

Wireless Video Surveillance: A Survey

YUN YE¹ (Student Member, IEEE), SONG CI¹ (Senior Member, IEEE),
AGGELOS K. KATSAGGELOS² (Fellow, IEEE), YANWEI LIU³, AND YI QIAN¹ (Senior Member, IEEE)

¹Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Omaha, NE 68182, USA

²Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA

³Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

Corresponding author: Y. Ye (yye@huskers.unl.edu)

This work was supported by the National Science Foundation under Grant 1145596.

ABSTRACT A wireless video surveillance system consists of three major components: 1) the video capture and preprocessing; 2) the video compression and transmission in wireless sensor networks; and 3) the video analysis at the receiving end. A myriad of research works have been dedicated to this field due to its increasing popularity in surveillance applications. This survey provides a comprehensive overview of existing state-of-the-art technologies developed for wireless video surveillance, based on the in-depth analysis of the requirements and challenges in current systems. Specifically, the physical network infrastructure for video transmission over wireless channel is analyzed. The representative technologies for video capture and preliminary vision tasks are summarized. For video compression and transmission over the wireless networks, the ultimate goal is to maximize the received video quality under the resource limitation. This is also the main focus of this survey. We classify different schemes into categories including unequal error protection, error resilience, scalable video coding, distributed video coding, and cross-layer control. Cross-layer control proves to be a desirable measure for system-level optimal resource allocation. At the receiver's end, the received video is further processed for higher-level vision tasks, and the security and privacy issues in surveillance applications are also discussed.

INDEX TERMS Video surveillance, wireless sensor networks, multimedia communications, cross-layer control, video analysis.

I. INTRODUCTION

Video Surveillance over wireless sensor networks (WSNs) has been widely adopted in various cyber-physical systems including traffic analysis, healthcare, public safety, wildlife tracking and environment/weather monitoring. The unwired node connection facility in WSNs comes with some typical problems for data transmission. Among them are line-of-sight obstruction, signal attenuation and interference, data security, and channel bandwidth or power constraint. A vast amount of research work has been presented to tackle these problems, and many have been successfully applied in practice and have become industrial standards. However, for video surveillance applications, especially those with real-time demands, the processing and transmission process at each wireless node for a large amount of video data is still challenging.

In current state-of-the-art wireless video surveillance systems, each source node is usually equipped with one or more cameras, a microprocessor, the storage unit, a transceiver, and a power supply. The basic functions of each node include video capture, video compression and data transmission.

The process of video analysis for different surveillance purposes is implemented either by the sender or by the receiver, depending on their computational capability. The remote control unit at the receiver's end can also provide some useful information feedback to the sender in order to enhance the system performance. The major functional modules of a video surveillance system are illustrated in Figure 1.

The existing WSN technologies are utilized in all kinds of wireless video surveillance applications. One popular application is traffic analysis. For example, the traffic signal system deployed by the transportation department in the city of Irving, Texas (Irving, 2004) [1] implemented seventy pan-tilt-zoom (PTZ) CCTV (closed-circuit television) cameras to cover about two hundred intersections. One smart camera capable of video codec and video over IP functions was installed at each traffic site together with a radio/antenna unit. The on-site signal is transmitted to the base stations ringed in a 100 Mbps wireless backbone operating at the licensed frequencies of 18-23 GHz. The traffic monitoring system at the University of Minnesota (UMN, 2005) [2], and the system

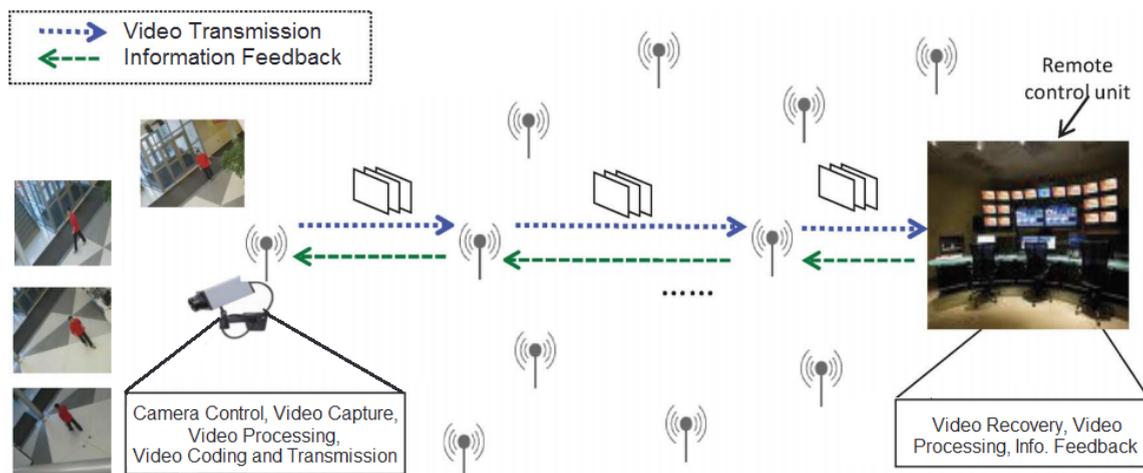


FIGURE 1. A wireless video surveillance system.

TABLE 1. Wireless video surveillance systems.

Surveillance system	Surveillance environment	MAC protocol	Carrier frequency	Maximum throughput	Network topology	Camera number	Camera control	Video delivery
Panoptes, 2003 [8]	indoor	802.11	2.4 GHz	12 Mbps	star	multiple	fixed	unicast
PRISMATICA, 2003 [7]	indoor/ outdoor	802.11	2.44 GHz	N. A.	star	multiple	fixed	unicast
Irving, 2004 [1]	highway	802.11a 802.16	5.8, 18, 24/23 GHz	20 - 60 Mbps for on-site, 100 Mbps for backbone	star, ring	multiple	PTZ	unicast
UMN, 2005 [2]	highway	802.16	5.4 GHz	3/30 Mbps	star	multiple	fixed	simulcast
FireWxNet, 2006 [4]	wildland	802.11	924Mhz	10 Mbps	tree	multiple	PTZ	unicast
MeshEye, 2007 [9]	indoor/ outdoor	802.15.4	2.4 GHz	11 Kbps	point-to-point	multiple	fixed	simulcast
Tsinghua, 2009 [10]	indoor	802.15.4	900 MHz	19.2 Kbps	star, tree, mesh	multiple	fixed	unicast
UML, 2010 [6]	indoor	802.11b/g	2.4 GHz	54 Mbps	tree	multiple	fixed	unicast
SCNS, 2011 [5]	railway	802.11g 802.11j	2.4, 5 GHz	20 Mbps	mesh	multiple	fixed, PTZ	unicast
UNT, 2011 [3]	campus	802.11 a/n	5.4, 5.8 GHz	20 Mbps	chain	multiple	PTZ	unicast

at the University of North Texas (UNT, 2011) [3] are among other examples of wireless traffic surveillance.

Video surveillance in other wireless communication applications is also intensively studied, such as the remote weather monitoring system (FireWxNet, 2006) initially developed for the fire fighting community in the Bitterroot National Forest in Idaho to monitor the lightning stricken forest fire [4], the smart camera network system (SCNS, 2011) used for security monitoring in a railway station [5], and the indoor surveillance system in a multi-floor department building at the University of Massachusetts-Lowell [6]. The common problems considered in these systems include the sensor deployment and the system configuration for video communications.

For surveillance in a wide social area like metropolis, the sensor deployment is more complex. An example is the multi-sensor distributed system developed at Kingston University, named proactive integrated systems for security management

by technological institutional and communication assistance (PRISMATICA, 2003) [7]. Both wired and wireless video and audio subsystems were integrated in the centralized network structure. The data processing module at the operation center supported multiple real-time intelligent services, such as overshadowing and congestion detection upon received video.

The power efficiency problem is another major concern for some wireless video surveillance applications. In the system (Panoptes, 2003) described in [8], a central node received data from other client nodes and performed video aggregation to detect unusual events. The energy saving strategy employed by the client node included data filtering, buffering, and adaptive message discarding. In the work presented in [9], the hybrid-resolution smart camera mote (MeshEye, 2007) was designed to perform stereo vision at the sensor node with low energy cost. The location of the targeted object was

first estimated from the image data by the two low resolution cameras. Then the high resolution camera marked the position in its image plane and transmitted only the video data inside the target region. The multiresolution strategy was also adopted in the multiview target surveillance system developed at Tsinghua University (Tsinghua, 2009) [10].

These surveillance systems are built upon the existing wireless video communication technologies, especially the WSN infrastructure and video codec. Compared to traditional systems adopting a wired connection, the advantage of network mobility greatly facilitates the system deployment and expansion. Some technical parameters of these systems are listed in Table 1.

While the well-established WSN infrastructure and video communication standards can be utilized in a surveillance system, many new technologies have been proposed to accommodate the special requirements of the surveillance applications, such as target object tracking, content-aware resource allocation, and delay or power constrained video coding and transmission. This paper presents a review of these proposals based on the analysis of the technical challenges in current systems, especially on the video delivery part in an unsteady wireless transmission environment, aiming to provide some beneficial insights for future development. The rest of the paper is organized as follows. Section II introduces the network infrastructure for a wireless video surveillance system, including the standard channel resource, and the network topology. Section III describes some examples of video capture and preliminary vision tasks that can be operated by the sensor node. Section IV summarizes a number of video coding and transmission techniques dedicated to unequal error protection, error resilience, and scalable and distributed data processing. The cross-layer control mechanism is introduced as an efficient way for optimal resource allocation. Section V briefly introduces several video analysis algorithms designed for wireless surveillance systems with single or multiple cameras. Section VI discusses the security and privacy issues, and conclusions are drawn in Section VII.

II. NETWORK INFRASTRUCTURE

Data transmission in a wireless video surveillance system is regulated by wireless communication standards. Before network deployment, comprehensive on-site investigation needs to be conducted to avoid signal interference and equipment incompatibility. This section discusses the channel resource and network topology for the configuration of a wireless video surveillance system. Detailed implementation of the sensor network deployment procedures can be found at [3].

A. CHANNEL RESOURCE

In the U.S.A., the Federal Communication Commission (FCC) is responsible for regulating radio spectrum usage [11]. The most commonly used license-exempt frequency bands in current wireless surveillance systems include 900MHz, 2.4GHz, and 5.8GHz. The 4.9GHz frequency band is reserved for Intelligent Transportation Systems (ITS)

for public safety and other municipal services [12]. The specific communication parameters are defined in several groups of standards including IEEE 802.11/WiFi, IEEE 802.16/WiMax, IEEE 802.15.4/ZigBee, etc. The properties of operation with these frequency bands are summarized in Table 2. The higher frequency band demonstrates better range and interference performance, with lower penetration capability.

TABLE 2. Common radio frequency bands for wireless surveillance.

Frequency band	License	Line of sight	Advantage	Disadvantage
900 MHz	Unlicensed	Not required	Penetrates trees, dry wall	Lower throughput for video compared to other bands
2.4 GHz	Unlicensed	Required	Better penetration than 5 GHz	Interference from consumer devices
4.9 GHz	Licensed	Required	Reserved for public safety; less interference	Requires frequency coordination with other agencies
5 GHz	Unlicensed	Required	Better range and less interference than 2.4 GHz	Lower penetration than 2.4 GHz

B. NETWORK TOPOLOGY

As in WSNs, the network topology in a wireless video surveillance system could be a one hop or relayed point-to-point connection for single view video transmission, or a chain, star, tree or mesh structure for multiview surveillance. The network topology is application dependent. The resource constraints, cost efficiency, as well as the terrain and ecological condition of the surveillance environment are among the factors considered in adopting a suitable topology.

In the campus traffic monitoring system (UMN, 2005) [2], the surveillance environment was relatively small-scale, and the runtime video delivery was the primary concern. Therefore the point-to-point communication was realized by simulcasting multiple synchronized video sequences to the remote base station for real-time observation, as displayed in Figure 2(a). For surveillance in a large public area, different types of sensors might need to be installed at multiple distant locations, and hence a centralized star structure is preferred, such as the PRISMATICA system (PRISMATICA, 2003) [7] illustrated in Figure 2(b). The centralized network connection results in high throughput at the center node, which has to meet stringent standard for both the performance and stability requirements.

When the energy conservation is the major consideration, the sensors need to be organized in a more efficient manner. The work presented in [10] tested the surveillance system under different WSN topologies and demonstrated that, when collaboration among different sensor nodes is required, a tree or mesh network could achieve higher system performance compared to a star structure, in terms of power efficiency and data throughput. Figure 2(c) shows the tree structure of SensEye [13], a multitier surveillance system with different data processing and power consumption patterns devised on

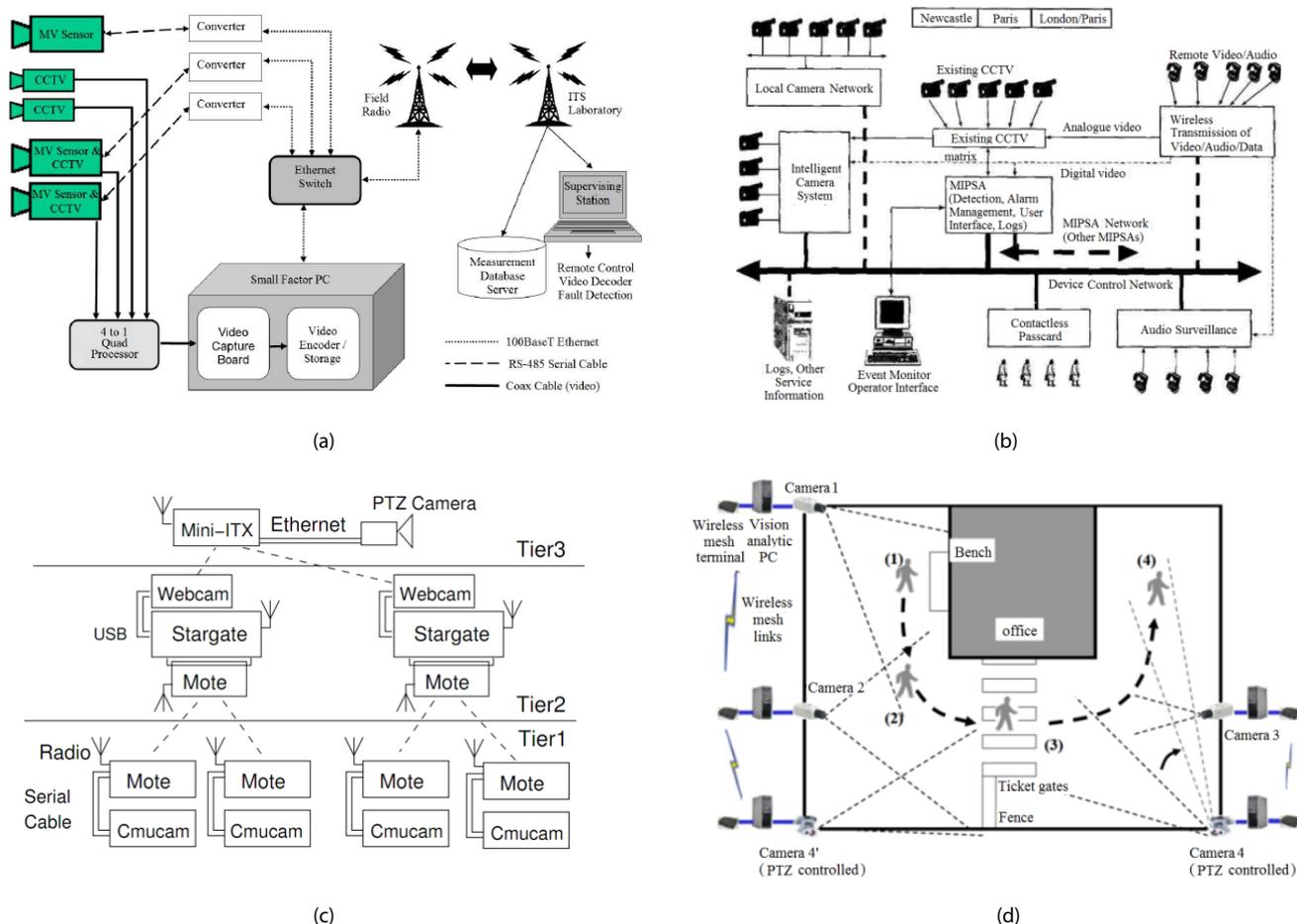


FIGURE 2. Network topology. (a) Point-to-point (UMN, 2005). (b) Star (PRISMATICA, 2003). (c) Tree (SenseEye, 2005). (d) Mesh (SCNS, 2011).

each level of the tree. The sensor nodes at the lower tiers consisting of low power devices worked at a longer duty cycle than the nodes at the higher tiers which consumed more power and executed more complex functions only upon receiving the signals from its child nodes at the lower tier.

If the functionality and computational capability are equally distributed among the sensor nodes, a mesh network is more appropriate. The mesh structure of the multiview object tracking system SCNS [5] using the Ad-hoc On-Demand Distance Vector (AODV) routing protocol is demonstrated in Figure 2(d). In this system, each node was able to communicate with others to determine the target position and to select the nearest camera for object tracking.

Another interesting issue in designing an efficient network topology is how to choose a proper amount of surveillance nodes for full-view coverage of the moving target. The camera barrier coverage in an existing network deployment was analyzed in [14], [15]. An optimal subset of the camera sensors is selected for video capture, while the distance between the camera and the target is sufficiently close, and the angle between the camera view direction and the target's face direction is within acceptable scope. The work presented in [16] studied the coverage problem with active cameras.

The camera's pan and zoom parameters were configured to support full-view coverage with a smaller number of selected nodes. The coverage schedule leads to better utilization of the network resources. In a wireless surveillance system, the camera selection procedure also needs to consider other critical issues including how to effectively identify the target location and to coordinate the distributed sensors over the air, under limited resources.

III. VIDEO CAPTURE AND PRELIMINARY VISION TASKS

The surveillance video is recorded by the sensor node at the monitor site for further data processing and transmission. Some preliminary vision tasks can be performed by a smart camera or the integrated processing unit at the sensor node.

For the surveillance systems using fixed cameras, object detection and localization are among the most popular functions performed at the sensor node. Object detection with a fixed camera often takes advantage of the static background. A commonly used technique is background subtraction. The background image can be obtained through periodically updating the captured data [9], [17], or through adaptive background modeling based on the Gaussian Mixture Model (GMM) learning process [18], [19]. This temporal

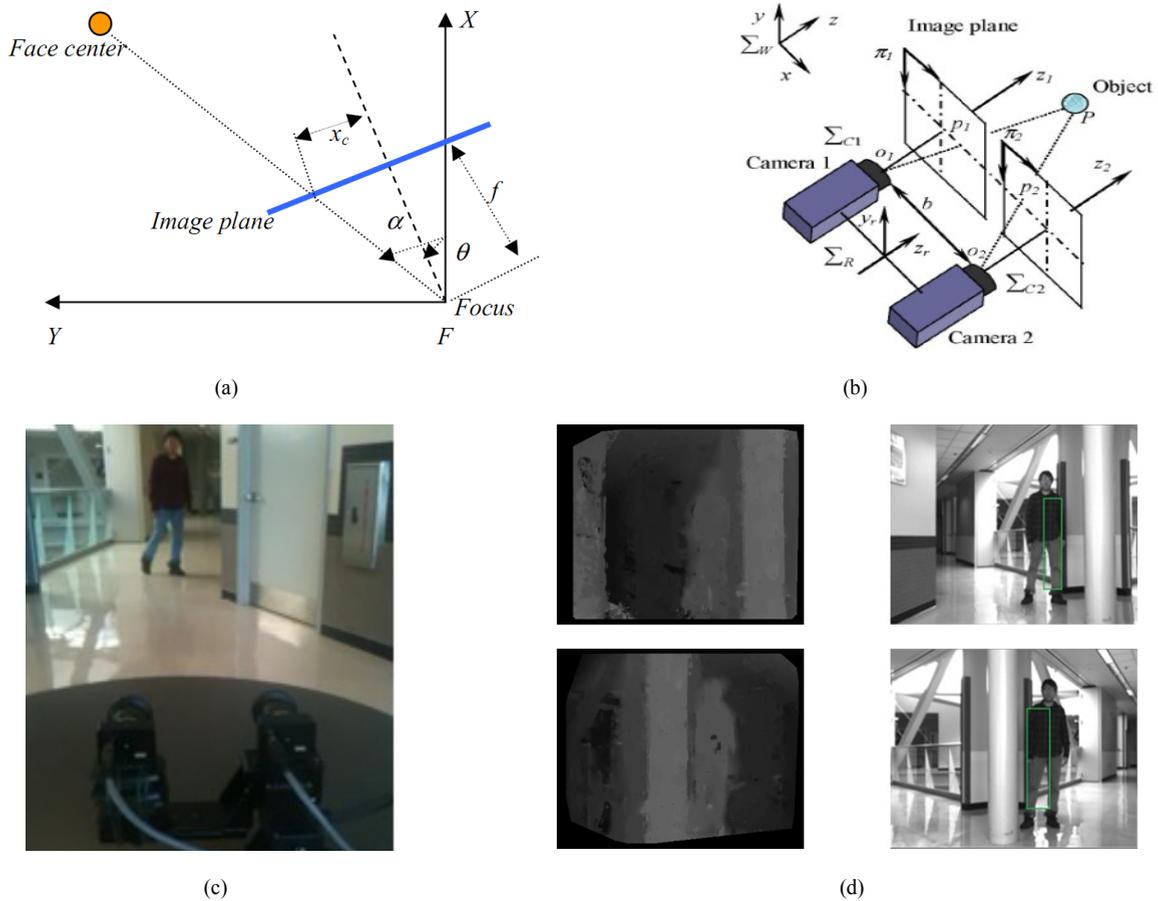


FIGURE 3. PTU camera for object tracking. (a) PTU camera control. (b) Binocular distance measure. (c) Binocular PTU cameras. (d) Disparity estimation and window size adjustment.

learning process models different conditions of a pixel at a certain position as a mixture of Gaussian distributions. The weight, mean, and variance values of each Gaussian model can be updated online, and pixels not conforming to any background model are quickly detected. The adaptive learning property makes this technique suitable for real-time applications, and a variety of detection methods are developed combining other spatiotemporal processing techniques [10], [20], [21]. With the object detection results provided by two or more cameras, the 3-D object position can be localized through vision analysis using the calibrated camera parameters and the object feature correlation [5], [9], [10], [17].

When the number of sensor nodes is restricted, a pan-tilt unit (PTU) or a PTZ camera provides more flexible view coverage than the stationary camera does. A PTU camera is capable of pan and tilt movements with a fixed focus position. The camera control can be manually performed by the remote receiver through information feedback [1], [3], [4], or automatically by the source node based on the vision analysis by the integrated processing unit [5], [22], [23]. The traffic surveillance system developed at the University of North Texas [3] had an Axis 213PTZ camera and a radio device installed at each of the three control center through

a daisy chain network. The operator at the control center was able to adjust the PTZ camera motion and the focal length, and to estimate the vehicle speed on a roadway parallel to the image plane.

The automatic camera control is closely related to the vision task performed by the processing unit. For example, object detection is often integrated with the camera control process. Figure 3(a) displays the PTU camera control algorithm described in [22] for object tracking. The focus O denotes the projection center. The image plane is viewed down along its y axis, and is projected onto the $X - Y$ world coordinate plane. α is the angle between the detected object center and the X axis, θ is the camera angle between the image center and the X axis, f is the focal length, and x_c the distance between the projected object center and the image center along the x axis of the image plane. Only the pan control algorithm is displayed in the figure. It applies to the tilt control similarly.

In the camera control process, the camera angle θ is updated at each time instance, aiming to minimize x_c and the difference between the estimated object speed and the actual object speed measured by a local tracker using the Mean Shift algorithm [24]. The exterior camera parameters

correlated to the pan and tilt movement (hand-eye calibration) were investigated in the binocular vision system introduced in [23]. In the tracking process, two PTU cameras were used to measure the distance campus sites. The traffic video was transmitted to the remote between the detected object and the image plane, as shown in Figure 3(b). The tracking region was scaled according to the estimated distance at the next detection process using the Continuously Adaptive Mean Shift Algorithm (CAMShift) algorithm [25]. To better obtain the distance information, in our binocular video surveillance system described in [26], the depth map for the entire frame is generated using a fast disparity estimation algorithm. Figure 3(c) and (d) demonstrate the binocular PTU cameras, and the tracking window adjustment using the generated depth information. The resulting 3D video data can be delivered to the receiver for further processing.

A PTU/PTZ camera is usually expensive and consumes much energy [13]. To reduce the cost and the technical complexity, some surveillance systems also used combined fixed and PTU/PTZ cameras for video capture and object tracking [5], [13], such as the systems illustrated in Figure 2 (c) and (d). Under some circumstances, special lenses can be adopted to further reduce the number of cameras. For example, the ultra wide-angle Fisheye and Panomorph lenses are used for panoramic or hemispherical viewing. The distorted images can be rectified using the camera parameters. The extra computation and communication resource consumption for processing the captured images can not be ignored in designing a wireless video surveillance system.

IV. VIDEO CODING AND TRANSMISSION

In a wireless video surveillance system, the captured video data are encoded and transmitted over the error prone wireless channel. Most of the current systems adopt a unicast or simulcast video delivery, as shown in Table 1. Each camera output is encoded independently using well-established image or video coding standards including JPEG, JPEG2000, motion JPEG (MJPEG), MPEG and H.26x. To better adapt to typical surveillance applications, a variety of techniques has been proposed for the video coding and transmission process in WSNs.

A. OBJECT BASED UNEQUAL ERROR PROTECTION

When the communication resources are limited, an alternative to heavier compression is to implement unequal error protection (UEP) for different parts of the video data. The idea of UEP is to allocate more resources to the parts of the video sequence that have a greater impact on video quality, while spending fewer resources on parts that are less significant [27]. In the surveillance video, the moving target object is of greater interest than the background. Hence the region of interest (ROI) based UEP mechanism is a natural way to optimize resource allocation.

An object based joint source-channel coding (JSCC) method over a differentiated service network was presented in [27]. The system scheduler considered the total energy

consumption and the transmission delay as the channel resource constraints for the video coding and transmission process. Discriminative coding decisions were applied to the shape packets and the texture packets in the video object coding in MPEG-4, as illustrated in Figure 4(a). Packets were selectively transmitted over different classes of service channels such that the optimal cost-distortion state was achieved under the energy and delay constraint.

The ROI based wireless video streaming system introduced in [28] adopted multiple error resilience schemes for data protection. The ROI region was assigned more resources than other areas including higher degree of forward error correction (FEC) and automatic repeat request (ARQ). For example, in the interleaving process displayed in Figure 4(b), the chessboard interleaving was performed on the ROI region with increased code rate and better error concealment result compared to the error concealment scheme with slice interleaving on background area.

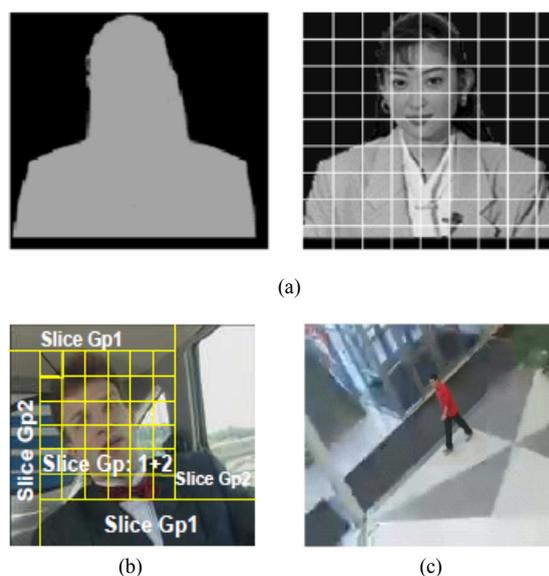


FIGURE 4. Unequal error protection. (a) Shape (left) and texture (right). (b) Interleaving. (c) Target region.

The system designed for surveillance video delivery over wireless sensor and actuator networks (WSANs) proposed in [29] extended the UEP mechanism from source data processing to network management. Each intermediate sensor node in the selected transmission path put the target packets ahead of all the background packets in the queue. Thus the target packet had a lower packet loss rate (PLR) than the background packet when the sensor node started dropping packets with a higher expected waiting time than the packet delay limit. The visual result of a received reconstructed frame can be observed from Figure 4(c).

Current video coding standards provide different interfaces for ROI data processing. For example, the object based video representation is supported in the MPEG-4 standard, and is more efficient when it is incorporated with the rate-distortion

estimation technique [30]. A contour free object shape coding method compatible with the SPIHT (Set Partition In Hierarchical Trees) codec [31] was introduced in [32]. In the latest H.264/AVC standard, several tools intended for error resilience like Flexible Macroblock Ordering (FMO) and Arbitrary Slice Ordering (ASO) can be used to define the ROI [33]. These interfaces enable convenient incorporation of the object based UEP mechanism into the coding process.

B. ERROR RESILIENCE

To cope with the signal distortion over the wireless channel, error resilience has been extensively studied to protect data transmission over WSNs. Some popular techniques include FEC, ARQ, adaptive modulation and coding (AMC), and channel aware resource allocation [34]–[39]. While traditional methods mainly focus on channel distortion, and are independent of the video coding process, more advanced error resilience techniques consider the end-to-end data distortion as the auxiliary information for making coding and/or transmission decisions, such as the JSCC method, the cross-layer control, and the multiple description coding. Multiple error resilience technologies have been adopted in the video codec standards H.263 and MPEG-4, as described in [39].

The JSCC method determines the coding parameters by estimating the end-to-end video distortion. In packetized video transmission over wireless networks, video compression and packet loss are two major causes for the data distortion observed by the receiver. Incorporating the packet loss information in the end-to-end distortion estimation process has been shown to be an efficient measure to improve the coding efficiency. In [34], a recursive optimal per-pixel estimate (ROPE) method was presented for the coding mode decision in block based coding process. This statistical model demonstrated a new way to adjust coding decisions according to both source coding and channel distortion. Another JSCC method introduced in [35] adopted random intra refreshing for error resilience. The source coding distortion was modeled as a function of the intra macro block (MB) refreshing rate, while the channel distortion was calculated in a similar recursive fashion as was done in [34]. This method also took into account the channel coding rate and FEC in the rate-distortion (RD) model. A further evolved type of channel aware WSN techniques that are considered efficient to deal with the packet loss is through the cross-layer control [36], [38], [40]. Both the source coding parameters and the transmission parameters are coordinated by the cross-layer controller to achieve the optimal end-to-end performance. More details about the cross-layer control mechanism will be introduced in Section IV-E. These techniques can be built upon current network protocols supporting video streaming, including TCP (Transmission Control Protocol), UDP (User Datagram Protocol), RTP (Real-time Transport Protocol)/RTCP (RTP control protocol), and RSVP (Resource ReSerVation Protocol) [41]–[43].

Another class of error resilience technique closely related to the video coding process is the multiple description coding (MDC). The main concept of MDC is to create

several independent descriptions that contribute to one or more characteristics of the original signal: spatial or temporal resolution, signal-to-noise ratio (SNR), or frequency content [43]–[46]. For video coding, the subdivision is usually performed in the spatial domain or in the temporal domain, such as separating the odd and even numbered frames [47], the spatial chessboard MB decomposition [28], and the spatiotemporal slice interleaving [48]. The descriptions can be generated in a way that each description is equally important to the reconstructed content. An example of constructing four balanced descriptions using spatial down-sampling to separate odd/even numbered rows and columns is displayed in Figure 5(a) [49]. The descriptions can also be constructed with unequal importance or with UEP. The asymmetric MDC (AMDC) scheme designed in [50] used layered coding to create unbalanced descriptions for several available channels with different bandwidths and loss characteristics in a heterogeneous network. The optimal description data length and FEC code rate for each channel were determined by an AMDC controller, as shown in Figure 5(b).

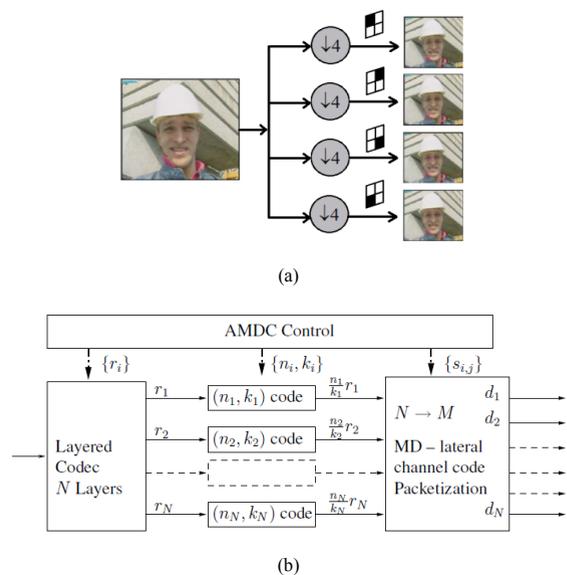


FIGURE 5. Balanced and unbalanced MDC. (a) Spatial downsampling. (b) Unbalanced descriptions.

MDC is considered to be an efficient measure to counteract bursty packet losses. Its robustness lies in the fact that it is unlikely the portions of the whole set of descriptions corresponding to the same part of the picture all corrupt during transmission. Each description can be independently decoded and the visual quality is improved when more descriptions are received. The compression efficiency of MDC is affected due to reduced redundancy in each description. The extra overhead is largely ignored when otherwise the complex channel coding schemes or the complex communication protocols have to be applied in the presence of high PLR. A comparison on the performance of the two description MDC coding (MD2) and the single description coding with Reed-Solomon code (SD+FEC) under the same

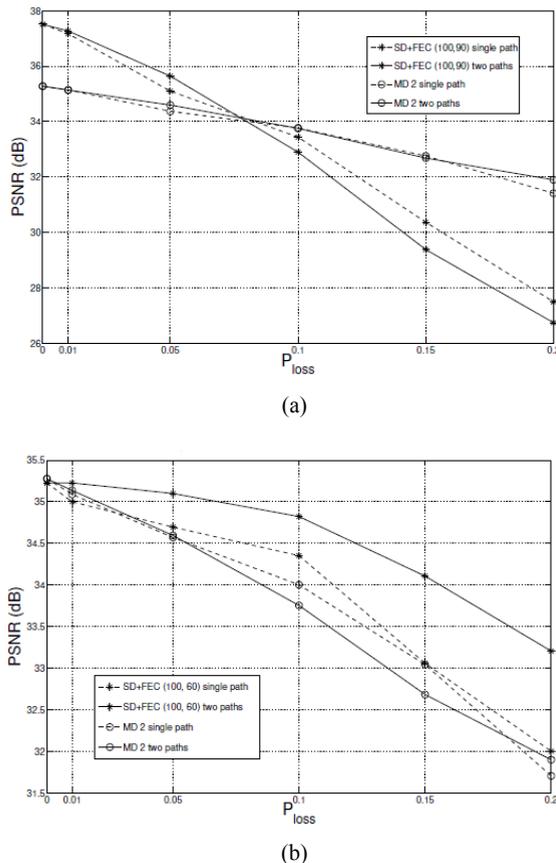


FIGURE 6. Comparison of MDC and SD+FEC. (a) Mean burst length = 10 packets. (b) Mean burst length = 20 packets.

data rate (*Foreman*, CIF, 30 fps, 850 kbps) is demonstrated in Figure 6 [51]. When the mean burst length is small, and the PLR is high, the MDC schemes outperform the FEC schemes.

Due to many advantages, MDC is favorably adopted in advanced coding paradigms including scalable video coding (SVC) and distributed video coding (DVC), as will be discussed in the following subsections. These coding techniques could enhance the adaptability of the wireless surveillance systems, especially when multiple cameras are recording simultaneously and the channel resource is strictly constrained.

C. SCALABLE VIDEO CODING

The development of SVC is intended for adaptive video delivery over heterogeneous networks. The basic idea is to encode the video into a scalable bitstream such that videos of lower qualities, spatial resolutions and/or temporal resolutions can be generated by simply truncating the scalable bitstream to meet the bandwidth conditions, terminal capabilities and quality of service (QoS) requirements in streaming video applications such as video transcoding and random access [52]. An SVC bit-stream consists of a base layer and one or more enhancement layers. The SVC feature is supported in several video coding standards including MPEG-2, MPEG-4, MJPEG 2000 and H.264/AVC [53]–[56].

The quality scalability of SVC is based on the progressive refinement data, such as the higher bit planes of the transform coefficients [52], and the prediction data with coarser quantization parameters (QPs) [57], added to the base layer. The spatial scalability is achieved by generating enhancement layers using video sequences of different resolutions. The data on higher layers are predicted from a scaled version of the reconstructed data on a lower layer [58]. The temporal scalability uses hierarchical B pictures for temporal decomposition. The pictures of the coarsest temporal resolution are encoded as the base layer, and B pictures are inserted at the next finer temporal resolution level in a hierarchical manner to construct the enhancement layers [56]. To improve the coding efficiency and granularity, a combination of SNR and spatiotemporal scalabilities is often adopted [59]–[63].

Compression efficiency and computation complexity are two major concerns in SVC applications. An influential concept to achieve efficient scalable coding is Motion Compensated Temporal Filtering (MCTF) based on the wavelet lifting scheme [64]. Figure 7(a) illustrates a two-channel analysis filter bank structure of MCTF consisting of the polyphase operation, prediction and update steps [65]. The output signals H_k and L_k can be viewed as high-pass and low-pass bands with motion compensation (MC) that spatially aligns separated input signals S_{2k} and S_{2k+1} towards each other. A three-band MCTF scheme was proposed in [60] to enhance the rate adaptation to the bandwidth variations in heterogeneous networks. The MCTF scheme proposed in [62] incorporated spatial scalability to further reduce inter-layer redundancy. A frame at a certain high-resolution layer was predicted both from the up-sampled frame at the next lower resolution layer, and the temporal neighboring frames within the same resolution layer through MC. For mobile devices with constrained computational resources, the SVC coding complexity scalability was considered in the work presented in [66]. Closed-form expressions were developed to predict the complexity measured in terms of the number of motion estimation (ME) computations, such that optimized rate-distortion-complexity tradeoffs can be achieved.

The dependency between nested layers is another hindrance for the application of SVC in wireless communications. The data on an enhancement layer are decodable only when the data on depended lower resolution layers are correctly recovered. To reduce the dependency, the work introduced by Crave *et al.* [67] applied MDC and Wyner-Ziv (WZ) coding [68], [69] in MCTF. The coding structure is displayed in Figure 7(b). Each description contained one normally encoded subsequence and another WZ encoded subsequence. The system achieved both enhanced error resilience ability and distributed data processing.

SVC has been applied in many video streaming systems [37], [70], [71]. The rate-distortion model related to the enhancement layer truncation, drift/error propagation, and error concealment in the scalable H.264/AVC video is discussed in details in [71]. A possible application in video surveillance is the interactive view selection. The user could

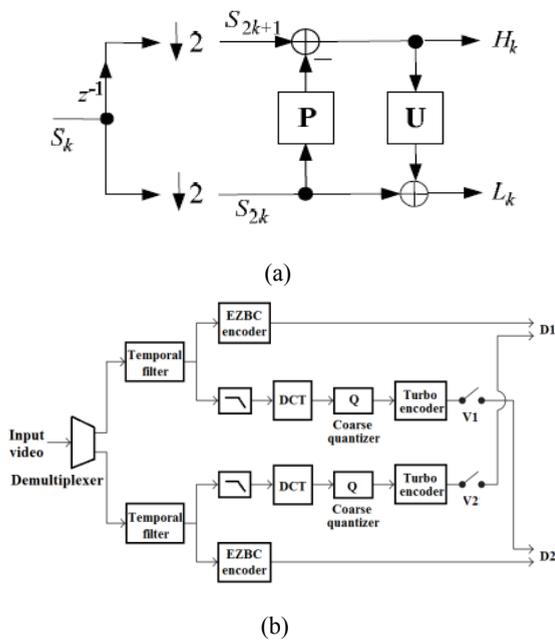


FIGURE 7. SVC coding structure. (a) MCTF. (b) MCTF with MDC.

randomly access the video data at any region, resolution, and/or view direction. To enable this function in the real-time video play, random access points are devised in the coding structure to realize smooth switching between different video streams. In H.264/AVC standard, the SP/SI slices are defined to achieve identical reconstruction of temporally co-located frames in different bitstreams coded at different bit-rates without causing drift [72]. This feature is especially useful for free-viewpoint applications [70], [73], [74].

D. DISTRIBUTED VIDEO CODING

DVC refers to the video coding paradigm applying the Distributed Source Coding (DSC) technology. DSC is based on the Slepian-Wolf (SW) [75] and WZ [68] theorems. In DSC, correlated signals are captured and compressed independently by different sensors, and are jointly decoded by the receiver [83]. Due to the many advantages of distributed data processing and the inherent spatiotemporal correlation in video data, DSC is applied in video coding in order to reduce encoder complexity and to maintain desirable error resilience ability [76].

Two representative architectures of DVC for single view video coding are the PRISM (Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding) [77] and the WZ coding structure [78]. In both schemes, part of the video data was compressed using the conventional intra coding method, and was prone to channel distortion. The rest was WZ encoded with coarser quantization and the error detection/correction code. At the decoder, the WZ data were restored using the correctly received intra coded data as side information. In [78], a feedback channel was adopted to request the error control information. The DISCOVER (DIStributed CODing for Video sERVICES) project presented in [79]

improved this coding structure with multiple enhancements including rate estimation and applying motion compensated temporal interpolation (MCTI) to obtain the side information. The work in [80] studied the effect of Group of Picture (GOP) size on the performance of DISCOVER with Low-Density Parity-Check (LDPC) codes for single view video coding. Based on the statistical analysis for the encoder time complexity and the RD performance, the DISCOVER encoder attained similar visual quality to the H.264/AVC encoder while the processing time was reduced by thirty percent on average. This feature makes the WZ coding scheme a competitive option for real-time video communication applications.

For multiview video coding, the inter-view correlation is also utilized by the decoder to restore the WZ data [76], [81]–[85]. The multiview DVC developed within the DISCOVER project applied both MCTI and homography compensated inter-view interpolation (HCII) in decoding [81], [76]. Similar coding structures for data processing in wavelet transform domain were reported [82]. The PRISM based multiview DVC presented in [85] incorporated disparity search (PRISM-DS) and view synthesis search in the decoding process. From the performance comparison with several simulcast coding schemes and the DISCOVER DVC scheme on visual quality under different PLR, the proposed coding scheme achieved better visual quality than the DISCOVER DVC scheme under low PLR, an average 2dB gain in PSNR when $PLR < 5\%$. The experimental results also revealed that under low level packet loss, the conventional MPEG/H.26x codec is still superior to the DVC schemes, and the additionally introduced complexity for reducing inter-view redundancy needs to be balanced with the coding efficiency.

E. CROSS-LAYER CONTROL

In the wireless video communication system, the limited channel resources are managed through configuring the options at different layers in the network architecture, such as the coding and error control at the application layer, the congestion control and reliability protocol at the transport layer, the routing at the network layer, the contention scheme at the MAC (medium access control) layer, and the MCS at the physical layer [86]. To jointly implement the configuration procedure, the cross-layer control methodology is developed to optimize the system-level resource allocation [87]. Given the channel state information (CSI), the controller is able to coordinate decision making at different layers in order to maximize the visual quality of the received video. The general optimization framework is formulated as a distortion minimization problem under certain resource constraints, typically the delivery delay constraint [29], [36], [38], [40], [88], and the transmission power constraint [27], [89]–[91]. The video coding scheme described in [40] considered the physical layer MCS in estimating the dynamic PLR in a Rayleigh fading channel. For video streaming over multi-hop WSNs, the systems demonstrated in [36] and [38] enabled adaptive configuration for both the physical layer MCS and the link layer path selection. A priority based packet

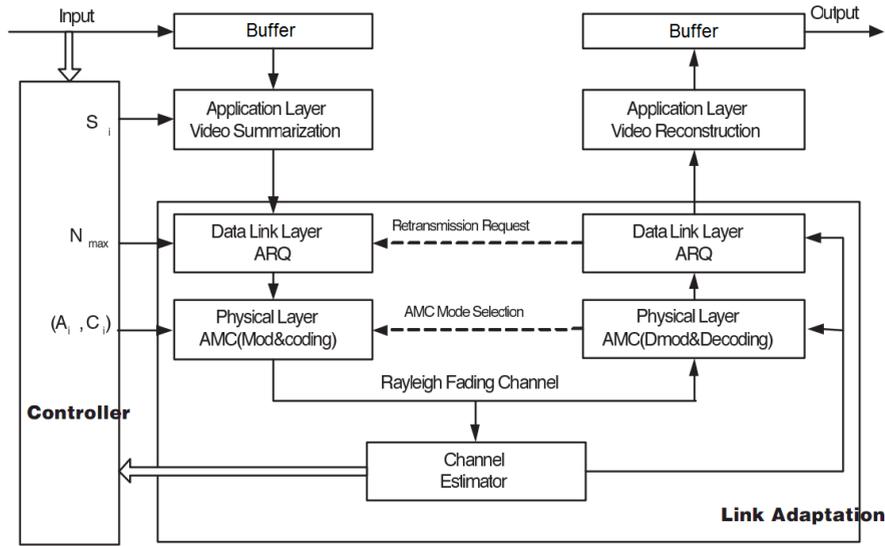


FIGURE 8. Cross-layer control model for wireless video streaming.

TABLE 3. Video technologies for wireless surveillance.

Video technology	Implementation	Advantage	Disadvantage
Object based UEP	Shape coding, priority queuing	Discriminative resource allocation	Extra computation for ROI detection and processing
Error resilience	FEC, ARQ, MDC, JSCC, cross-layer control	Better error control with CSI	Increased complexity for CSI collection and distortion estimation
Scalable video coding	Quality/spatial/temporal scalability	Rate adaptivity	Coding efficiency and complexity, dependence between layers
Distributed video coding	Single view/multiview DVC	Low encoding complexity	Deteriorated video quality

queuing mechanism at each sensor node is adopted in the video surveillance system designed in [29] to implement UEP for the target packets and the background packets at the transport layer. The work introduced in [88] incorporates congestion control with link adaptation for real-time video streaming over ad hoc networks. Power constraint is another consideration for energy efficient mobile devices. In [89], node cooperation is applied to optimally schedule the routing in order to minimize the energy consumption and delay. The cross-layer design presented in [90] jointly configured the physical, MAC, and routing layers to maximize the lifetime of energy-constrained WSNs. The object based video coding and transmission scheme developed in [27] performed UEP for the shape data and the texture data in the rate and energy allocation procedure. The optimal rate allocation policies introduced in [91] are developed to maximize aggregate throughput or to minimize queuing delays.

A standard formulation for the cross-layer optimization procedure can be expressed as

$$\begin{aligned} \min \quad & \sum E\{D(\psi_1, \psi_2, \dots, \psi_n, p)\} \\ \text{s.t.} \quad & C(\psi_1, \psi_2, \dots, \psi_n) \leq C^{\max} \end{aligned} \quad (1)$$

where $E\{D\}$ is the expected video data distortion under the system configuration set $\psi_1, \psi_2, \dots, \psi_n, p$ is the expected data loss over the WSN given the same configuration, $C(\psi_1, \psi_2, \dots, \psi_n)$ is the vector of corresponding consumed resources, and C^{\max} represents the resource constraints. The most challenging part in the procedure is to accurately predict the data loss information based on the system configuration set and the collected CSI from the time varying wireless network, in order to estimate the received data distortion. In online video communication application, the computational complexity of the solution procedure is also a primary concern. Figure 8 shows a paradigm of the cross-layer optimized video streaming scheme described in [38]. A summary of above potential technologies for a wireless video surveillance system is provided in Table 3.

V. VIDEO ANALYSIS

After transmission over the lossy channel, the video data is recovered by the receiver for observation and further analysis. In advanced video surveillance systems, two commonly studied applications are object detection and object tracking. A variety of techniques have been developed for related vision

tasks. For example, with fixed cameras, object detection takes advantage of static background. A popular technique is the background subtraction based on a Gaussian Mixture Model [18]. This temporal learning process models different conditions of a pixel at certain positions as a mixture of Gaussian distributions. The weight, mean, and variance values of each Gaussian model can be updated online, and pixels not conforming to any background model are quickly detected. The adaptive learning property makes this technique suitable for real-time applications [20], [21], [92]. Other detection methods include the region segmentation based graph cut [93], edge detection based variational level set [94], and compressive sensing [95].

With active cameras such as PTZ cameras, the detection method needs to consider the changing background in the recorded video. The feature point matching algorithm has been widely studied for the purpose of robust object detection and tracking, including the scale invariant feature transform [96], and the kernel filtering algorithm [97]. These point matching methods are costly to implement and hence are not suitable for real-time applications. The RANSAC (RANDOM SAMPLE CONSENSUS) algorithm is often adopted for fast implementation of the point matching process [98], [99]. In the video object detection scheme presented in [99], the moving target is detected through subtracting the background image synthesized by the homography-RANSAC algorithm, which is based on the special property of the PTU camera movement. The object detection procedure is illustrated in Figure 9.

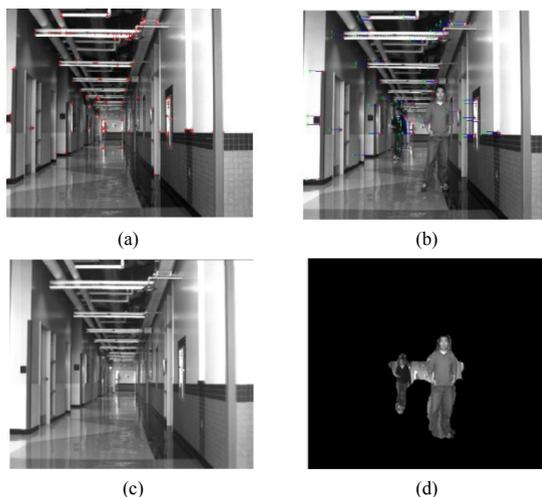


FIGURE 9. Object detection procedure: (a) feature point detection on precaptured background image; (b) feature point correspondence on the recorded image with homography-RANSAC; (c) background synthesis using the estimated homography; (d) object segmentation with background subtraction and level-set contour.

Another well known motion detection technique under dynamic scene is the optical flow [100]. The affine transformation between consecutive frames is estimated such that the motion area not conforming to the transformation stands out. Real-time computation of optical flow was presented in

[101]. In [102], the disparity information was combined with optical flow for binocular view object tracking. Other popular methods for dynamic object tracking include Lucas-Kanade-Tomasi tracker [103], Mean Shift [24], level set contour [104], and the techniques fusing multiple properties of the video data [105], [106]. In multiview object detection/tracking, the problem of object correspondence in different view was discussed in [17], [107]. The camera control algorithm based on the object detection result was also rigorously studied for tracking with active cameras [22], [70].

Other vision technologies, such as super resolution [108], view synthesis [109], and 3D model reconstruction [110], can be possibly applied to a video surveillance system. However, most of these technologies are either based on undistorted video data, or are independent of the error control procedure at the transmitter. The impact of video compression on RD performance was considered in several vision applications for optimal source coding decisions at the transmitter, including view synthesis [111], [112], object tracking [113], and super resolution [114]. Some JSCC schemes were embedded in the coding structure for optimal resource allocation based on the end-to-end distortion estimation [27], [35], [115]–[117]. The channel distortion model for more complex vision applications remains a challenging research topic.

VI. OTHER ISSUES

Data security is an important issue in secret communications in sensor networks [118]. For video data, the encryption can be performed on the compressed bitstream using well established cryptographic algorithms, such as the built-in authentication and AES (Advanced Encryption Standard) encryption defined in the IEEE 802.16/WiMax standard [119]. For a large amount of video data, the resource allocated for security protection has to be balanced with the error control effort supported by the wireless communication system, in order to achieve the optimal end-to-end secrecy.

The encryption can also be performed within the coding process using the video scrambling technique [120], without adverse impact on error resilience. Moreover, video watermarking has been proved to be an efficient measure for data protection and authentication in WSNs [121], [122]. These security measures often come with the reduced coding efficiency, and the requirement of more advanced error concealment techniques for recovering the corrupted video.

Privacy is another issue gaining increasing attention in video surveillance systems [123]. A major concern regarding this issue is that some contents of the surveillance video, such as those involving personal identity, are inappropriate or illegal to be displayed directly in front of the audience. Current methods applied to address this issue are based on object detection techniques [124], especially the facial recognition techniques. The content-aware coding method proposed in [125] utilized the spatial scalability features of the JPEG XR (JPEG extended range) codec for face masking. The face regions were detected and scrambled in the transform domain. In another shape coding scheme [126], the object region was

encrypted independently in the SPIHT based coding process, with enhanced coding efficiency compared to the contour based block coding in MPEG-4. The implementation of privacy measures in a real-time surveillance application could be very difficult, as the prerequisite to identify the sensitive content or to detect the unusual event is a challenging task itself.

VII. CONCLUSION

Wireless video surveillance is popular in various visual communication applications. IMS Research has predicted that the global market for wireless infrastructure gear used for video surveillance applications will double up from 2011 to 2016 [127]. This paper presents a survey on the technologies dedicated to different functional modules of a video surveillance system. A comprehensive system design would require interdisciplinary study to seamlessly incorporate different modules into an optimal system-level resource allocation framework. While the advanced WSN infrastructure provides a strong support for surveillance video communications, new challenges are emerging in the process of compressing and transmitting large amounts of video data, and in the presence of run time and energy conservation requirements for mobile devices. Another trend in this field is the 3D signal processing technology in more advanced multiview video surveillance. The wireless communication environment posts greater difficulty for this kind of applications. How to efficiently estimate the distortion for the dedicated vision task at the receiving end using the compressed and concealed video data is essential to the system performance.

REFERENCES

- [1] S. Leader. (2004). "Telecommunications handbook for transportation professionals—The basics of telecommunications," Federal Highway Administration, Washington, DC, USA, Tech. Rep. FHWA-HOP-04-034 [Online]. Available: http://ops.fhwa.dot.gov/publications/telecomm_handbook/telecomm_handbook.pdf
- [2] J. Hourdakis, T. Morris, P. Michalopoulos, and K. Wood. (2005). "Advanced portable wireless measurement and observation station," Center for Transportation Studies in Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. CTS 05-07 [Online]. Available: <http://conservancy.umn.edu/bitstream/9591/CTS-05-07.pdf>
- [3] N. Luo, "A wireless traffic surveillance system using video analytics," M.S. thesis, Dept. Comput. Sci. Eng., Univ. North Texas, Denton, TX, USA, 2011.
- [4] C. Hartung, R. Han, C. Seielstad, and S. Holbrook, "FireWxNet: A multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments," in *Proc. 4th Int. Conf. Mobile Syst., Appl. Services*, 2006, pp. 28–41.
- [5] A. Kawamura, Y. Yoshimitsu, K. Kajitani, T. Naito, K. Fujimura, and S. Kamijo, "Smart camera network system for use in railway stations," in *Proc. Int. Conf. Syst., Man, Cybern.*, 2011, pp. 85–90.
- [6] N. Li, B. Yan, G. Chen, P. Govindaswamy, and J. Wang, "Design and implementation of a sensor-based wireless camera system for continuous monitoring in assistive environments," *J. Personal Ubiquitous Comput.*, vol. 14, no. 6, pp. 499–510, Sep. 2010.
- [7] B. P. L. Lo, J. Sun, and S. A. Velastin, "Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems," *Acta Autom. Sinica*, vol. 29, no. 3, pp. 393–407, 2003.
- [8] W. Feng, B. Code, M. Shea, and W. Feng, "Panoptes: A scalable architecture for video sensor networking applications," in *Proc. ACM Multimedia*, 2003, pp. 151–167.
- [9] S. Hengstler, D. Prashanth, S. Fong, and H. Aghajan, "MeshEye: A hybrid-resolution smart camera mote for applications in distributed intelligent surveillance," in *Proc. Int. Symp. Inf. Process. Sensor Netw.*, 2007, pp. 360–369.
- [10] X. Wang, S. Wang, and D. Bi, "Distributed visual-target-surveillance system in wireless sensor networks," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1134–1146, Oct. 2009.
- [11] *Electronic Code of Federal Regulations* [Online]. Available: http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=1143b55e16daf5dce6d225ad4dc6514a&tpl=/ecfrbrowse/Title47/47cfr15_main_02.tpl
- [12] M. Intag. (2009). *Wireless Video Surveillance—Challenge or Opportunity?* [Online] Available: <http://www.bicsi.org/pdf/conferences/winter/2009/presentations/Wireless%20Security%20and%20Surveillance%20-%20Challenge%20or%20Opportunity%20-%20Mike%20Intag.pdf>
- [13] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "SensEye: A multi-tier camera sensor network," in *Proc. 13th Annu. ACM Multimedia*, 2005, pp. 229–238.
- [14] Y. Wang and G. Cao, "On full-view coverage in camera sensor networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1781–1789.
- [15] Y. Wang and G. Cao, "Barrier coverage in camera sensor networks," in *Proc. 12th ACM Int. Symp. Mobile Ad Hoc Network. Comput.*, 2011, pp. 1–10.
- [16] M. Johnson and A. Bar-Noy, "Pan and scan: Configuring cameras for coverage," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1071–1079.
- [17] T. J. Ellis and J. Black, "A multi-view surveillance system," in *Proc. IEEE Symp. Intell. Distrib. Surveill. Syst.*, London, U.K., Feb. 2003, pp. 11/1–11/5.
- [18] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.
- [19] M. Xu and T. J. Ellis, "Illumination-invariant motion detection using color mixture models," in *Proc. BMVC*, Manchester, U.K., Sep. 2001, pp. 163–172.
- [20] S. Babacan and T. Pappas, "Spatiotemporal algorithm for background subtraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2007, pp. 1-1065–1-1068.
- [21] J. Gallego, M. Pardo, and G. Haro, "Bayesian foreground segmentation and tracking using pixel-wise background model and region based foreground model," in *Proc. Int. Conf. Image Process.*, Nov. 2009, pp. 3205–3208.
- [22] P. Petrov, O. Boumbarov, and K. Muratovski, "Face detection and tracking with an active camera," in *Proc. 4th Int. Conf. Inf. Syst.*, Sep. 2008, pp. 14-34–14-39.
- [23] T. W. Yang, K. Zhu, Q. Q. Ruan, and J. D. Han, "Moving target tracking and measurement with a binocular vision system," in *Proc. Int. Conf. Mech. Mach. Vis. Pract.*, Dec. 2008, pp. 85–91.
- [24] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, May 2003.
- [25] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Oct. 1998, pp. 214–219.
- [26] Y. Ye, S. Ci, Y. Liu, H. Wang, and A. K. Katsaggelos, "Binocular video object tracking with fast disparity estimation," in *Proc. Int. Conf. Adv. Video Signal-Based Surveill.*, Aug. 2013.
- [27] H. Wang, F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Cost-distortion optimized unequal error protection for object-based video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1505–1516, Dec. 2005.
- [28] R. Chakravorty, S. Banerjee, and S. Ganguly, "MobiStream: Error-resilient video streaming in wireless WANs using virtual channels," in *Proc. INFOCOM*, Apr. 2006, pp. 1–14.
- [29] D. Wu, S. Ci, H. Luo, Y. Ye, and H. Wang, "Video surveillance over wireless sensor and actuator networks using active cameras," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2467–2472, Oct. 2011.
- [30] A. K. Katsaggelos, L. P. Kondi, F. W. Meier, J. Ostermann, and G. M. Schuster, "MPEG-4 and rate-distortion-based shape-coding techniques," *Proc. IEEE*, vol. 86, no. 6, pp. 1126–1154, Jun. 1998.
- [31] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

- [32] K. Martin, R. Lukac, and K. N. Plataniotis, "SPIHT-based coding of the shape and texture of arbitrarily shaped visual objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 10, pp. 1196–1208, Oct. 2006.
- [33] Y. Dhondt, P. Lambert, S. Notebaert, and R. Van de Walle, "Flexible macroblock ordering as a content adaptation tool in H.264/AVC," *Proc. SPIE*, vol. 6015, pp. 601506.1–601506.9, Oct. 2005.
- [34] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *J. Sel. Areas Commun.*, vol. 18, pp. 966–976, Jun. 2000.
- [35] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
- [36] Y. Andreopoulos, N. Mastrorade, and M. van der Schaar, "Cross-layer optimized video streaming over wireless multi-hop mesh networks," *J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2104–2115, Nov. 2006.
- [37] P. Pahalawatta, R. Berry, T. Pappas, and A. K. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 749–759, 2007.
- [38] D. Wu, S. Ci, and H. Wang, "Cross-layer optimization for video summary transmission over wireless networks," *J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 841–850, May 2007.
- [39] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 4, pp. 61–82, Jul. 2000.
- [40] Z. Chen and D. Wu, "Rate-distortion optimized cross-layer rate control in wireless video communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 352–365, Mar. 2012.
- [41] J. Postel. (1980, Aug.). *RFC 768—User Datagram Protocol*, USC/Information Sciences Inst., Marina del Rey, CA, USA [Online] Available: <http://tools.ietf.org/html/rfc768>
- [42] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. (1996, Jan.). *RFC 1889—RTP: A Transport Protocol for Real-Time Applications* Audio-Video Transport Working Group [Online] Available: <http://www.freesoft.org/CIE/RFC/1889/>
- [43] (1997). *Resource Reservation Protocol* [Online]. Available: <http://www.isi.edu/rsvp/>
- [44] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.
- [45] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 3, pp. 321–334, Mar. 2005.
- [46] E. Akyol, A. Murat Tekalp, and M. Reha Civanlar, "Scalable multiple description video coding with flexible number of descriptions," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, pp. 712–715.
- [47] J. G. Apostolopoulos and S. J. Wee, "Unbalanced multiple description video communication using path diversity," in *Proc. Int. Conf. Image Process.*, Oct. 2001, pp. 966–969.
- [48] Y. Wang, J. Y. Tham, W. S. Lee, and K. H. Goh, "Pattern selection for error-resilient slice interleaving based on receiver error concealment technique," in *Proc. Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–4.
- [49] A. Vitali. (2007, Oct.). *Multiple Description Coding—A New Technology for Video Streaming over the Internet*, EBU Technical Review [Online]. Available: http://tech.ebu.ch/docs/techreview/trev_312-vitali_streaming.pdf
- [50] J. R. Taal and I. L. Lagendijk, "Asymmetric multiple description coding using layered coding and lateral error correction," in *Proc. Symp. Inf. Theory Benelux*, Jun. 2006, pp. 39–44.
- [51] R. Bernardini, M. Durigon, R. Rinaldo, and A. Vitali, "Comparison between multiple description and single description video coding with forward error correction," in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, Oct./Nov. 2005, pp. 1–4.
- [52] H. Sun, A. Vetro, and J. Xin, "An overview of scalable video streaming," *J. Wireless Commun. Mobile Comput.*, vol. 7, no. 2, pp. 159–172, 2007.
- [53] *Generic Coding of Moving Pictures and Associated Audio*, ISO/IEC Standard JTC1 IS 13818 (MPEG-2), 1994.
- [54] *Generic Coding of Moving Pictures and Associated Audio*, ISO/IEC Standard JTC1 IS 14386 (MPEG-4), 2000.
- [55] "Information technology—JPEG 2000 image coding system: Motion JPEG 2000," T.802, 2000.
- [56] *Annex G of H.264/AVC/MPEG-4 Part 10: Scalable Video Coding (SVC)*, Standard ISO/IEC 14496-10, 2007.
- [57] T. Schierl, K. Ganger, C. Hellge, T. Wiegand, and T. Stockhammer, "SVC-based multisource streaming for robust video transmission in mobile ad hoc networks," *IEEE Wireless Commun.*, vol. 13, no. 5, pp. 96–103, Oct. 2006.
- [58] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [59] L. Luo, J. Li, S. Li, Z. Zhuang, and Y. Zhang, "Motion compensated lifting wavelet and its application in video coding," in *Proc. IEEE Int. Conf. Multimedia Expo*, Aug. 2001, pp. 365–368.
- [60] C. Tillier and B. Pesquet-Popescu, "3D, 3-band, 3-TAP temporal lifting for scalable video coding," in *Proc. Int. Conf. Image Process.*, vol. 2, 2003, pp. 779–782.
- [61] N. Mehrseresh and D. Taubman, "A flexible structure for fully scalable motion-compensated 3-D DWT with emphasis on the impact of spatial scalability," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 740–753, Mar. 2006.
- [62] R. Xiong, J. Xu, and F. Wu, "In-scale motion compensation for spatially scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 145–158, Feb. 2008.
- [63] S. Xiang and L. Cai, "Scalable video coding with compressive sensing for wireless videocast," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–5.
- [64] W. Sweldens, "A custom-design construction of biorthogonal wavelets," *J. Appl. Comput. Harmonic Anal.*, vol. 3, no. 2, pp. 186–200, 1996.
- [65] R. Schafer, H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "MCTF and scalability extension of H.264/AVC and its application to video transmission, storage, and surveillance," in *Proc. Int. Conf. Vis. Commun. Image Process.*, Jul. 2005, pp. 1–12.
- [66] D. S. Turaga, M. van der Schaar, and B. Pesquet-Popescu, "Complexity scalable motion compensated wavelet video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 982–993, Aug. 2005.
- [67] O. Crave, C. Guillemot, B. Pesquet-Popescu, and C. Tillier, "Distributed temporal multiple description coding for robust video transmission," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, article id 183536, pp. 1–13, Jul. 2007.
- [68] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [69] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.
- [70] J. G. Lou, H. Cai, and J. Li, "A real-time interactive multi-view video system," in *Proc. ACM Multimedia*, Nov. 2005, pp. 161–170.
- [71] E. Maani and A. K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 407–416, Mar. 2010.
- [72] M. Karczewisz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 637–644, Jul. 2003.
- [73] Y. Liu, Q. Huang, D. Zhao, and W. Gao, "Low-delay view random access for multi-view video coding," in *Proc. Int. Symp. Circuits Syst.*, May 2007, pp. 997–1000.
- [74] Z. Pan, Y. Ikuta, M. Bandai, and T. Watanabe, "A user dependent system for multi-view video transmission," in *Proc. IEEE Int. Conf. Adv. Inf. Neww. Appl.*, Mar. 2011, pp. 732–739.
- [75] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Mar. 1973.
- [76] F. Dufaux, M. Ouaret, and T. Ebrahimi, "Recent advances in multi-view distributed video coding," *Proc. SPIE*, vol. 6579, pp. 657902-1–657902-11, May 2007.
- [77] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. Commun., Control Comput.*, Allerton, IL, USA, Oct. 2002, pp. 1–10.
- [78] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Conf. Rec. 36th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2002, pp. 240–244.
- [79] X. Artigas, J. Ascenso, M. Dalai, S. Klomp D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Proc. Picture Coding Symp.*, 2007, pp. 1–4.

- [80] F. Pereira, J. Ascenso, and C. Brites, "Studying the GOP size impact on the performance of a feedback channel-based Wyner-Ziv video codec," in *Proc. PSIVT*, 2007, pp. 801–815.
- [81] X. Artigas, E. Angeli, and L. Torres, "Side information generation for multiview distributed video coding using a fusion approach," in *Proc. 7th Nordic Signal Process. Symp.*, 2006, pp. 250–253.
- [82] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multi-view video coding," *Proc. SPIE*, vol. 6077, pp. 60770T-1–60770T-8, Jan. 2006.
- [83] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, "Distributed monoview and multiview video coding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 67–76, Sep. 2007.
- [84] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Iterative multiview side information for enhanced reconstruction in distributed video coding," *EURASIP J. Image Video Process.*, vol. 2009, pp. 1–17, Mar. 2009.
- [85] C. Yeo and K. Ramchandran, "Robust distributed multiview video compression for wireless camera networks," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 995–1008, Apr. 2010.
- [86] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks," *IEEE Commun. Surv. Tuts.*, vol. 10, no. 4, pp. 18–39, Jan. 2009.
- [87] M. Van der Schaar and S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms," *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 50–58, Aug. 2005.
- [88] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 59–65, Aug. 2005.
- [89] S. Cui and A. J. Goldsmith, "Cross-layer optimization of sensor networks based on cooperative MIMO techniques with rate adaptation," in *Proc. IEEE 6th Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2005, pp. 960–964.
- [90] R. Madan, S. Cui, S. Lall, and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3142–3152, Nov. 2006.
- [91] A. Scaglione and M. Van der Schaar, "Cross-layer resource allocation for delay constrained wireless video transmission," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2005, pp. 909–912.
- [92] P. Suo and Y. Wang, "An improved adaptive background modeling algorithm based on Gaussian mixture model," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 1436–1439.
- [93] P. Tang and L. Gao, "Video object segmentation based on graph cut with dynamic shape prior constraint," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
- [94] M. Ristivojević and J. Konrad, "Space-time image sequence analysis: Object tunnels and occlusion volumes," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 364–376, Feb. 2006.
- [95] H. Jiang, W. Deng, and Z. Shen, "Surveillance video processing using compressive sensing," *Inverse Problems Imag.*, vol. 6, no. 2, pp. 201–214, 2012.
- [96] D. G. Lowe, "Distinctive image features from scale-invariant key points," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [97] B. Georgescu and P. Meer, "Point matching under large image deformations and illumination changes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 674–688, Jun. 2004.
- [98] Y. Jin, L. Tao, H. Di, N. Rao, and G. Xu, "Background modeling from a free-moving camera by multi-layer homography algorithm," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 1572–1575.
- [99] Y. Ye, S. Ci, Y. Liu, and H. Tang, "Dynamic video object detection with single PTU camera," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.*, Nov. 2011, pp. 1–4.
- [100] 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.
- [101] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr, "Real-time optic flow computation with variational methods," in *Proc. Int. Conf. Images Patterns*, 2003, pp. 222–229.
- [102] T. Dang, C. Hoffmann, and C. Stiller, "Fusing optical flow and stereo disparity for object tracking," in *Proc. Int. Conf. Intell. Transp. Syst.*, 2002, pp. 112–117.
- [103] C. Tomasi and T. Kanade, "Detection and tracking of point features," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, Apr. 1991.
- [104] N. Paragiosa and R. Deriche, "Geodesic active regions and level set methods for motion estimation and tracking," *Comput. Vis. Image Understand.*, vol. 97, no. 3, pp. 259–282, 2005.
- [105] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.
- [106] A. Suga, K. Fukuda, T. Takiguchi, and Y. Ariki, "Object recognition and segmentation using SIFT and graph cuts," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
- [107] G. Mohammadi, F. Dufaux, T. H. Minh, and T. Ebrahimi, "Multi-view video segmentation and tracking for video surveillance," *Proc. SPIE*, vol. 7351, pp. 735104-1–735104-11, Apr. 2009.
- [108] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*. San Rafael, CA, USA: Morgan & Claypool, Jan. 2007.
- [109] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, 2004.
- [110] J.-Y. Guillemot, J. Kilner, and A. Hilton, "Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes," in *Proc. IEEE 12th ICCV*, Sep./Oct. 2009, pp. 809–816.
- [111] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *Proc. Picture Coding Symp.*, Apr. 2006, pp. 38–39.
- [112] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model," *Signal Process., Image Commun.*, vol. 24, no. 8, pp. 666–681, Aug. 2009.
- [113] E. Soyak, S. A. Tsiftaris, and A. K. Katsaggelos, "Quantization optimized H.264 encoding for traffic video tracking applications," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 1241–1244.
- [114] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 898–911, Jul. 2004.
- [115] A. S. Tan, A. Aksay, G. B. Akar, and E. Arıkan, "Rate-distortion optimization for stereoscopic video streaming with unequal error protection," *EURASIP J. Appl. Signal Process.*, vol. 2009, pp. 1–14, Jan. 2009.
- [116] L. X. Liu, G. Cheung, and C.-N. Chuah, "Rate-distortion optimized joint source/channel coding of WWAN multicast video for a cooperative peer-to-peer collective," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 1, pp. 39–52, Jan. 2011.
- [117] C. Hou, W. Xiang, and F. Wu, "Channel distortion modeling for multi-view video transmission over packet-switched networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1679–1692, Nov. 2011.
- [118] C.-Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.
- [119] D. Johnston. *AES-CCM Encryption and Authentication Mode for 802.16* [Online]. Available: http://www.ieee802.org/16/tge/contrib/C80216e-04_12.pdf
- [120] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 118–129, Mar. 2003.
- [121] N. Checcacci, M. Barni, F. Bartolini, and S. Basagni, "Robust video watermarking for wireless multimedia communications," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 3, Sep. 2000, pp. 1530–1535.
- [122] M. Chen, Y. He, and R. L. Lagendijk, "A fragile watermark error detection scheme for wireless video communications," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 201–211, Apr. 2005.
- [123] C. S. Regazzoni, V. Ramesh, and G. L. Foresti, "Special issue on video communications, processing, and understanding for third generation surveillance systems," *Proc. IEEE*, vol. 89, no. 10, pp. 1355–1365, Oct. 2001.
- [124] J. Wickramasuriya, M. Alhazzazi, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy-protecting video surveillance," *Proc. SPIE*, vol. 5671, pp. 64–75, Mar. 2005.
- [125] H. Sohn, W. De Neve, and R. Y. Man, "Privacy protection in video surveillance systems: Analysis of subband-adaptive scrambling in JPEG XR," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 2, pp. 170–177, Feb. 2011.
- [126] K. Martin and K. N. Plataniotis, "Privacy protected surveillance using secure visual object coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1152–1162, Aug. 2008.
- [127] IMS Research. (2013). *Market for Wireless Infrastructure Gear for Video Surveillance Set to More than Double by 2016*, Wellingborough, U.K. [Online]. Available: http://www.imsresearch.com/news-events/press-template.php?pr_id=3387



YUN YE received the B.S. degree in electrical engineering from Sun Yat-Sen University, Guangzhou, China, in 2005, the master's degree in telecommunications from Shanghai Jiaotong University, Shanghai, China, in 2008, and the Ph.D. degree in computer engineering from the University of Nebraska-Lincoln, Omaha, NE, USA, in 2013. Currently, she is a Research Associate with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln. Her research interests include video surveillance, wireless multimedia communications, and 3D multimedia signal processing.



SONG CI (S'98–M'02–SM'06) received the B.S. degree from the Shandong University of Technology (now Shandong University), Jinan, China, in 1992, the M.S. degree from the Chinese Academy of Sciences, Beijing, China, in 1998, and the Ph.D. degree from the University of Nebraska-Lincoln, Omaha, NE, USA, in 2002, all in electrical engineering.

He is currently an Associate Professor with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln. Prior to joining University of Nebraska-Lincoln, he was an Assistant Professor of computer science with the University of Massachusetts, Boston, MA, USA and the University of Michigan, Flint, MI, USA. His current research interests include dynamic complex system modeling and optimization, green computing and power management, dynamically reconfigurable embedded system, content-aware quality-driven cross-layer optimized multimedia over wireless, cognitive network management and service-oriented architecture, and cyber-enable e-healthcare.

Dr. Ci serves as a Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE Network Magazine, and an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, an Associate Editor in the editorial board of Wiley Wireless Communications and Mobile Computing, and an Associate Editor of the Journal of Computer Systems, Networks, and Communications, Journal of Security and Communication Networks, and Journal of Communications. He serves as the technical program committee (TPC) co-chair, the TPC vice chair, or TPC member for numerous conferences. He won the Best Paper Award of the 2004 IEEE International Conference on Networking, Sensing, and Control. He is a recipient of the 2009 Faculty Research and Creative Activity Award at the College of Engineering of the University of Nebraska-Lincoln.



AGGELOS K. KATSAGGELOS (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1981 and 1985, respectively.

He joined the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, in 1985, where he is currently a Professor. He held the Ameritech Chair of information technology from 1997 to 2003. He is the Director of the Motorola Center for Seamless Communications, a member of the academic staff at NorthShore University Health System, and an affiliated faculty with the Department of Linguistics and the Argonne National Laboratory. He has published extensively and he holds

16 international patents. He is the co-author of *Rate-Distortion Based Video Compression* (Norwell, MA: Kluwer, 1997), *Super-Resolution for Images and Video* (San Rafael, CA: Claypool, 2007), and *Joint Source-Channel Video Transmission* (San Rafael, CA: Claypool, 2007).

Dr. Katsaggelos was the Editor-in-Chief of the IEEE SIGNAL PROCESSING MAGAZINE from 1997 to 2002, a BOG Member of the IEEE Signal Processing Society from 1999 to 2001, and a Publication Board Member of the IEEE Proceedings from 2003 to 2007. He became a Fellow of the SPIE in 2009 and was a recipient of the IEEE Third Millennium Medal in 2000, the IEEE Signal Processing Society Meritorious Service Award in 2001, the IEEE Signal Processing Society Best Paper Award in 2001, the IEEE ICME Paper Award in 2006, the IEEE ICIP Paper Award in 2007, and the ISPA Paper Award in 2009. He was a Distinguished Lecturer of the IEEE Signal Processing Society from 2007 to 2008.



YANWEI LIU received the B.S. degree in applied geophysics from Jiangnan Petroleum University, Jingzhou, China, in 1998, the M.S. degree in computer science from China Petroleum University, Beijing, China, in 2004, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2010.

He joined the Institute of Acoustics, Chinese Academy of Sciences, in 2010, as an Assistant Researcher. His research interests include digital image/video processing, multiview and 3D video coding, and wireless video communication.



YI QIAN is an Associate Professor in the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln (UNL), Lincoln, NE, USA. Prior to joining UNL, he involved in the telecommunications industry, academia, and the government. Some of his previous professional positions include serving as a senior member of scientific staff and a technical advisor at Nortel Networks, a senior systems engineer and a technical advisor at several start-up companies, an Assistant Professor with the University of Puerto Rico at Mayaguez, Puerto Rico, Brazil, and a Senior Researcher with the National Institute of Standards and Technology, Gaithersburg, MD, USA. His research interests include information assurance and network security, network design, network modeling, simulation and performance analysis for next generation wireless networks, wireless ad-hoc and sensor networks, vehicular networks, broadband satellite networks, optical networks, high-speed networks, and the Internet. He has a successful track record to lead research teams and to publish research results in leading scientific journals and conferences. His recent journal articles on wireless network design and wireless network security are among the most accessed papers in the IEEE Digital Library. He is a member of ACM.

•••