Efficient and Privacy Preserving Video Transmission in 5G-Enabled IoT Surveillance Networks: Current Challenges and Future Directions

Efficient and Privacy Preserving Video Transmission in 5G-Enabled IoT Surveillance Networks: Current Challenges and Future Directions

Khan Muhammad[1, *], Member, IEEE, Tanveer Hussain[1], Student Member, IEEE, Joel J. P. C. Rodrigues[2, 3], Senior Member, IEEE, Paolo Bellavista[4], Senior Member, IEEE, Antônio Roberto L. de Macêdo[5], Victor Hugo C. de Albuquerque[5], Senior Member, IEEE
[1]Department of Software, Sejong University, Seoul, Republic of Korea
[2] Instituto de Telecomunicações, Covilhã, Portugal
[3] Federal University of Piauí, Teresina – PI, Brazil
[4] University of Bologna, 40136 Bologna, Italy
[5] University of Fortaleza, Fortaleza/CE, Brazil

Abstract

Vision sensors in Internet of Things (IoT)-connected smart cities play a vital role in the exponential growth of video data, thereby making its analysis and storage comparatively tough and challenging. Those sensors continuously generate data for 24 hours, which require huge storage resources, dedicated networks for sharing with data centers, and most importantly, it makes browsing, retrieval, and event searching a difficult and time-consuming job. Video Summarization (VS) is a promising direction of solution to these problems, which analyzes the visual contents acquired from a vision sensor and prioritize them based on events, saliency, person's appearance, etc. However, the current VS literature still lacks focusing on resource-constrained devices that can summarize data over the edge and upload it to data repositories efficiently for instant analysis. Therefore, in this paper, we carry out a survey of VS methods functional to understand their pros and cons for resource-constrained devices, with the ambition to provide a compact tutorial to the community of researchers in the field. Further, we present a novel saliency-aware VS framework, incorporating 5G-enabled IoT devices, which keeps only important data thereby saving storage resources and providing representative data for immediate exploration. Keeping privacy of data on second priority, we intelligently encrypt the salient frames over resource-constrained devices before transmission over the 5G network. The reported experimental results show that our proposed framework has additional benefits of faster transmission (1.8~13.77% frames of a lengthy video are considered for transmission), reduced bandwidth, and real-time processing compared to state-of-the-art methods in the field.

Keywords

Anomaly Detection, Energy-Efficient Devices, Internet of Things, Intelligent Surveillance Networks, Video Summarization, Machine Learning, 5G, and Security

I. Introduction

The major sources for exponential growth of data in smart cities are various sensors connected in an IoT setup to undertake different missions such as water, transport, energy management, and many other applications [1]. These sensors generate different dimensions' data with varied nature of processing complexity. Pictorial data generated from vision sensors are considered as one of the most complex types of data for various applications [2], compared to other sensors such as smart meters and weather forecast sensors. Despite the practice of vision sensors in smart cities for surveillance and many other applications, it is a fact that video data pose various challenges such as computational complexity, storage, and transmission for meaningful usage. Stowing 24-hours video data from vision sensors create huge repositories in few days and the size breeds exponentially if the captured data are kept for weeks or months. Similarly, browsing, searching, and managing such Big Data are definitely challenging and time consuming. In addition, most recorded video data in smart cities are not very significant, whose transmission over network and storage on servers results in resource waste. Furthermore, huge video repositories make it tough for data analysts and end-users in smart cities to search for desired contents such as particular events or people in different applications, such as anomaly detection, face/facial expression recognition, and action/activity recognition. Thus, automatic techniques are required to prudently squeeze and filter the video data by keeping salient information and suppressing redundant contents. These are known as Video Summarization (VS) techniques, which are able to transform lengthy videos in a short and representative form (keyframes or video skims), with condensed contents that are trustworthy to be kept for further analysis, playing a key role in several applications of healthcare, entertainment, security and monitoring, tracking, and intelligent transportation domains.

VS methods are divided into two categories based on the number of vision sensors used to generate and summarize video data. Single vision sensor data summarized by VS techniques are called Single-view Video Summarization (SVS), while distributed cameras providing a larger coverage

generate Multi-view Video Summary (MVS). A single vision sensor provides inadequate coverage and is not suitable enough for consideration in smart cities. Distributed cameras in contrast consider different viewpoints to generate visual data, but their exploitation is particularly challenging for various purposes, such as summarization due to correlation among different views, variations of angles and light, and probably lack of synchronization. The aforementioned challenges and useful applications of VS in smart cities instigate the need of SVS and MVS that can prominently contribute to the intelligence and greenery of the current smart cities. VS methods can assist in useful data preservation generated from single and multi-viewed vision sensors connected in an IoT setup. In addition, it makes searching for data of interest easier and ensures real-time processing in an energy-friendly manner. Another big issue in IoT-assisted smart cities is the transmission of video data over the said wireless network [3], which is well-recognized to suffer from limited security support, slow data rates, and huge traffic. Since our focus is smart cities with 5G network support for devices, thus we consider the 5G network for transmission, which has high data rates but still with some recognized privacy leakage concerns [4].

To acquire effective summarization of video data from vision sensors in smart cities and ensure its secure transmission over 5G network to data centers, we propose an efficient and secure VS framework, whose main contributions are summarized as follows:

- We survey video summarization methods that are specifically suitable to be run at resource-constrained devices in IoT or deployed over the edge. Further, we provide a technical review of these methods and their possible future directions of exploitation in 5G environments.
- We introduce a novel framework to extract salient frames from Big Data generated by vision sensors and to transmit them over 5G network in an efficient way to achieve the complete advantage of 5G swiftness. Our framework reduces the size of big video data at high proportion, making it a best fit for further analysis and distribution.
- In order to ensure secure transmission over 5G network, we encrypt the salient frames before transmission. Alongside privacy, encryption has a bonus of lower bandwidth, reduced transmission time as well as it ensures real-time processing in IoT-assisted smart cities.
- We validate our framework by experimenting on real-world surveillance videos acquired from YouTube and video summarization datasets. The results prove that our framework is a valuable and potentially impactful contribution.

The remainder of the paper is structured as follows. In the next section, we present our proposed framework for VS in IoT environments. Further, we provide a compact overview of video analytics, SVS, and MVS methods from the smart cities' perspective. Following this, we provide future research directions for VS and data transmission techniques over 5G networks for smart cities. The final section concludes this article by compactly presenting its key findings and limitations.

II. Edge-Intelligence based Privacy Preserving VS Framework

Smart cities consist of an enormous number of vision sensors generating data for analysis and effective future usage. A high-resolution vision sensor usually generates 30 frames per second having minimum size of 1~2 MB for each and a whole day video will reserve 2.7~4.9 TB of space. Therefore, the high volumes of data stored in datacenters need to be reduced exponentially to make its storage and analysis smoother and easily implementable to gain thorough advantages of vision sensors in smart cities. Furthermore, its transmission over 5G network to data centers or cloud servers for instant analysis and efficient storage needs to be private and secure to keep the information of citizens highly confidential. To ensure secrecy of information over broadband network and preserve only useful data, we present a novel edge-intelligence based VS framework.

Our proposed framework, depicted in **Figure 1**, analyzes the video content generated from vision sensor used for smart surveillance to create a compact and representative form of long day videos called a "summary". This framework filters video frames and transmit the salient ones to cloud servers and data centers in a secure way over the 5G broadband network. The resultant data in summarized form is available on servers for users, authorities, and data analysts for purposeful processing. There are three main working layers in our framework: (1) data acquisition, (2) data prioritization/saliency extraction and transmission, and (3) data analysis and distribution layer. The data acquisition layer consists of a network of distributed vision sensors. The vision sensor is a special one with an attached 5G-enabled resource-constrained device that is interconnected with other smart sensors and devices. Compared to a single vision sensor, distributed vision sensors have several advantages and some extra challenges of processing. The key advantage of a network with numerous vision sensors is the broader coverage of events from different angles, making it comparatively easy to analyze and extract contextual information from a scene. The major challenges of a network of cameras include cross-view overlapping, which creates huge

redundancy plus extra computational complexity for initial scrutiny of video data. Therefore, we use a network of vision sensors to provide better coverage and effective data for analysis. The resource-constrained device records live frames via the attached vision sensor and passes it to the next layer of processing and transmission.

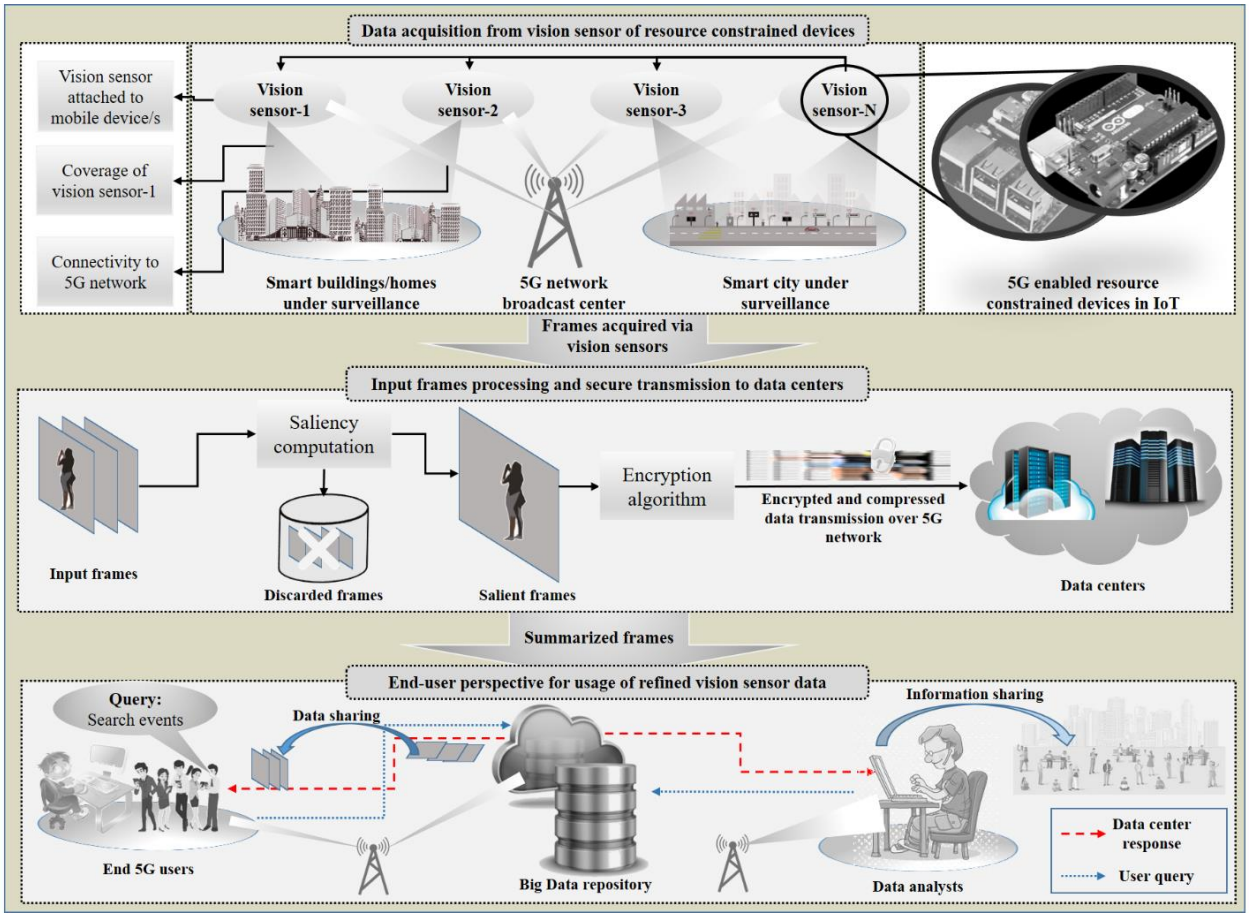

Figure 1: The proposed layered framework: first layer is data capturing, second one is saliency extraction and data transmission, and the final one is the data storage, distribution, and analysis layer.

The middle layer of our framework is central for saving resources and ensuring secrecy of data. As a first priority, we compute saliency of the live frames and check the level of information inside them to decide whether it deserves to be stored for future usage or not. We compute saliency of each input frame by extracting information and compare it with a certain threshold. The saliency computation algorithm is considered from our recent work on MVS, which is publicly available on Github. It outputs a real-valued output within the range of 0~4. The proposed framework is

flexible for salient frames threshold selection, where a higher value of threshold results in less but highly classified keyframes. In contrast, lower threshold outputs a reasonable number of keyframes and minimizes the chances of redundancy. The current optimal threshold used in our experiments is 0.22, which is decided after extensive analysis and experimentation. For more information on saliency computation, readers are referred to reference [5]. This step significantly suppresses the frames that are not enough informative to be kept for future and hence reduces the data size for transmission to data centers as proven from the experiments whose primary results are shown in **Figure 2**. Data reduction on a high proportion is clearly observable from **Figure 2**, where 1775 frames from video-1 are discarded by keeping only 32 keyframes with richer events and deserving information for further analysis. The number of discarded frames has a direct impact on size of the data being discarded. The salient data size in **Figure 2** seems to be significantly reduced when compared to the original data size. The data of video 1 is normalized to the range of 0~1800 from 0~26955 frames due to very large size in **Figure 2**.

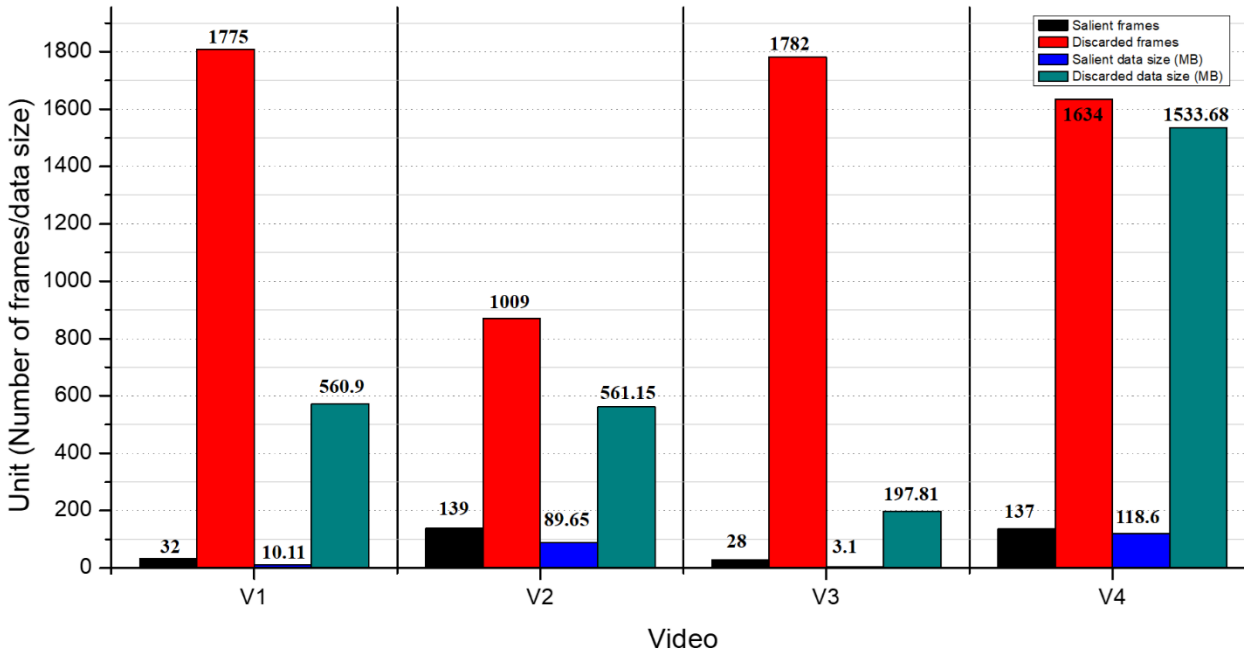

Figure 2: Statistical analysis of our proposed method for saving computation, storage, and network (5G) transmission resources. The network resources are comparatively reduced by our framework where we transmit summarized data over the 5G network.

Table 1: The implemented algorithms in our framework with their corresponding execution time over a resource-constrained device and a personal computer.

		Saliency computation	**Encryption**
Pseudo Code		1. Acquire HSV from RGB frame 2. Separate H, S, and V 3. Quantize H into 8 and S into 3 histogram bins (H and V are same) 4. Compute occurrence of each bin (h_x and s_x) in H and V 5. Compute P_x via h_x/size of frame 6. Compute P_{x1} through s_x/ size of frame 7. Compute log of P_x and P_{x1} and add with H and S, respectively 8. Add the resultants and normalize to range [0,1] 9. **Return**: the normalized value as final output	**At source:** 1. Acquire input RGB frame as I 2. Generate 16-byte random key as K 3. Initialize Cipher class for I and K 4. Acquire ET via Encrypt (I) 5. Transmit key and encrypted text (K, ET) **At destination:** 6. Receive (K, Encrypted text) 7. Initialize Cipher class for ET and K 8. Attain image I via Decrypt (ET) 9. Check the status of I via Verify (I) 10. **If** Status is True: Get Decrypted RGB frame 11. **Else**: **Return**: incorrect key or corrupted message
Source		https://github.com/tanveer-hussain/Embedded-Vision-for-MVS/blob/master/entropy.py	https://pycryptodome.readthedocs.io/en/latest/src/cipher/aes.html
Execution time (seconds)	Core i5-4670 CPU@ 3.40 Ghz	0.03	0.02
	Raspberry Pi-ARM Cortex A53	0.32	0.21

It is evident from several studies that 5G networks are still threats prone and there is a chance of leakage of information, therefore, preventive steps are needed to eliminate or reduce the risk of data outflow. To address this challenge, we utilize an encryption and decryption technique for secure transfer of salient frames. As the overall processing is executed over resource-constrained device, so we utilize a simple yet secure algorithm. We encrypt the salient frames in real-time through an encryption key that is randomly generated after a specific number of frames

transmission. After encryption, we transmit the scrambled file over 5G. When the file reaches destination, the same random key is used to decrypt it. We have employed a built-in cipher encryption technique in Python programming language. It is an advanced encryption standard (AES) based symmetric block cipher scheme with fixed data block size of 16 bytes with 128, 192, and 256 bits long key size. The first generated key is sent along with frame to the destination, which works fine up to a user-defined interval. After the interval, a new random key is generated at source location and transmitted alongside frames to the destination. The time complexity analysis and Pseudo Code of our adopted implementation of saliency computation and encryption/decryption techniques are given in **Table 1**.

Our framework is implemented in Python 3.5. The experimental results over MVS dataset and YouTube surveillance videos are reported in **Table 2**, where a huge gap among the overall data storage, and summarized frames can be observed.

Table 2: Data reduction rates for storage in data centers provided by our solution.

Video	**# total frames/ size (MB)**	**# salient frames/ size (MB)**	**Discarded frames without saliency**	**Remarks and salient frames extraction percentage**
1	1810/571.96	32/10.11	1775	This is office multi-view dataset video (view-0) lasting 14 minutes and 58 secs but the frames are normalized to 1810. Since this video has huge amount of non-salient frames, so our framework only considers 1.8% of total frames.
2	1009/650.8	139/89.65	870	It is surveillance video from YouTube. It saves salient data with 13.77% from overall frames.
3	1810/200.91	28/3.1	1782	It is also surveillance video. This video has very less amount of information, so our framework only saves 1.54% of frames from the whole video.
4	1771/1533.68	137/118.6	1634	Industrial surveillance video, which saves 7.73% frames

				only and discards others due to non-saliency.

III. Video Analytics in Smart Cities: Current Achievements and Challenges

In this section, we describe various video analytics methods from the perspective of smart cities, adopting approaches that exploit from statistical features to deep learning-based strategies, as shown in **Table 3**. We cover both single and multi-view literature and highlight the limitations of existing methods in these categories, with special focus on suitability for resource-constrained IoT environments. As a general consideration, the existing VS literature can be classified into two major categories: (1) employing low and mid-level features such as shape, motion, color, visual attention schemes; and (2) using high-level features such as deep learning-based techniques and learned features in prerequisite steps for summary generation. An example of the first category is presented by Meng et al. [6] to select representative object proposals generated from visual frames to summarize a video into fewer frames with only salient objects. This research is based on statistical features and the experiments are performed using movies data without any focus on surveillance, making it limited to many real-world scenarios and is not representative enough to be made general to all domains. Deep features-based shot segmentation followed by keyframes selection mechanism is given in [7], which falls in the high-level features category for VS. This research involves the usage of two deep learning models, making it computationally expensive and hard to be employed in smart city surveillance.

The literature about MVS is comparatively scarce due to the extra challenges posed by multi-view video data. For instance, Fu et al. [8] used the concept of spatio-temporal graphs to generate shots of input videos, followed by Gaussian entropy fusion scheme to give importance score to these shots and generate a multi-view representative summary. Learned features integrated with bi-directional LSTM are used to generate multi-view video skims in our recent work for MVS. In another follow-up work, we utilized lightweight CNNs to transmit reduced sized data generated from vision sensors to a master resource limited device in industrial IoT environments for final MVS. Considering the limitation of wireless networks, we encoded the transmitted frames to reduce the frame size during transmission.

Till date, the methods presented for SVS and MVS have still several drawbacks that need to be covered, particularly while dealing with Big Data from distributed vision sensors in IoT-assisted

smart cities. For instance, the majority of existing VS techniques are based on low or mid-level features, which are difficult to be generalized for smart city surveillance and are limited to only the tested scenarios. Almost all the VS techniques do not pay attention to the transmission of summarized frames, waste bandwidth, and create huge traffic and congestion on wireless networks. Indeed, the main problem of existing methods is their suitability and efficient applicability to smart cities because state-of-the-art VS techniques do not usually provide edge-based mechanisms running over resource-constrained devices that can be used over variable locations for better coverage [9]. Similarly, the targeted data volumes are huge, thus making hard to store full data on data portals for usage and analysis.

Table 3: Concise overview of the existing VS literature, with highlights about recent trends on adopted techniques.

Period	Trend	Applications orientation	Adaptability to 5G/IoT
2009-2011	• Most techniques are based on low and mid-level features • Saliency and motion are considered as prerequisite for VS • Statistical strategy-based object detection has special role • No proper shot segmentation mechanism used as prerequisite	• Some methods have diverse objectives of VS without any focus on surveillance, sports, etc. • Many methods are specific for road surveillance and general-purpose security applications	• Methods are not malleable due to the diverse nature of summarization and cannot be considered in smart city deployment environments with resource-constrained devices
2012-2013	• Shot segmentation is found in rare methods (spatial and temporal decomposition) • Summarization through clustering, events, and saliency-based track grouping • Person's detection and multiple moving object tracking based on feature matching and blobs detection • Event classification into different sub-types to assist further steps of summarization • Ranking mechanisms for keyframe selection	• Law enforcement • School security • Surveillance and road accident analysis	• The methods with person's detection, objects tracking, and event-based summarization can be extended to smart cities. However, the rest of them seems limited to certain domains

2014-2017	• Main focus on object tracking and motion-based VS • Some methods use salient motion detection based on principle of human cognition • Selection of keyframes using attention score, which is fusion of different metrics • Key object-based frames selection for final summary • Fuzzy rule-based mechanism for shots detection with events • Clustering techniques • Deep learning-based mechanisms in various stages of summarization	• Keyframe extraction for indexing and retrieval • Indoor/Outdoor surveillance	• The key objects-based keyframes and salient motion detection can be made applicable in IoT with cloud-assisted servers for huge processing. These methods do not usually provide a distributed implementation with parts that run at edge nodes in a smart city scenario
2018-2019	• Primary objective is to generate representative summaries by considering shots • Convolutional features and person's appearance-based shots segmentation strategies • Various supervised and semi-supervised learning mechanisms for VS • Cloud computing involvement and first rare contributions in edge intelligence-based VS	• Domain independent • Surveillance • Industrial sector where latency and reliability requirements are strict	• Certain methods in this trend are specifically proposed for smart cities and (industrial) IoT scenarios, while majority of them can be extended with 5G features, such as network slicing and edge-hosted execution, over 5G

To handle the limited processing capability issue in edge computing, Li et al. [10] introduced deep learning for IoTs in edge computing and posed a novel offloading strategy to optimize the execution performance of deep learning processes over edge. Similarly, M. S. Hossain et al., presented a cloud-assisted private video transmission framework for smart cities, where keyframes are extracted using genetic algorithms and a two-layer protection mechanism is utilized for secure transmission. A cost-effective VS in smart cities for IoT surveillance networks is presented by Muhammad et al, where keyframes are extracted after a hierarchical weighted fusion of different features i.e., aesthetics, memorability, and entropy. A multimedia data analytics framework is presented in [11], which uses traffic patterns-based intelligent models for data flow traffic classification. Muhammad et al., posed event-based keyframes extraction system from vision

sensor data in IoT environment, followed by lightweight encryption algorithm before transmission. Microsoft and NVIDIA are collaborating to cover video analytics on the edge [12] by altering raw and high-bandwidth consuming frames into lightweight transmission: they achieved good results in terms of real-time performance with reduced computational cost for end-users. The usage of multi-view videos in industrial IoT environments is studied in [5] by using statistical features, with a specific focus on the compression of frames for transmission over wireless network.

Intelligent VS techniques play a vital role in the greenery of smart cities. Despite the maturity of VS literature in surveillance, sports, and news, certain challenges still prevent these methods from generating summaries that could completely satisfy the requirements of green smart city users. We summarize these open challenges in **Figure 3** in a compact form with requirements and applications of computationally intelligent techniques for green future of smart cities. To handle these issues, we have presented a novel framework that is specifically designed and appropriate for smart cities IoT, as discussed in Section 2.

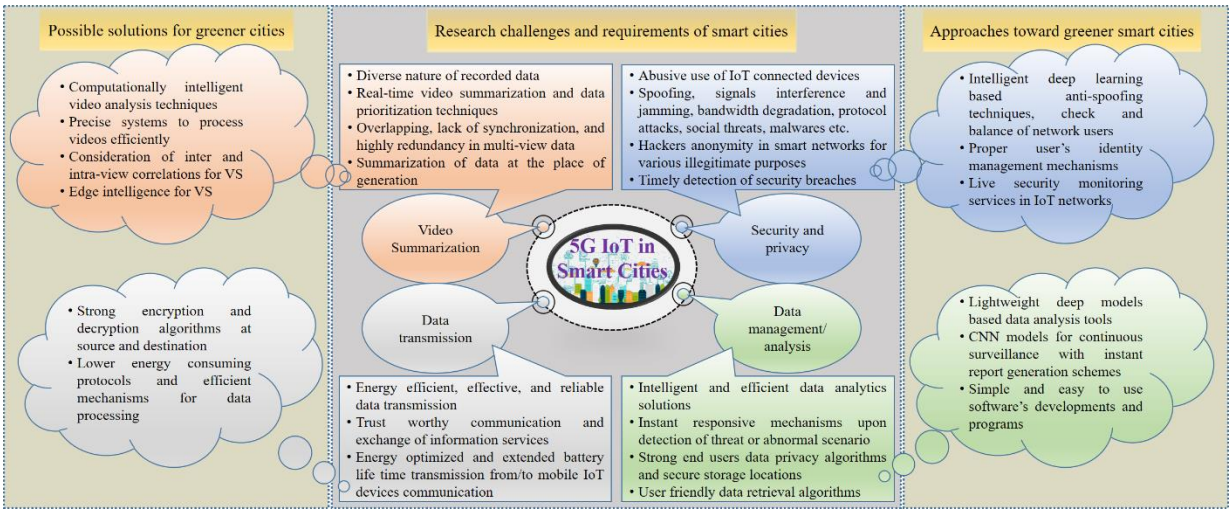

Figure 3: Main components of smart cities assisted by 5G and IoT, their major challenges and possible solutions for making them greener

IV. Future Research Directions

The major challenges that prevent the effective usage and analysis of big multimedia data generated in smart cities are given in **Figure 3** along with possible solutions. It clearly shows that extensive research is required in video analytics, in particular from the perspective of the adoption of these techniques in smart cities.

a) Edge Intelligence for Video Analytics

Edge intelligence remarks the processing of data at the place of its generation. Video analytics in general and particularly video summarization literature extremely needs research contributions that could analyze video frames at the edge. Resource-constrained devices can be integrated with vision sensors to process and prioritize video data, e.g., by discarding frames without an event or salient objects. This concept can also be applied to embedded devices to make the hardware capable of summarizing video data. These concepts are also applicable to mobile cameras and similarly adoptable for vehicle to vehicle (V2V) communication with edge devices in vehicular cloud computing to effectively save the prioritized traffic data for further analysis [13].

b) Data Secrecy during Transmission

One crucial feature while transmitting data over 5G broadband networks is the secrecy of information [14]. Since 5G network is prone to threats, where data can be exposed and accessed illegitimately by hackers, data privacy is a primary concern in the field, to be addressed by efficient techniques capable of being executed also at resource-constrained devices, with the support of edge nodes, and with the scalability requirements that are typical of large-scale smart city deployment environments. The current VS technologies still under-estimate the relevance of privacy during transmission, also because this research sub-area needs solid encryption algorithms suitable to be executed at resource-constrained and edge nodes in smart cities.

c) End-to-End Deep Learning Strategies

In state-of-the-art VS, majority of the recent techniques use deep learning or high-level features as preprocessing or in intermediate steps to generate a qualitative summary. To take benefit of the success of deep learning models in different computer vision domains, end-to-end neural networks for summary generation need to be investigated. Such networks should input small sequence of frames and pass it through various convolutional, pooling, etc. layers leading to only representative frames selection. End-to-end deep learning models are proving to be beneficial in smart cities due to the higher accuracy of neural networks for problems such as events detection and activity classification. Similar strategies, such as explainable and federated learning, are emerging in the computer vision domain and can be ported to video analytics for smart cities in the near future (see the survey by A. B. Arrieta et al [15]).

d) Benchmark Datasets

The publicly available data with ground truth for VS techniques is very limited, particularly while dealing with the problem of surveillance summary generation. As VS is subjective in nature, so datasets should be made available along with ground truth so that objective evaluation is easy and VS algorithms can be further matured. Furthermore, the available datasets are limited to only certain scenarios and do not contain very challenging cases such as smoky and foggy videos. Future solutions for smart city surveillance should not be limited to only certain types of regular deployment environments but should be adaptable thanks to flexible mechanisms that are able to deal with uncertain situations intelligently while performing the regular job of summary generation.

V. Conclusive Remarks

In this article, we offer a concise and easily accessible overview of VS techniques and their recent trends of evolution, by pointing out the major limitations of the existing solutions. We provided coverage of both single-view and multi-view video summarization state-of-the-art schemes for the last decade. We overviewed the strategies and steps followed by representative techniques in VS literature such as preprocessing, features extraction, and summary generation.

Considering the drawbacks of existing techniques, we proposed an energy-friendly framework for video summarization by incorporating saliency extraction and data encryption algorithms under an umbrella to intelligently filter out big video data generated from vision sensors. The presented framework has a high-level of adaptability for 5G-enabled smart city surveillance, with good capabilities to efficiently exploit the execution resources at resource-constrained devices and edge nodes. In addition, we have evaluated our proposed framework by conducting experiments on real-world surveillance data from YouTube and other multi-view datasets. Experimental results show that our framework is energy-efficient, flexible, and applicable to smart cities. In addition, following the challenges in VS literature and contributions of our framework, we highlighted the major challenges of this domain from different perspectives and envisioned future research directions for scientists for possible contributions to the greenery of smart cities, supported by IoT, 5G, and big data analytics.

Acknowledgments

The corresponding author is Khan Muhammad (khan.muhammad@ieee.org, khan@sejong.ac.kr). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2016R1A2B4011712); by National Funding from the FCT - Fundação para a Ciência e a Tecnologia, through the UID/EEA/50008/2019 Project; by the Government of Russian Federation, Grant 08-08; by RNP, with resources from MCTIC, Grant No. 01250.075413/2018-04, under the Centro de Referência em Radiocomunicações - CRR project of the Instituto Nacional de Telecomunicações (Inatel), Brazil; and by Brazilian National Council for Research and Development (CNPq) via Grant No. 309335/2017-5; by Brazilian National Council for Research and Development (CNPq) via grants no. 304315/2017-6 and 430274/2018-1.

References

[1] X. Li, S. Cheng, Z. Lv, H. Song, T. Jia, and N. Lu, "Data analytics of urban fabric metrics for smart cities," Future Generation Computer Systems, 2018/02/16/ 2018.

[2] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," IEEE Internet of Things Journal, vol. 6, pp. 1375-1383, 2018.

[3] M. Chen, Y. Miao, X. Jian, X. Wang, and I. Humar, "Cognitive-LPWAN: Towards intelligent wireless services in hybrid low power wide area networks," IEEE Transactions on Green Communications and Networking, vol. 3, pp. 409-417, 2018.

[4] M. A. Ferrag, L. Maglaras, A. Argyriou, D. Kosmanos, and H. Janicke, "Security for 4G and 5G cellular networks: A survey of existing authentication and privacy-preserving schemes," Journal of Network and Computer Applications, vol. 101, pp. 55-82, 2018.

[5] T. Hussain, K. Muhammad, J. D. Ser, S. W. Baik, and V. H. C. d. Albuquerque, "Intelligent Embedded Vision for Summarization of Multi-View Videos in IIoT," IEEE Transactions on Industrial Informatics, pp. 1-1, 2019.

[6] J. Meng, H. Wang, J. Yuan, and Y.-P. Tan, "From keyframes to key objects: Video summarization by representative object proposal selection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1039-1048.

[7] K. Muhammad, T. Hussain, and S. W. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices," Pattern Recognition Letters, 2018/08/07/ 2018.

[8] Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z.-H. Zhou, "Multi-view video summarization," IEEE Transactions on Multimedia, vol. 12, pp. 717-729, 2010.

[9] L. Zhao, J. Wang, J. Liu, and N. Kato, "Optimal Edge Resource Allocation in IoT-Based Smart Cities," IEEE Network, vol. 33, pp. 30-35, 2019.

[10] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," IEEE Network, vol. 32, pp. 96-101, 2018.

[11] A. Canovas, J. M. Jimenez, O. Romero, and J. Lloret, "Multimedia Data Flow Traffic Classification Using Intelligent Models Based on Traffic Patterns," IEEE Network, vol. 32, pp. 100-107, 2018.
[12] G. Ananthanarayanan, V. Bahl, L. Cox, A. Crown, S. Nogbahi, and Y. Shu, "Video Analytics-Killer App for Edge Computing," in Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, 2019, pp. 695-696.
[13] P. Gomes, C. Olaverri-Monreal, and M. Ferreira, "Making Vehicles Transparent Through V2V Video Streaming," IEEE Transactions on Intelligent Transportation Systems, vol. 13, pp. 930-938, 2012.
[14] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, "Overview of 5G security challenges and solutions," IEEE Communications Standards Magazine, vol. 2, pp. 36-43, 2018.
[15] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, et al., "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," arXiv preprint arXiv:1910.10045, 2019.

Khan Muhammad [S'16, M'18] is an assistant professor at Department of Software, Sejong University, South Korea. His research interests include video summarization, computer vision, big data analytics, IoT, 5G, and video surveillance. He has authored over 70 papers in peer-reviewed international journals, such as IEEE COMMAG, NETWORK, TII, TIE, IoJT, TNNLS and TSMC-Systems, and is a reviewer of over 40 SCI/SCIE journals, including IEEE COMMAG, WCOMM, NETWORK, IoTJ, TIP, TII, TCYB, Access and ACM TOMM. He is a member of the ACM.

Tanveer Hussain [S'19] received his Bachelor's degree in Computer Science from Islamia College Peshawar (with Gold Medal distinction), Peshawar, Pakistan in 2017. He is currently pursuing his M.S. leading to Ph.D. degree from Sejong University, Seoul, Republic of Korea and serving as Research Assistant at Intelligent Media Laboratory (IM Lab). His major research domains are features extraction (learned and low-level features), video analytics, image processing, pattern recognition, medical image analysis, multimedia data retrieval, deep learning for multimedia data understanding, single/multi-view video summarization, IoT, IIoT, and resource-constrained programming. He has published several journal articles in these areas in reputed journals including IEEE TII, IoTJ, Elsevier PRL, and Wiley IJDSN.

Joel J. P. C. Rodrigues [S'01, M'06, SM'06] is a professor at the National Institute of Telecommunications (Inatel), Brazil and senior researcher at Instituto de Telecomunicações, Portugal. He is the leader of the Internet of Things Research Group (CNPq), Director for

Conference Development (IEEE ComSoc Board of Governors), IEEE Distinguished Lecturer, and Past Chair of the IEEE ComSoc eHealth and Communications Software TCs. He is the Editor-in-Chief of the International Journal of E-Health and Medical Communications, and he has authored or coauthored over 700 publications.

Paolo Bellavista (SM'06) received the Ph.D. degree in computer science engineering from the University of Bologna, Italy, in 2001, where he is currently a Full Professor. His research interests include mobile agent-based middleware solutions, and pervasive wireless computing to location/context-aware services and management of cloud systems. He serves on the Editorial Board of the IEEE TNSM, IEEE TSC, Elsevier PMC, Springer WINET, and Springer JNSM.

Antônio Roberto L. de Macêdo is a Ph.D. student at Unifor, Fortaleza. He Graduated in Electronic Engineering in 2004 from the University of Fortaleza / UNIFOR. He received his specialization in Petroleum Engineering in 2014 from UNIFOR and graduated from the College of War in Defense Resource Management in the same year. He received Master in Engineering in 2016 from IFCE. His research interests include medical data analysis using computationally intelligent techniques. He has various publications in reputed journals including European journal of physical and rehabilitation medicine, clinical biomechanics.

Victor Hugo C. de Albuquerque [M'17, SM'19] completed Ph.D. in Mechanical Engineering from the Federal University of Paraiba (UFPB, 2010), M.Sc. in Teleinformatics Engineering from the Federal University of Ceará (UFC, 2007). He is currently Assistant VI Professor of the Graduate Program in Applied Informatics at the University of Fortaleza (UNIFOR). He has authored or co-authored over 160 papers in refereed international journals, conferences, four book chapters, and four patents. He is an Editorial Board Member and Lead Guest Editor of several high-reputed journals, and TPC for many international conferences.