted FOI presented in Table 5 lead to the same conclusions as for PSNR measurement. The privacy of individuals has been taken into account by considering the face and human (full body) at L3 (refer to Use Case 2). If the detection algorithm failed to detect the face (see Figure 10d) due to real-time factors such as camera location, zooming, and FOV, or various lighting and environmental conditions, then, in these scenarios, humans (see Figure10e) is selected as an FOI. Besides, at L3, FOI encryption (i.e. only face encrypted, only skin encrypted or human encrypted) leads to the privacy protection of individuals, while preserving their shape/structure (see Figure 10d and Figure10e). The latter helps in event/action recognition within the scene.

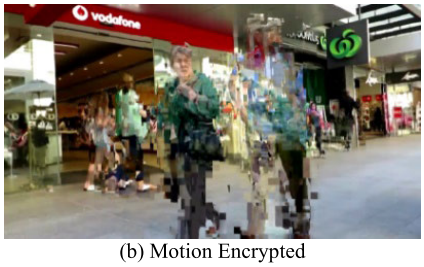
C. PERFORMANCE EVALUATION

This Section contains a detailed performance evaluation of the system.

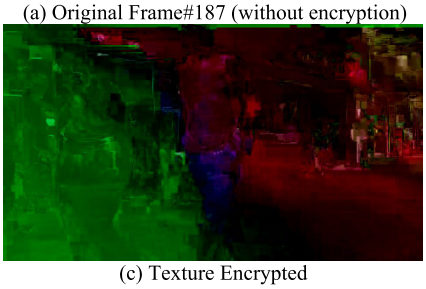
1) FOI DETECTION ACCURACY ANALYSIS

To assess the encryption accuracy aspect of the system, the detection percentage was calculated. The average face

(a) Original Frame#187 (without encryption)

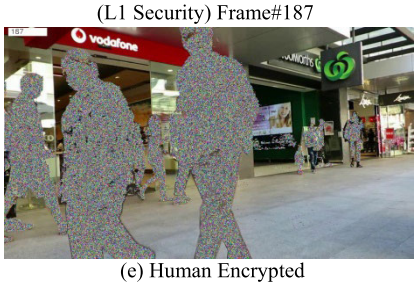


(b) Motion Encrypted (L1 Security) Frame#187

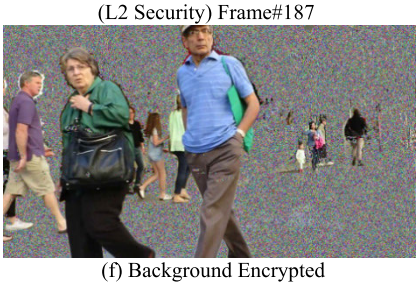


(c) Texture Encrypted (L2 Security) Frame#187

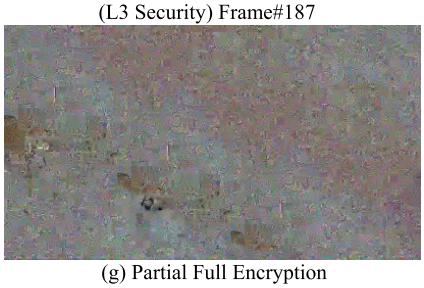
(d) Face Encrypted (L3 Security) Frame#187



(e) Human Encrypted (L3 Security) Frame#187



(f) Background Encrypted (L4 Security)Frame #187



(g) Partial Full Encryption (L5 Security) Frame #187

FIGURE 9. Visual results of Multi-Level FOI detection and encryption on the MOT17-09 video.

TABLE 4. Resulting security levels w.r.t devices in operation in a surveillance system.

Device_Id	Storage	Throughput	Resolution	Frame Rate	Compression Format	Band-width	Power	Security Rank	**Output Security_ Level**
d1	Medium	Medium	Medium	Medium	H.264	Low	Medium	Low	**L3**
d2	Critical	Very Low	Very Low	Very Low	H.264	Medium	Critical	High	**L1**
d3	Large	High	High	High	H.264	Medium	High	Medium	**L4**
d4	Low	Low	Low	Low	H.264	High	Low	Medium	**L1 and L2**
d5	Low	Low	Low	Low	H.264	Low	Low	Low	**L2**
d6	Medium	Medium	Medium	Medium	H.264	Medium	Medium	Medium	**L3**
d7	Medium	Medium	Medium	Medium	H.264	High	Medium	High	**L1 and L3**
d8	Unlimited	Very High	Very High	Very High	H.264	High	Unlimited	Very High	**L5**
d9	Large	High	High	High	H.264	High	High	High	**L1 and L4**
d10	Critical	Very Low	Very Low	Very Low	H.264	Low	Critical	Very Low	**L1**
d11	Medium	Medium	Medium	Medium	H.264	Low	Medium	Low	**L3**
d12	Medium	Medium	Medium	Medium	H.264	Medium	Medium	Medium	**L3**
d13	Critical	Very Low	Very Low	Very Low	H.264	High	Critical	High	**L1 and L4**
d14	Low	Low	Low	Low	H.264	Low	Low	Low	**L2**
d15	Low	Low	Low	Low	H.264	High	Low	Medium	**L1 and L2**
d16	Large	High	High	High	H.264	High	High	High	**L1 and L4**
d17	Unlimited	Very High	Very High	Very High	H.264	High	Unlimited	Very High	**L5**
d18	Medium	Medium	Medium	Medium	H.264	High	Medium	High	**L1 and L3**
d19	Large	High	High	High	H.264	Medium	High	Medium	**L4**
d20	Low	Low	Low	Low	H.264	Low	Low	Low	**L2**

detection rate as a percentage using the NPD face detector [38] algorithm adopted in this study (see Section III.3), for selected frames of the MOT17-09 video sequence is shown in Table 6. Table 7 averages the face detection rate for a number of reference video sequences. To compute the detection percentage of faces in the test videos, the True Positive Rate (TPR) and False Discovery Rate (FDR) are calculated by processing each frame of the video stream as (4) and (5).

$$TPR = TP/P \tag{4}$$

$$FDR = FP/(FP + TP) \tag{5}$$

TABLE 5. Average distortion in term of PSNR, SSIM and MSE for FOI encryption of sample surveillance videos.

Video	FOI Encrypted	PSNR (dB)			SSIM	MSE
		Y	U	V		
MOT17-09	Motion Encrypted	19.04226	33.02529	33.2722	0.77843	810.43
	Texture Encrypted	8.87031	23.57742	18.78884	0.29857	8434.25
	Face Encrypted	22.17336	34.46513	33.28676	0.92931	394.18
	Human (Full Body) Encrypted	18.88960	29.94356	28.96743	0.82797	839.36
	Only Background Encrypted	12.53844	25.19116	23.70526	0.18889	3624.49
	Motion and Background Encrypted	12.03817	25.09244	23.82488	0.16845	4066.99
	Partial Full Encryption	7.43616	8.20258	8.2064	0.042495	15557.08
PETS09-S2L1	Motion Encrypted	24.6406	38.8236	39.5858	0.8591	223.33
	Texture Encrypted	8.00488	23.82790	25.04036	0.27066	10294.17
	Face Encrypted	100.00	100.00	100.00	1	0.00000
	Human (Full Body) Encrypted	21.93251	36.2957	37.2552	0.8869	416.48
	Only Background Encrypted	12.68795	21.0085	24.3578	0.0607	3501.48
	Motion and Background Encrypted	12.93410	21.5822	26.1471	0.0629	3308.48
	Partial Full Encryption	7.2075	7.3485	7.7559	0.0893	12426.75
Atrium	Motion Encrypted	19.3393	35.4809	38.3059	0.7712	756.82
	Texture Encrypted	9.18209	24.6590	23.4647	0.2465	8995.18
	Face Encrypted	29.08717	33.55725	32.97499	0.97837	80.23483
	Human (Full Body) Encrypted	21.20160	28.44293	28.01466	0.8631	492.88
	Only Background Encrypted	10.80974	24.4509	27.3361	0.1516	5396.57
	Motion and Background Encrypted	10.59309	24.6891	28.3319	0.1545	5672.60
	Partial Full Encryption	6.75743	6.5195	6.9884	0.0539	15633.49
Cricket	Motion Encrypted	23.13694	38.9144	39.4745	0.8188	315.64
	Texture Encrypted	8.68836	17.43088	23.70684	0.19470	8795.15
	Face Encrypted	24.14696	26.69562	38.93598	0.96391	260.70893
	Human (Full Body) Encrypted	25.38987	36.7513	35.8129	0.9187	187.96
	Only Background Encrypted	14.55583	26.2100	24.8937	0.1065	12133.71
	Motion and Background Encrypted	14.83823	28.2437	26.6284	0.1023	12277.18
	Partial Full Encryption	7.13058	7.5247	7.3242	0.1167	14308.97

where 'TP' is the number of true positives i.e. number of faces correctly identified and encrypted, 'FP' is the number of false positives i.e. number of faces incorrectly identified and encrypted' and P is the total number of faces within the frame. The Precision is the proportion of the true positives results against all positive results and calculated as (6). The Accuracy is calculated as (7) for the test videos.

$$Precision = TP/(TP + FP) \tag{6}$$

$$Accuracy = (TP + TN)/(TP + FP + TN + TP) \tag{7}$$

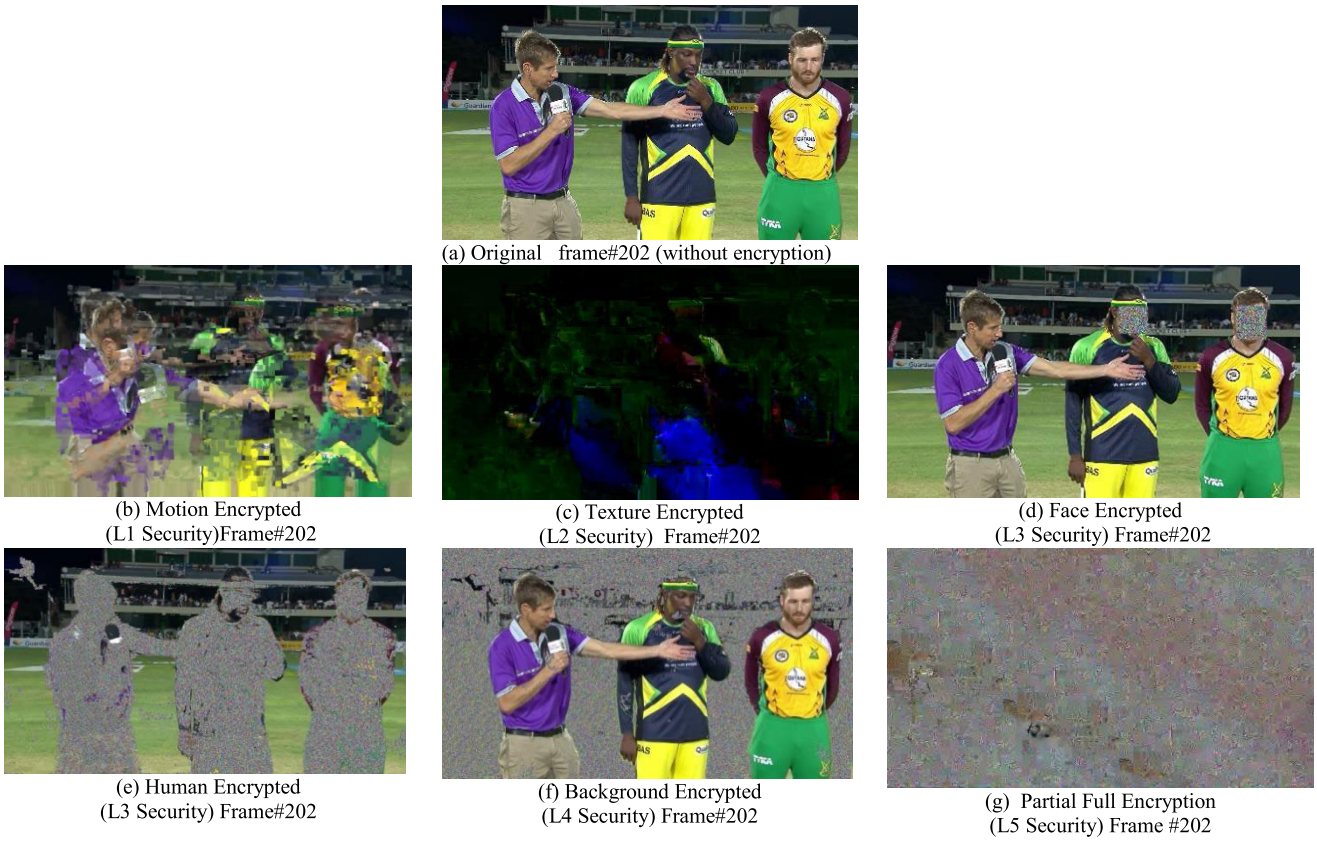
Likewise, the human count accuracy over the entire video was measured by investigating the missed number of human/people and incorrectly detected human. For the human count accuracy, the Measurement Multiple Object Count (MOC) metric is utilized. MOC is calculated as [85]:

$$MOC = 1 - ((f_n + f_p)/T_f) \tag{8}$$

where f_n is the total number of humans not identified or missed and f_p is the total number of incorrectly detected humans, whereas, 'T_f' is the number of humans (ground truth) present in the entire video. The average human count accuracy of the sample test videos appears in Figure 12. The results show that the performance of the method for human detection is better in the test videos where detection accuracy has an average of $\approx$ 89.9 %. Therefore, where the face detection algorithm suffers when faces are blurred, out of focus, or the camera capturing the video is installed far away (at a long distance or at a height) from the surveillance location (refer to Use Case 2 and see Figure 11 (d)) than human is selected for sufficient security at L3.

TABLE 6. Average Face detection rate (%) of MOT17-09 video.

FRAME#	DETECTION RATE (%)	FALSE DETECTION RATE (%)	PRECISION (%)
71	83.3 %	16.6 %	82.7%
109	80%	20 %	80%
187	66.6%	16.6 %	80%
190	71.4%	14.2 %	83.3%
227	75%	25 %	75%
415	100%	0 %	75%
443	100%	33.3 %	100 %
487	100%	0 %	100 %
AVERAGE	**86.5**	**13.7 %**	**84.26 %**



(a) Original frame#202 (without encryption)

(b) Motion Encrypted (L1 Security)Frame#202

(c) Texture Encrypted (L2 Security) Frame#202

(d) Face Encrypted (L3 Security) Frame#202

(e) Human Encrypted (L3 Security) Frame#202

(f) Background Encrypted (L4 Security) Frame#202

(g) Partial Full Encryption (L5 Security) Frame #202

FIGURE 10. Visual results of multi-level FOI detection and encryption on the Cricket video.

TABLE 7. Average face detection rate (%) for 100 frames of a number of reference video sequences.

SR.NO	VIDEO	TPR (%)	FPR (%)	PRECISION (%)
1.	MISS AMERICA	100%	0%	100%
2.	CREW	94.72%	3.49%	96.44%
3.	VIDEO 1	87%	5%	94.56%
4.	VIDEO 2	17.46%	0%	100%
5.	PETS09-S2L1	4.31%	0%	100%
6.	PETS09-S2L2	7.01%	1.85%	77.54%
7.	ATRIUM	94%	4.75%	93.91%
8.	AVG-TOWN CENTRE	8.52%	0%	100%
9.	CRICKET	89.53%	11.98%	87.19%
10	MOT17-09	86.5%	13.7%	84.26%

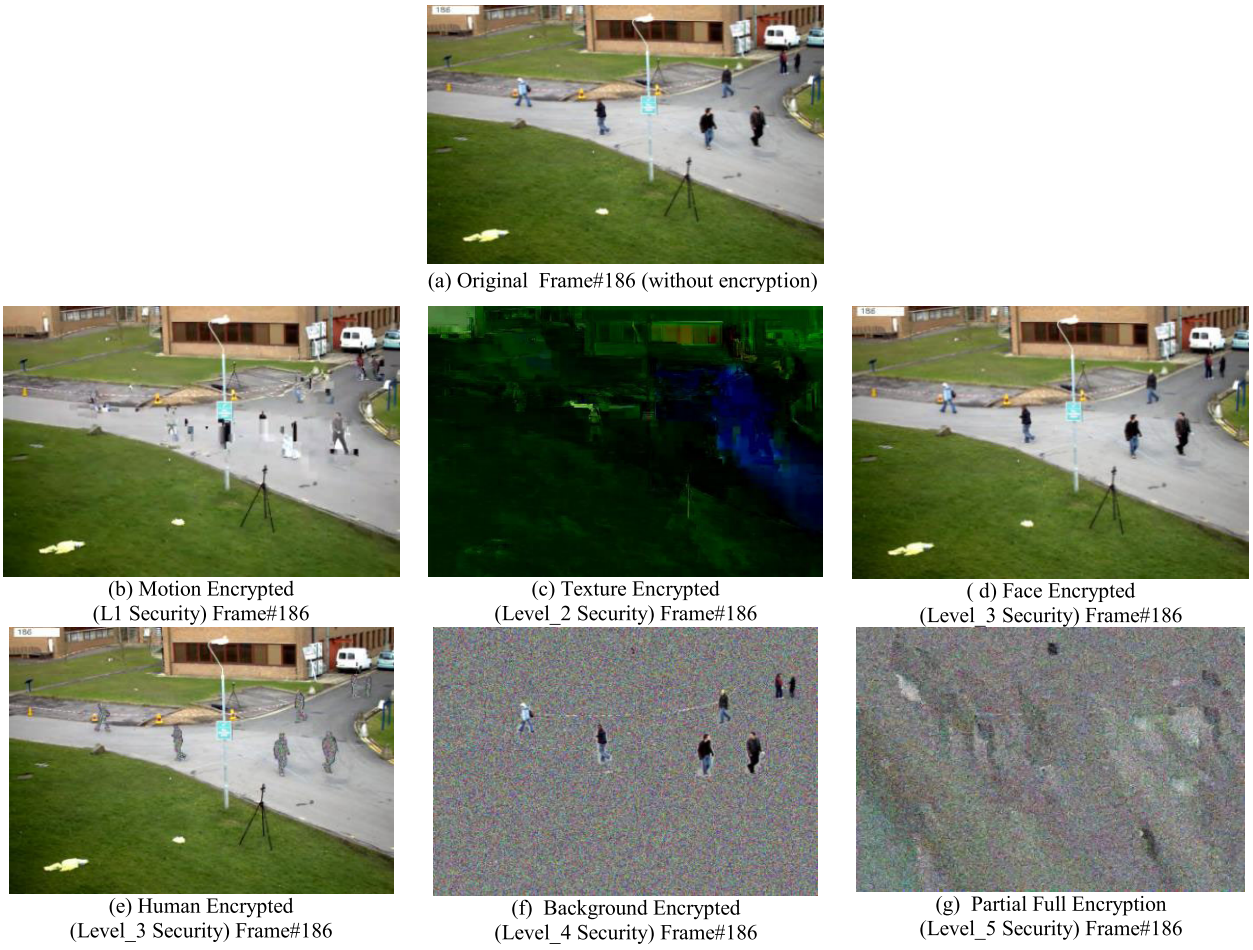
2) ENCRYPTION SPACE RATIO ANALYSIS

To evaluate the performance of the system, the average Encryption Space Ratio (ESR), i.e. the ratio of bits encrypted to bits not encrypted in a selective or partial encryption scheme, see [86] or [87] for an example of ESR in use. The average ESR of sample test videos for each level is provided in Table 8. For more clarity, the ESR for MOT17-09 is illustrated in Figure. 13. It can be noticed from Figure 13 that the average ESR of only the motion syntax element encrypted at L1 for constrained devices is only 0.04 %, i.e. scarcely any bits, as a proportion of the whole, are encrypted.

The comparative results are shown in Table 8 imply that the ESR of the FOI (i.e. with only MVD encryption) at L1 is very small for all the videos. However, the results also illustrate that the ESR of background encryption is much higher (an average of $\sim$ 86%) as compared to the ESR of the FOIs encrypted at L1, L2, and L3. However, in some scenarios, the background is considered to be a sensitive FOI (refer to Use Case 3). Hence, the background is considered as an FOI but only for devices with considerable resources and with high computational performance (see Figures. 9 (f), 10 (f) and 11 (f)).

3) BITRATE OVERHEAD ANALYSIS

Furthermore, FOI encryption has different bitrate overhead impacts at every security level. The bitrate overhead as a result of encryption for smart devices should be negligible, as compared to that of medium to very high storage capacity devices. Thus, the estimated bitrate overhead on average for each video at each level is illustrated in Figure 14. The results imply that the encryption performed for low-resource devices does not generally affect the bitrate. However, encryption for medium to very high capacity storage devices introduces some bitrate overhead. Moreover, the encryption overhead introduced is video content dependent. The results show that the average bitrate overhead of MOIT09 video stream for



(a) Original Frame#186 (without encryption)

(b) Motion Encrypted (L1 Security) Frame#186

(c) Texture Encrypted (Level_2 Security) Frame#186

(d) Face Encrypted (Level_3 Security) Frame#186

(e) Human Encrypted (Level_3 Security) Frame#186

(f) Background Encrypted (Level_4 Security) Frame#186

(g) Partial Full Encryption (Level_5 Security) Frame#186

FIGURE 11. Visual results of multi-level FOI detection and encryption on PETS09-S2L1.

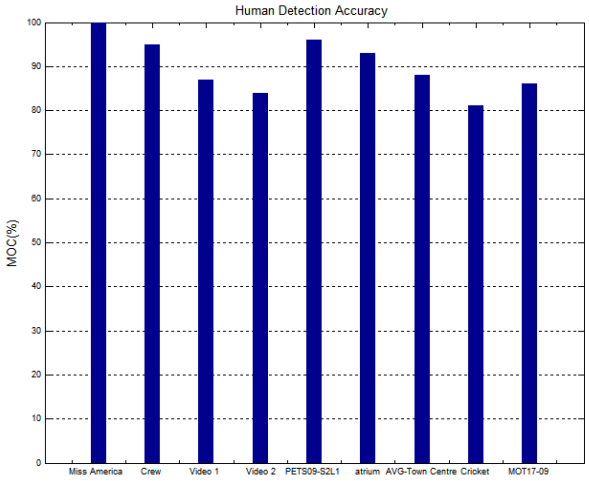


FIGURE 12. Comparison of average human count accuracy over the entire sample test videos.

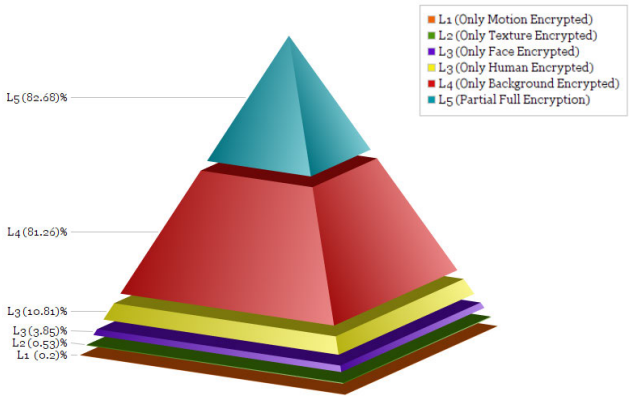


FIGURE 13. Average ESR at each security level of the MuLViS system for the MOT17-09 video.

only motion and texture encryption (which provides L1 and L2 security) is zero, for face and human encryption (which provides L3 security) is an average of 0.24 % and 3.1 % respectively, which is quite reasonable. However, results show that bitrate overhead for background (which provides L4 security) is 8.3 % for the MOT17-09 bit-stream, which is high as compared to L1 encryption for constrained resources devices. However, for high resources and high performance devices this bitrate overhead is manageable.

TABLE 8. Average ESR at each security level of MuLViS.

Sr. #	Video	ESR (%)					
		L1	L2	L3		L4	L5
		Motion Ency.	Texture Ency.	Face Ency.	Human Ency.	Background Ency.	Partial Full Encryption
1.	Video 1	0.02%	0.11%	1.03%	1.63%	92.04%	93.27%
2.	Video 2	0.03%	0.33%	2.21%	2.8%	90.72%	91.66%
3.	PETS09-S2L1	0.31%	0.07%	0%	1.89%	94.37%	94.94%
4.	PETS09-S2L2	0.19%	0.69%	0%	3.73%	91.45%	89.26%
5.	Atrium	0.02%	0.057%	1.01%	3.17%	94.14%	96.00%
6.	AVG-Town Centre	0.04%	0.03%	0.0002%	3.35%	86.64%	87.35%
7.	Cricket	0.20%	0.24%	2.33%	11.43%	78.91%	79.46%
8.	MOT17-09	0.20%	0.53%	3.85%	10.81%	81.26%	82.68%
	Average	**0.13%**	**0.26%**	**1.30%**	**4.85%**	**88.69%**	**86.20%**

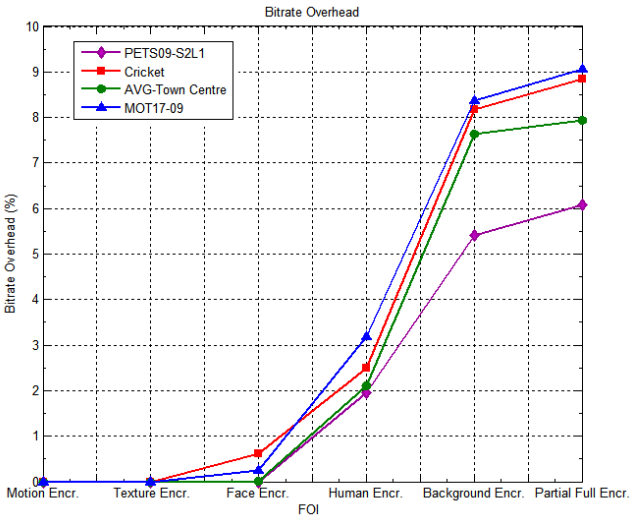


FIGURE 14. Comparative average bitrate overhead introduced at each security level of MuLViS for text videos.

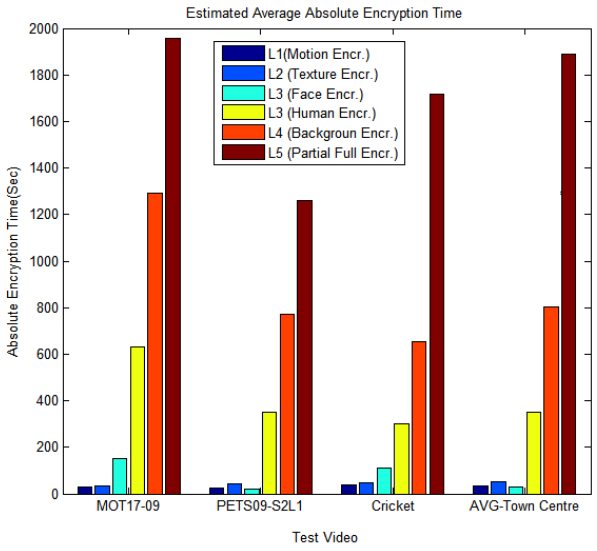


FIGURE 15. Average absolute encryption time of Partial FOI encryption.

4) TIME COMPLEXITY ANALYSIS

Finally, the performance of MuLViS is evaluated in terms of time complexity. However, notice that these timings are intended to be indicative rather than definitive, since different timings could result from employing other hardware than the laptop specified in Section IV.1. The computational time itself is calculated by adding the FOI extraction time (for background, face, motion, human, or objects), the compression time of the H.264/AVC codec and the absolute encryption time. The absolute encryption time for each level is illustrated in Figure 15. The Absolute Encryption Time (AET) is calculated as:

$$AET = Total\ Execution\ Time - Detection\ Time \quad (9)$$

The results show that the AET for security levels L1 and L2 (i.e. only motion or texture syntax elements are encrypted) adopted for constrained resources devices is significantly low, being on average 2.7% of the total execution time (i.e. motion detection, compression time, and encryption time). Similarity, the average absolute encryption time ratio is an average of 9.7% of the total execution time of L3 (i.e. human encrypted) considered for medium capacity resource devices. While the average absolute encryption time ratio for L4 (background encrypted) adopted for high capacity resource devices is an average of 16.4 % of the total execution time. The average absolute encryption time for L5 security (i.e. partial encryption) adopted for very high capacity resource devices is an average of 25.2% of total execution time, which is a significant amount.

The results imply that an additional 2.7% time is needed for the encryption of selected FOIs (from 1033.43 to 1062.23 seconds) for low capacity devices. That time can easily be tolerated by the constrained devices. However, the average absolute encryption time going between low-capacity devices and very high capacity devices increases from 28.80 to 1959.14 seconds which is significantly higher than the absolute encryption time at L1 and L2 (i.e. only motion or only faces are encrypted), when adopted for the MOT17-09 video sequence. The small encryption time reflects the fact that partial encryption (i.e. full FOI encryption) adopted for low

capacity device is much simpler, as compared to the measures adopted for high-performance devices.

5) COMPARATIVE ANALYSIS

In this Section, the proposed scheme and other security approaches proposed by researchers to protect visual content in a smart infrastructure are compared. Comparison parameters were chosen that indicate the positive features of MuLViS in terms of security, as well as how well other schemes also meet those parameters. The parameters chosen for the comparison and justifications for their choice are as follows:

Confidentiality: This parameter indicates whether encryption is applied to the surveillance data or not. Notice that other forms of privacy protection such as pixilation or blurring can be applied and are applied in a number of market-based systems. There are many other forms of privacy protection such as mosaic, masking, and morphing. For a description of these and other privacy protection methods refer to [13].

Computational Overhead: Computational overhead specifies the total processing time required for identifying and encrypting the FOI (see also Section IV.C.4). A higher level of computing time means that the computational overhead is also high and vice versa. A higher computational time could impede real-time responses or require costly hardware to achieve a real-time response.

Compression: This implies that compression is applied, which is an important consideration for network transmission and storage, particularly as even with compression, video streams occupy considerable bandwidth and, when stored, considerable memory.

Format Compliance: This parameter defines whether the encrypted video sequences are consistent with the standard for an H.264/AVC or HEVC decoder. Standard compliance allows intermediate devices in a network path to handle video streams without a need to decrypt those video streams [82]. Example intermediate devices include: video transcoders to change the bitrate of a video stream when a network link has reduced bandwidth and video splicers to insert logos or watermarks in the compressed domain.

Intelligibility: This indicates whether further processing (computer vision tasks such as recognitions of events i.e. is walking, running, fighting etc.) is achievable on protected videos, without the need for decryption of the protected parts of a video.

Reversibility: The reversibility parameter indicates whether the protected video can be decrypted by an authorized person possessing the encryption/decryption key. Notice that some forms of privacy protection, including blurring and pixilation are not reversible. This implies that such video might not be acceptable in a court. Notice also that encryption is the only reversible protection solution designated in the articles of the European Union's GDPR [13].

A comparison of the proposed system against other approaches is summarized in Table 9. The proposed scheme meets a good number of the requirements that might be required for a secure smart surveillance system.

V. CONCLUSION AND FUTURE WORK

The video data generated by surveillance cameras and sensors on a daily basis require effective security measures to ensure data security and privacy. Some prior studies on the cryptographic protection of surveillance video may be insufficient (though not redundant) because they do not allow both for the amount of data that must now be processed and the diversity of devices in operation within the surveillance system. This work presented an innovative, multi-level security system in which ontology is integrated for selection of a suitable security level. This is achieved by judging a sensor device's characteristics and network requirements. In that way, at least in terms of preserving the privacy of objects, people, or locations within surveillance video streams, the required encryption processing can be matched to device capabilities and scaled according to the amount of data that the device is capable of processing. The proposed framework is a unique blend of technologies i.e. ontology, computer vision and encryption over visual data for real-time smart surveillance systems.

Extensive experiments were used to evaluate different aspects of the performance of the proposed framework. The objective quality metrics i.e. PSNR, SSIM and MSE were calculated for statistical visual degradation of videos. The performance of MuLViS was evaluated by the detection accuracy of all FOI, i.e. face, human, background etc. with respect to the ESR on each security level. The positive detection rate of human faces in MOT17-09 video was found to be 86.5% and the false detection rate was 13.7% with the NPD face detection algorithm. The human accuracy count results via the MOC metric on all tested videos showed that the performance of the proposed method for human detection was better in the test videos with average detection accuracy $\approx$ 89.9 %. ESR was used as a tool to detect the ratio of encrypted bits vs. non-encrypted bits of videos on each security level. The comparative results of average ESR implied that the ESR of the motion FOI at L1 was just 0.13% (minimal) for all tested videos while the face and human encryption at L3 was 1.30% and 4.85% respectively, which was again quite low and easily computed by constrained surveillance devices. However, the results also illustrated that the ESR of background encryption was high (an average of $\sim$ 86%) as compared to the ESR of the FOIs encrypted at L1, L2, and L3. Hence, the background L4 and full partial encryption L5 can be considered for moderate to high computational devices.

For the effective computation of MuLViS on smart cameras, the bit-rate overhead and results with absolute encryption times were calculated, which were demonstrated to be manageable for constrained devices for L1, L2, and L3 security levels. In this paper, the most recent industry standard cipher AES-OFB with a 128-bit key was implemented to provide a non-breakable secure solution for smart

TABLE 9. Comparative Evaluation of the Proposed Solution.

Encry. scheme	FOI	Confidentiality	Comput. overhead	Compression	Format compliance	Intelligibility	Reversibility
[54]	Fixed	Yes	Medium	Yes	Yes	No	Yes
[55]	Fixed	Yes	High	Not defined	Not defined	No	Yes
[88]	Fixed	Yes	Low	Yes	Yes	No	Yes
[89]	Fixed	Yes	Medium	Not defined	No	No	Yes
Proposed MuLViS	**Adaptive**	**Yes**	**Flexible**	**Yes**	**Yes**	**Yes**	**Yes**

surveillance systems. However, AES can be computationally expensive for constrained devices. Therefore, to avoid the impact of cipher complexity, a single round-cipher, i.e. real-time eXclusive OR (XOR) cipher can alternatively be deployed (for moderate security strength) to further reduce the bit-rate overhead and also reduce the absolute encryption time.

From all these findings, it can be concluded that this paper provides a practical solution for privacy-protected surveillance systems in accordance with the current data protection laws within the EU related to visual surveillance data, namely the GDPR framework. The novelty of this data protection-by-design solution is that multi-level privacy protection for each surveillance video has not so far been targeted by previous research. All countries in the EU are obliged to comply with the GDPR. Thus, the proposed solution can be adopted by Smart Cities with confidence that the solution fulfills the data protection laws concerned with individuals' privacy within Europe.

There is the possibility of enlarging the scope of the ontology to include aspects other than encryption. For example, the level of authentication checks or the level of encryption key management can be incorporated. It also appears that computational intelligence may have a role in better selecting the security level according to device characteristics. The whole is a way forward in the context of research into surveillance and the security that evidently needs to be put in place. In future work the deep learning based detected algorithms will be incorporated in MuLViS. Alternatively, as a lighter-weight, real-time reasoning system, fuzzy logic can be adopted, as it already includes expert-derived modelling and rules for combining individual models.

APPENDIX A

The parameter threshold ranges for the *Storage capacity* attribute of the Storage_Media class are classified as follows:

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID, ?s) ^swrl: lessThanOrEqual((?s,1) -> Critical (?s)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID, ?s) ^swrl: greaterThan((?s,1)^swrl: lessThanOrEqual((?s,64) ->Low (?s)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID, ?s) ^swrl: greaterThan((?s,64)^swrl: lessThanOrEqual((?s,500) ->Medium (?s)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID, ?s) ^swrl: greaterThan((?s,500)^swrl: lessThanOrEqual((?s, 80000) ->Large (?s)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID, ?s) ^swrl: lessThan((?s, 80000) ->Unlimited (?s)

The threshold ranges for the *Power* attribute of Storage_Media are classified as follows:

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasBattery_Power(?Device_ID, ?p)^swrl: lessThanOrEqual((?p,5) -> Critical (?p)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID,?p) ^swrl: greaterThan((?p,5)^swrl: lessThanOrEqual((?p,15) ->Low (?p)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID,?p) ^swrl: greaterThan((?p,15)^swrl: lessThanOrEqual((?p,50) ->Medium(?p)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID,?p) ^swrl: greaterThan((?p,50)^swrl: lessThanOrEqual((?p, 75) ->High (?p)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t) ^hasStorage_Capacity(?Device_ID,?p) ^swrl: lessThan((?p,75) ->Unlimited (?p)

The threshold ranges for the *Throughput* attribute of the Storage_Media class are classified as follows:

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t)^hasthroughput(?Device_ID,?r)^swrl: lessThanOrEqual((?b,2.5) ->Very Low (?b)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t)^ hasthroughput (?Device_ID,?r)^swrl: greaterThan((?r,2.5) ^swrl: lessThanOrEqual((?r,50) -> Low(?r)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t)^ hasthroughput (?Device_ID,?r)^swrl: greaterThan((?r,50) ^swrl: lessThanOrEqual((?r,100) ->Medium(?r)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t)^ hasthroughput (?Device_ID,?r)^swrl: greaterThan((?r,100) ->High(?r)

The threshold ranges for the *Bandwidth* attribute of *Network* class are classified as follows:

SSSO:Network(?Network_ID)^hasDevice_ID (Network_ID,?Device_ID)^hasBandwidth(?Device_ID,?b) ^swrl: lessThanOrEqual((?b,5) ->Low (?b)

SSSO: Network(?Network_ID)^hasDevice_ID (Network_ID,?Device_ID)^hasBandwidth(?Device_ID,?b) ^swrl: greaterThan((?b,5) ^swrl: lessThanOrEqual((?b,10) -> Medium(?r)

SSSO:Storage_Media(?Device_ID)^hasDevice_Type (?Device_ID,?t)^hasBandwidth(?Device_ID,?r)^swrl: greaterThan((?b,10) ->High(?b)

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments in improving the quality of the article and in enhancing the readability and usefulness of our manuscript.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013, doi: 10.1016/J.FUTURE.2013.01.010.

[2] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient IoT-based sensor BIG data collection–processing and analysis in smart buildings," *Future Gener. Comput. Syst.*, vol. 82, pp. 349–357, May 2018, doi: 10.1016/j.future.2017.09.082.

[3] M. Rouse, IoT Agenda. (2015). *IoT Security (Internet of Things Security)?* Accessed: Aug. 5, 2020. [Online]. Available: https://internetofthingsagenda.techtarget.com/definition/IoT-security-Internet-of-Things-security

[4] J. L. Hernández-Ramos, M. V. Moreno, J. B. Bernabé, D. G. Carrillo, and A. F. Skarmeta, "SAFIR: Secure access framework for IoT-enabled services on smart buildings," *J. Comput. Syst. Sci.*, vol. 81, no. 8, pp. 1452–1463, Dec. 2015, doi: 10.1016/j.jcss.2014.12.021.

[5] M. N. Asghar, N. Kanwal, B. Lee, M. Fleury, M. Herbst, and Y. Qiao, "Visual surveillance within the EU general data protection regulation: A technology perspective," *IEEE Access*, vol. 7, pp. 111709–111726, Aug. 2019, doi: 10.1109/access.2019.2934226.

[6] *General Data Protection Regulation (GDPR)–Official Legal Text*. Accessed: Aug. 5, 2020. [Online]. Available: https://gdpr-info.eu/

[7] T. Zhang, A. Chowdhery, P. Bahl, K. Jamieson, and S. Banerjee, "The design and implementation of a wireless video surveillance system," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2015, pp. 426–438, doi: 10.1145/2789168.2790123.

[8] S. Fleck and W. Strasser, "Smart camera based monitoring system and its application to assisted living," *Proc. IEEE*, vol. 96, no. 10, pp. 1698–1714, Oct. 2008, doi: 10.1109/JPROC.2008.928765.

[9] Y. Wang, S. Velipasalar, and M. Casares, "Cooperative object tracking and composite event detection with wireless embedded smart cameras," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2614–2633, Oct. 2010, doi: 10.1109/TIP.2010.2052278.

[10] B. Furht and D. Kirovski, *Multimedia Encryption and Authentication Techniques and Applications*. Boca Raton, FL, USA: Taylor & Francis, Auerbach Publications, 2006.

[11] M. Collotta, G. Pau, and D. G. Costa, "A fuzzy-based approach for energy-efficient Wi-Fi communications in dense wireless multimedia sensor networks," *Comput. Netw.*, vol. 134, pp. 127–139, Apr. 2018, doi: 10.1016/j.comnet.2018.01.041.

[12] A. Shifa, M. N. Asghar, M. Fleury, and M. S. Afgan, "Ontology-based intelligent security framework for smart video surveillance," in *Proc. Future Technol. Conf. (FTC)*, 2019, pp. 118–126, doi: 10.1007/978-3-030-02683-7_10.

[13] D. Potoglou, F. Dunkerley, S. Patil, and N. Robinson, "Public preferences for Internet surveillance, data retention and privacy enhancing services: Evidence from a pan-European study," *Comput. Hum. Behav.*, vol. 75, pp. 811–825, Oct. 2017, doi: 10.1016/J.CHB.2017.06.007.

[14] A. Shifa, M. N. Asghar, and M. Fleury, "Multimedia security perspectives in IoT," in *Proc. 6th Int. Conf. Innov. Comput. Technol. (INTECH)*, Aug. 2016, pp. 550–555, doi: 10.1109/INTECH.2016.7845081.

[15] Cisco. (2019). *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper–Cisco*. Accessed: Aug. 5, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[16] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-Khah, and P. Siano, "Iot-based smart cities: A survey," in *Proc. IEEE 16th Int. Conf. Environ. Electr. Eng. (EEEIC)*, Jun. 2016, pp. 1–6, doi: 10.1109/EEEIC.2016.7555867.

[17] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015, doi: 10.1109/COMST.2015.2444095.

[18] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of cloud computing and Internet of Things: A survey," *Future Gener. Comput. Syst.*, vol. 56, pp. 684–700, Mar. 2016, doi: 10.1016/j.future.2015.09.021.

[19] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of IoT and cloud computing," *Future Gener. Comput. Syst.*, vol. 78, pp. 964–975, Jan. 2018, doi: 10.1016/j.future.2016.11.031.

[20] P. Hernandez-Leal, H. J. Escalante, and L. E. Sucar, "Towards a generic ontology for video surveillance," in *Applications for Future Internet* Cham, Switzerland: Springer, 2017, pp. 3–7.

[21] M. Y. Kazi Tani, A. Ghomari, A. Lablack, and I. M. Bilasco, "OVIS: Ontology video surveillance indexing and retrieval system," *Int. J. Multimedia Inf. Retr.*, vol. 6, no. 4, pp. 295–316, Dec. 2017, doi: 10.1007/s13735-017-0133-z.

[22] L. Calavia, C. Baladrón, J. M. Aguiar, B. Carro, and A. Sánchez-Esguevillas, "A semantic autonomous video surveillance system for dense camera networks in smart cities," *Sensors*, vol. 12, no. 8, pp. 10407–10429, Aug. 2012, doi: 10.3390/s120810407.

[23] N. Pahal, A. Mallik, and S. Chaudhury, "An ontology-based context-aware IoT framework for smart surveillance," in *Proc. 3rd Int. Conf. Smart City Appl. (SCA)*, 2018, pp. 1–7, doi: 10.1145/3286606.3286846.

[24] S. Martínez, D. Sánchez, and A. Valls, "Semantic adaptive microaggregation of categorical microdata," *Comput. Secur.*, vol. 31, no. 5, pp. 653–672, Jul. 2012, doi: 10.1016/j.cose.2012.04.003.

[25] A. Herzog, N. Shahmehri, and C. Duma, "An ontology of information security," *Int. J. Inf. Secur. Privacy*, vol. 1, no. 4, pp. 1–23, Oct. 2007, doi: 10.4018/jisp.2007100101.

[26] R. Luh, S. Marschalek, M. Kaiser, H. Janicke, and S. Schrittwieser, "Semantics-aware detection of targeted attacks: A survey," *J. Comput. Virol. Hacking Techn.*, vol. 13, no. 1, pp. 47–85, Feb. 2017, doi: 10.1007/s11416-016-0273-3.

[27] A. Razzaq, H. F. Ahmed, A. Hur, and N. Haider, "Ontology based application level intrusion detection system by using Bayesian filter," in *Proc. 2nd Int. Conf. Comput., Control Commun.*, Feb. 2009, pp. 1–6, doi: 10.1109/IC4.2009.4909223.

[28] H. Bannour and C. Hudelot, "Building and using fuzzy multimedia ontologies for semantic image annotation," *Multimedia Tools Appl.*, vol. 72, no. 3, pp. 2107–2141, Oct. 2014, doi: 10.1007/s11042-013-1491-z.

[29] D. T. Nguyen, W. Li, and P. O. Ogunbona, "Human detection from images and videos: A survey," *Pattern Recognit.*, vol. 51, pp. 148–175, Mar. 2016, doi: 10.1016/J.PATCOG.2015.08.027.

[30] S.-C. Huang, "An advanced motion detection algorithm with video quality analysis for video surveillance systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 1, pp. 1–14, Jan. 2011, doi: 10.1109/TCSVT.2010.2087812.

[31] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Comput. Sci. Rev.*, vol. 28, pp. 157–177, May 2018, doi: 10.1016/J.COSREV.2018.03.001.

[32] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proc. 4th IEEE Workshop Appl. Comput. Vis. (WACV)*, Oct. 1998, pp. 8–14, doi: 10.1109/ACV.1998.732851.

[33] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 43–60, Jan. 2016, doi: 10.1016/J.TRIT.2016.03.005.

[34] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, Feb. 1994, doi: 10.1007/BF01420984.

[35] A. C. Bovik, *The Essential Guide to Video Processing*. New York, NY, USA: Academic, 2009.

[36] C. Zhang and Z. Zhang. (Jun. ,1 2010). *A Survey of Recent Advances in Face Detection*. Accessed: Aug. 5, 2020. [Online]. Available: https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/

[37] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.

[38] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput. Vis. Image Understand.*, vol. 138, pp. 1–24, Sep. 2015, doi: 10.1016/J.CVIU.2015.03.015.

[39] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Advances in Biometrics*. Berlin, Germany: Springer, 2007, pp. 11–18.

[40] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[41] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 211–223, Feb. 2016, doi: 10.1109/TPAMI.2015.2448075.

[42] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2006, pp. 428–441.

[43] C. Conde, D. Moctezuma, I. Martín De Diego, and E. Cabello, "HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments," *Neurocomputing*, vol. 100, pp. 19–30, Jan. 2013, doi: 10.1016/J.NEUCOM.2011.12.037.

[44] F. Zhou and F. De la Torre, "Spatio-temporal matching for human detection in video," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2014, pp. 62–77.

[45] D. Cavaliere, V. Loia, and S. Senatore, "Towards an ontology design pattern for UAV video content analysis," *IEEE Access*, vol. 7, pp. 105342–105353, 2019, doi: 10.1109/ACCESS.2019.2932442.

[46] K. Goyal and J. Singhai, "Review of background subtraction methods using Gaussian mixture model for video surveillance systems," *Artif. Intell. Rev.*, vol. 50, no. 2, pp. 241–259, Aug. 2018, doi: 10.1007/s10462-017-9542-x.

[47] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, 2004, pp. 28–31, doi: 10.1109/ICPR.2004.1333992.

[48] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006, doi: 10.1016/J.PATREC.2005.11.005.

[49] R. P. Singh, P. Sharma, and J. Madarkar, "Compute-extensive background subtraction for efficient ghost suppression," *IEEE Access*, vol. 7, pp. 130180–130196, 2019, doi: 10.1109/ACCESS.2019.2937402.

[50] M. Tao, J. Zuo, Z. Liu, A. Castiglione, and F. Palmieri, "Multi-layer cloud architectural model and ontology-based security service framework for IoT-based smart homes," *Future Gener. Comput. Syst.*, vol. 78, pp. 1040–1051, Jan. 2018, doi: 10.1016/J.FUTURE.2016.11.011.

[51] A. Alti, A. Lakehal, S. Laborie, and P. Roose, "Autonomic semantic-based context-aware platform for mobile applications in pervasive environments," *Futur. Internet*, vol. 8, no. 4, pp. 1–26, 2016, doi: 10.3390/fi8040048.

[52] S. Ciftci, A. O. Akyuz, and T. Ebrahimi, "A reliable and reversible image privacy protection based on false colors," *IEEE Trans. Multimedia*, vol. 20, no. 1, pp. 68–81, Jan. 2018, doi: 10.1109/TMM.2017.2728479.

[53] W. L. Hoo, A. Miron, A. Badii, and C. S. Chan, "Skin-based privacy filter for surveillance systems," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Sep. 2015, pp. 269–272, doi: 10.1109/IWSSIP.2015.7314228.

[54] Y. Wang, M. O'Neill, F. Kurugollu, and E. O'Sullivan, "Privacy region protection for H. 264/AVC with enhanced scrambling effect and a low bitrate overhead," *Signal Process., Image Commun.*, vol. 35, pp. 71–84, Jul. 2015, doi: 10.1016/J.IMAGE.2015.04.013.

[55] X. Ma, L. T. Yang, Y. Xiang, W. K. Zeng, D. Zou, and H. Jin, "Fully reversible privacy region protection for cloud video surveillance," *IEEE Trans. Cloud Comput.*, vol. 5, no. 3, pp. 510–522, Jul. 2017, doi: 10.1109/TCC.2015.2469651.

[56] P. Carrillo, H. Kalva, and S. Magliveras, "Compression independent reversible encryption for privacy in video surveillance," *EURASIP J. Inf. Secur.*, vol. 2009, no. 1, pp. 1–13, 2009, doi: 10.1155/2009/429581.

[57] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, A. Javed, and A. Ahmadinia, "An intelligent real-time occupancy monitoring system with enhanced encryption and privacy," in *Proc. IEEE 17th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI CC)*, Jul. 2018, pp. 524–529, doi: 10.1109/ICCI-CC.2018.8482047.

[58] N. Cao, S. B. Nasir, S. Sen, and A. Raychowdhury, "Self-optimizing IoT wireless video sensor node with *in-situ* data analytics and context-driven energy-aware real-time adaptation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 9, pp. 2470–2480, Sep. 2017, doi: 10.1109/TCSI.2017.2716358.

[59] T. Sultana and K. A. Wahid, "IoT-guard: Event-driven fog-based video surveillance system for real-time security management," *IEEE Access*, vol. 7, pp. 134881–134894, 2019, doi: 10.1109/ACCESS.2019.2941978.

[60] D. E. Bell and L. J. La Padula. (1976). *Secure Computer System: Unified Exposition and Multics Interpretation*. Accessed: Mar. 23, 2019. [Online]. Available: https://apps.dtic.mil/docs/citations/ADA023588.

[61] I. Butun, M. Erol-Kantarci, B. Kantarci, and H. Song, "Cloud-centric multi-level authentication as a service for secure public safety device networks," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 47–53, Apr. 2016, doi: 10.1109/MCOM.2016.7452265.

[62] K. Sehairi, F. Chouireb, and J. Meunier, "Comparative study of motion detection methods for video surveillance systems," *J. Electron. Imag.*, vol. 26, no. 2, Apr. 2017, Art. no. 023025, doi: 10.1117/1.jei.26.2.023025.

[63] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," in *Proc. Semantic Web Work. Symp.*, 2001. Accessed: Aug. 5, 2020. [Online]. Available: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html

[64] M. O'Connor and A. Das, "SQWRL," in *Proc. 6th Int. Conf. OWL, Experiences Directions (OWLED)*, 2009, pp. 208–215. Accessed: Aug. 5, 2020. [Online]. Available: https://dl.acm.org/citation.cfm?id=2890072

[65] Z. Shahid, M. Chaumont, and W. Puech, "Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for i and p frames," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 565–576, May 2011, doi: 10.1109/TCSVT.2011.2129090.

[66] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003, doi: 10.1109/TCSVT.2003.815165.

[67] Z. Shahid and W. Puech, "Visual protection of HEVC video by selective encryption of CABAC binstrings," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 24–36, Jan. 2014, doi: 10.1109/TMM.2013.2281029.

[68] M. J. Dworkin, "FIPS 197, advanced encryption standard (AES)," *Netw. Secur. Nat. Inst. Stand. Technol.*, vol. 197, no. 12, p. 6028, 2001, doi: 10.6028/NIST.FIPS.197.

[69] M. N. Asghar and M. Ghanbari, "An efficient security system for CABAC bin-strings of H.264/SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 3, pp. 425–437, Mar. 2013, doi: 10.1109/TCSVT.2012.2204941.

[70] M. N. Asghar, M. Ghanbari, M. Fleury, and M. J. Reed, "Sufficient encryption based on entropy coding syntax elements of H.264/SVC," *Multimedia Tools Appl.*, vol. 74, no. 23, pp. 10215–10241, Dec. 2015, doi: 10.1007/s11042-014-2160-6.

[71] C. Saifurrab, S. Mirza, and M. Tech, "AES algorithm using advance key implementation in MATLAB," *Int. Res. J. Eng. Technol.*, vol. 3, no. 4, pp. 846–850, 2016.

[72] J. S. Khan and J. Ahmad, "Chaos based efficient selective image encryption," *Multidimens. Syst. Signal Process.*, vol. 30, pp. 943–961, May 2018, doi: 10.1007/s11045-018-0589-x.

[73] R. Hamza, A. Hassan, T. Huang, L. Ke, and H. Yan, "An efficient cryptosystem for video surveillance in the Internet of Things environment," *Complexity*, vol. 2019, pp. 1–11, Dec. 2019, doi: 10.1155/2019/1625678.

[74] S. M. Bellovin and R. Housley. (2005). *Guidelines for Cryptographic Key Management*. Accessed: Aug. 5, 2020. [Online]. Available: https://tools.ietf.org/html/rfc4107

[75] C. Paar and J. Pelzl, "The advanced enryption standard (AES)," in *Understanding Cryptography: A Textbook for Students and Practitioners*. Berlin, Germany: Springer-Verlag, 2010, pp. 87–117.

[76] X. Ma, W. K. Zeng, L. T. Yang, D. Zou, and H. Jin, "Lossless ROI privacy protection of H.264/AVC compressed surveillance videos," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 3, pp. 349–362, Jul. 2016, doi: 10.1109/TETC.2015.2460462.
[77] J.-P. Jodoin, G.-A. Bilodeau, and N. Saunier, "Urban tracker: Multiple object tracking in urban mixed traffic," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2014, pp. 885–892, doi: 10.1109/WACV.2014.6836010.
[78] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," in *Proc. 12th IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, Dec. 2009, pp. 1–6, doi: 10.1109/PETS-WINTER.2009.5399556.
[79] (2017). *MOT17: Multiple Object Tracking Benchmark*. Accessed: Mar. 23, 2019. [Online]. Available: https://motchallenge.net/data/MOT17/
[80] (2015). *ABODA - Abandoned Objects Dataset*. Accessed: Mar. 23, 2019. [Online]. Available: http://imp.iis.sinica.edu.tw/ABODA/index.html
[81] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*. London, U.K.: Institution Electrical Engineers, 2003.
[82] A. Shifa, M. Asghar, S. Noor, N. Gohar, and M. Fleury, "Lightweight cipher for H.264 videos in the Internet of multimedia things with encryption space ratio diagnostics," *Sensors*, vol. 19, no. 5, p. 1228, Mar. 2019, doi: 10.3390/s19051228.
[83] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[84] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004, doi: 10.1016/S0923-5965(03)00076-6.
[85] R. Kasturi, D. Goldgof, and P. Soundararajan. (2006). *Performance Evaluation Protocol for Face, Person and Vehicle Detection & Tracking, in Video Analysis and Content Extraction*. Accessed: Mar. 23, 2019. [Online]. Available: http://www.academia.edu/download/30794727/ClearEval_Protocol_v5.pdf
[86] M. Long, F. Peng, and H.-Y. Li, "Separable reversible data hiding and encryption for HEVC video," *J. Real-Time Image Process.*, vol. 14, no. 1, pp. 171–182, Jan. 2018, doi: 10.1007/s11554-017-0727-y.
[87] M. N. Asghar, R. Kousar, H. Majid, and M. Fleury, "Transparent encryption with scalable video communication: Lower-latency, CABAC-based schemes," *J. Vis. Commun. Image Represent.*, vol. 45, pp. 122–136, May 2017, doi: 10.1016/j.jvcir.2017.02.017.
[88] B. Boyadjis, C. Bergeron, B. Pesquet-Popescu, and F. Dufaux, "Extended selective encryption of H.264/AVC (CABAC)- and HEVC-encoded video streams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 892–906, Apr. 2017, doi: 10.1109/TCSVT.2015.2511879.
[89] X. Zhang, S.-H. Seo, and C. Wang, "A lightweight encryption method for privacy protection in surveillance videos," *IEEE Access*, vol. 6, pp. 18074–18087, 2018, doi: 10.1109/ACCESS.2018.2820724.

AMNA SHIFA received the M.S. degree in software engineering from Bahria University Islamabad and the master's degree in computer science from the Quaid-I Azam University Islamabad, Pakistan, and the Ph.D. degree in computer science from the Islamia University of Bahawalpur (IUB) Punjab, Pakistan, in 2019. She is currently working as a Lecturer with the Department of Computer Science and Information Technology (DCS&IT), The Islamia University of Bahawalpur. She is also an active Research Member with the Multimedia Research Group, DCS & IT, IUB. Her research interests include efficient video/image processing and compression, encryption, steganography, privacy, surveillance applications, and secure communication of multimedia data (images, audio, and video) over future networks. She has published ISI indexed journal articles along with International Conference papers.

MAMOONA N. ASGHAR received the Ph.D. degree with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K., in 2013. She has been a Marie Sklodowska-Curie (MSC) Career-Fit Research Fellow with the Software Research Institute, Athlone Institute of Technology (AIT), Ireland, since June 2018. As an MSC Principal Investigator (PI), her research targets the proposals and implementation of technological solutions for General Data Protection Regulation (GDPR) compliant CCTV Surveillance systems. She is currently a regular Faculty Member with the Department of Computer Science and Information Technology (DCS&IT), The Islamia University of Bahawalpur, Punjab, Pakistan, where she is also on Postdoctoral leave. She has more than 15 years of teaching and Research and Development experience. She has published several ISI indexed journal articles along with numerous International conference papers. She is also actively involved in reviewing for renowned journals and conferences. Her research interests include security aspects of multimedia (image, audio, and video), compression, visual privacy, encryption, steganography, secure transmission in future networks, video quality metrics, and key management schemes.

MARTIN FLEURY (Member, IEEE) received the degree in modern history from Oxford University, U.K., the degree in maths/physics from The Open University, Milton Keynes, U.K., the M.Sc. degree in astrophysics from the QMW College, University of London, U.K., in 1990, the M.Sc. degree in parallel computing systems from the University of South-West England, Bristol, in 1991, and the Ph.D. degree in parallel image-processing systems from the University of Essex, Colchester, U.K. He was a Senior Lecturer with the University of Essex, after which he became a Visiting Fellow. He is currently associated with the School of Engineering, Arts, Science, Technology, and Engineering (EAST), University of Suffolk, Ipswich, U.K. He is also a Free-Lance Consultant. He has authored or coauthored around 296 articles and book chapters on topics, such as document and image compression algorithms, performance prediction of parallel systems, software engineering, reconfigurable hardware, and vision systems. He has also published or edited books on high performance computing for image processing and peer-to-peer streaming. His current research interests include video communication over wireless networks and multimedia network security.

NADIA KANWAL (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Essex, Essex, U.K., in 2009 and 2013, respectively. She is currently a Marie Sklodowska-Curie Career-Fit Postdoctoral Fellow with the Software Research Institute, Athlone Institute of Technology (AIT), Ireland. The primary objective of this fellowship is to propose technological solutions for privacy protection of humans as per GDPR guidelines. She is also associated with the Lahore College for Women University, Pakistan, where she is also an Associate Professor and is also on leave to pursue Postdoctoral Fellowship. She is applying deep learning methods to improve the performance of vision algorithms for detection and matching tasks which can help to develop robust solutions for different vision-related applications. Her research interests include machine learning, image/video processing, medical imaging, and privacy. She remained a Student Member of the IEEE Computer Society, the Institution of Engineering and Technology, and the British Machine Vision Association. She has been actively involved in reviewing for reputed conferences and journals.

MOHAMMAD S. ANSARI (Senior Member, IEEE) received the B.Tech., M.Tech., and Ph.D. degrees in electronics engineering from Aligarh Muslim University, Aligarh, India, in 2001, 2007, and 2012, respectively. He is currently working as a Postdoctoral Research Fellow with the Software Research Institute, Athlone Institute of Technology, Ireland, while being on leave from the position of an Assistant Professor with the Department of Electronics Engineering, Aligarh Muslim University (AMU), Aligarh. Prior to joining AMU as a Lecturer, he has been associated with Siemens, Defence Research and Development Organization (DRDO), and with the Malaviya National Institute of Technology, Jaipur. His research interests include neural networks, machine learning, and analog signal processing. He has published around 100 research papers in reputed international journals and conferences and authored two books and contributed for book chapters. He was a recipient of the prestigious Young Faculty Research Fellowship by Department of Electronics and IT, Ministry of Communications and Information Technology, Government of India.

BRIAN LEE received the Ph.D. degree in application of programmable networking for network management from the Trinity College Dublin, Dublin, Ireland. He has more than 25 years Research and Development experience in telecommunications network monitoring, their systems and software design and development for large telecommunications products with very high-impact research publications. Formerly, he was the Director of research for LM Ericsson, Ireland, with responsibility for overseeing all research activities, including external collaborations and relationship management. He was an Engineering Manager with Duolog Ltd., where he was responsible for strategic and operational management of all research and development activities. He is currently the Director of the Software Research Institute, Athlone Institute of Technology, Athlone, Ireland.

MARCO HERBST received the degree in mechanical engineering from the Trinity College Dublin, in1999. He has successfully lead several software development teams. As the CEO and the CTO of Jobs.ie and Evercam, Marco was responsible for leading a team of software developers to build Ireland's most popular recruitment website and CCTV systems. He worked on CCTV hardware, software, and large-scale camera deployments, for seven years. He is currently an Innovator in time-lapse and monitoring software for construction projects and urban CCTV systems, positioning his company Evercam Ltd., as a Market Leader in the sector.

YUANSONG QIAO (Member, IEEE) received the B.Sc. degree and the M.Sc. degree in solid mechanics from Beihang University, Beijing, China, in 1996 and 1999, respectively, and the Ph.D. degree in computer applied technology from the Institute of Software, Chinese Academy of Sciences (ISCAS), Beijing, China, in 2008. He is currently a Senior Research Fellow with the Software Research Institute (SRI), Athlone Institute of Technology (AIT), Ireland. His research interests include future Internet architecture, blockchain systems, the IoT systems, and edge intelligence and computing. He is a member of IEEE (Communications and Computer societies and Blockchain Community) and ACM (SIGCOMM and SIGMM).