Compressed Sensing Based Low-Power Multi-View Video Coding and Transmission in Wireless Multi-Path Multi-Hop Networks

Nan Cen, *Member, IEEE*,
Zhangyu Guan, *Senior Member, IEEE*, and Tommaso Melodia, *Fellow, IEEE*

Abstract—Wireless Multimedia Sensor Network (WMSN) is increasingly being deployed for surveillance, monitoring and Internet-of-Things (IoT) sensing applications where a set of cameras capture and compress local images and then transmit the data to a remote controller. Such captured local images may also be compressed in a multi-view fashion to reduce the redundancy among overlapping views. In this paper, we present a novel paradigm for compressed-sensing-enabled multi-view coding and streaming in WMSN. We first propose a new encoding and decoding architecture for multi-view video systems based on Compressed Sensing (CS) principles, composed of cooperative sparsity-aware block-level rate-adaptive encoders, feedback channels and independent decoders. The proposed architecture leverages the properties of CS to overcome many limitations of traditional encoding techniques, specifically massive storage requirements and high computational complexity. Then, we present a modeling framework that exploits the aforementioned coding architecture. The proposed mathematical problem minimizes the power consumption by jointly determining the encoding rate and multi-path rate allocation subject to distortion and energy constraints. Extensive performance evaluation results show that the proposed framework is able to transmit multi-view streams with guaranteed video quality at lower power consumption.

Index Terms—Compressed sensing, multi-view video streaming, network optimization, Internet of Things

1 Introduction

Wireless Multimedia Sensor Networks (WMSNs) are composed of low-cost, battery-operated wireless camera sensors with the ability of acquiring, processing and transmitting visual data. By extending the capability of traditional Wireless Sensor Networks (WSNs), WMSNs play a paramount role in the evolution of the Internet-of-Things (IoTs) paradigm by enabling multi-media data gathering, processing and analysis, for example, disaster monitoring, pervasive surveillance, traffic and infrastructure monitoring [2] in the scenario of smart cities. However, WMSN poses additional challenges compared to traditional WSN because it requires intense processing ability and high-bandwidth availability. Given the fact that the sensor nodes in WMSNs are characterized by tight energy, limited processing and bandwidth, how to design a low-complexity low-power framework pertaining to data compression, processing and networking is a critical issue.

Recently, compressed sensing (CS) has been proposed as a possible solution to enable video streaming in resource constrained WMSNs. CS-based imaging systems are able to reconstruct image or video signals from a relatively "small" number of (random or deterministic) linear combinations of original image pixels, referred to as measurements, *without collecting the entire frame* [3], [4], thereby offering a promising alternative to traditional video encoders by *acquiring and compressing video or images simultaneously at very low computational complexity for encoders* [5]. This attractive feature motivated a number of works that have applied CS to video streaming in low-power wireless surveillance scenarios. For example, [6], [7], [8] mainly concentrate on single-view CS-based video compression, by exploiting temporal correlation among successive video frames [6], [7] or considering energy-efficient rate allocation in WMSNs with traditional CS reconstruction methods [8]. In [9], we showed that CS-based wireless video streaming can deliver surveillance-grade video for a fraction of the energy consumption of traditional systems based on predictive video encoding such as H.264. In addition, [8] illustrated and evaluated the *error-resilience* property of CS-based video streaming, which results in graceful quality degradation in wireless lossy links. A few recent contributions [10], [11], [12], [13] have proposed CS-based multi-view video streaming techniques, primarily focusing on an independent-encoder and joint-decoder paradigm, which exploits the implicit correlation among multiple views at the decoder side to improve the resulting video quality using complex joint reconstruction algorithms.

From a system view of multi-view video streaming, besides visual data acquiring and compressing, how to achieve power-efficient quality-assured data transmission over a multi-hop wireless sensor network is another

- *Nan Cen is with the Department of Computer Science, Missouri Univeristy of Science and Technology, Rolla, MO 65401 USA. E-mail: nancen@mst.edu.*
- *Zhangyu Guan is with the Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA. E-mail: guan@buffalo.edu.*
- *Tommaso Melodia is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA. E-mail: melodia@ece.neu.edu.*

Manuscript received 18 Feb. 2020; revised 7 Oct. 2020; accepted 22 Dec. 2020. Date of publication 8 Jan. 2021; date of current version 4 Aug. 2022. (Corresponding author: Nan Cen.)
Digital Object Identifier no. 10.1109/TMC.2021.3049797

important open problem. Very limited work has been reported in the literature to address this issue, especially for CS-based streaming system. For example, [14] and [15] have looked at this problem by considering traditional encoding paradigms, e.g., H.264 or MPEG4; these contributions focus on video transmission in single-hop wireless networks and provide a framework to improve power efficiency by adjusting encoding parameters such as quantization step (QS) size to adapt the resulting rate.

To the best of our knowledge, we propose for the first time a holistic paradigm of coding and transmitting for low-complexity low-power compressed-sensing-enabled multi-view video streaming in multi-hop wireless sensor networks. The objective is to efficiently deliver high-quality video on resource-limited video sensors. To achieve this objective, we first propose a novel CS-based multi-view coding and decoding architecture composed of cooperative encoders and independent decoders. Unlike existing works [10], [11], [12], the proposed system is based on independent encoding and independent decoding procedures with limited channel feedback information and negligible content sharing among camera sensors. Furthermore, we propose a power-efficient quality-guaranteed rate allocation algorithm based on a compressive Rate-Distortion (R-D) model for multi-view video streaming in multi-path multi-hop wireless sensor networks with lossy links. Our work makes the following contributions:

CS-based multi-view video coding architecture with independent encoders and independent decoders. Different from state-of-the-art multi-view coding architectures, that are either based on joint encoding or on joint decoding, we propose a new CS-based sparsity-aware independent encoding and decoding multi-view structure, that relies on lightweight feedback and inter-camera cooperation.

- *Sparsity Estimation.* We develop a novel adaptive approach to estimate block sparsity based on the reconstructed frame at the decoder. The estimated sparsity is then used to calculate the block-level measurement rate to be allocated with respect to a given frame-level rate. Next, the resulting block-level rates are transmitted back to the encoder through the feedback channel. The encoder that is selected to receive the feedback information, referred to as reference view (R-view), shares the content with other non-reference views (NR-views) nearby.

- *Block-Level Rate Adaptive Multi-View Encoders.* R-view and NR-views perform the block-level CS encoding independently based on the shared block-level measurement rate information. The objective is to not only implicitly leverage the considerable correlation among views, but also to adaptively balance the number of measurements among blocks with different sparsity levels. Our experimental results show that the proposed method outperforms state-of-the-art CS-based encoders with equal block-level measurement rate by up to 5 dB in terms of Peak Signal-to-Noise Ratio (PSNR).

Modeling framework for CS-based multi-view video streaming in multi-path multi-hop wireless sensor networks. We consider a rate-distortion model of the proposed streaming system that captures packet losses caused by unreliable links and playout deadline violations. Based on this model, we propose a two-fold (frame-level and path-level) rate control algorithm designed to minimize the network power consumption under constraints on the minimum required video quality for multi-path multi-hop multi-view video streaming scenarios.

The rest of the paper is organized as follows. In Section 2, we discuss related works. In Section 3, we review a few preliminary notions. In Section 4, we introduce the proposed CS-based multi-view video encoding/decoding architecture. In Section 5 we present a modeling framework to design optimization problems of multi-view streaming in multi-hop sensor networks and propose a solution algorithm. Finally, simulation results are presented in Section 6, while in Section 7 we draw the main conclusions and discuss future work.

2 Related Works

CS-Based Multi-View Video. More recently, several proposals have appeared for CS-based multi-view video coding based on Distributed Video Coding (DVC)[1] architecture [10], [13], [18], [19], [20], [21], [22], [23], [24]. In [10], a distributed multi-view video coding scheme based on CS is proposed, which assumes the same measurement rates for different views, and can only be applied together with specific structured dictionaries as sparse representation matrix. A linear operator [20] is proposed to describe the correlations between images of different views in the compressed domain. The authors then use it to develop a novel joint image reconstruction scheme. In [18], the authors propose a novel CS joint multi-view reconstruction method guided by the spatial correlation and low-rank background constraints. [19] presents a joint optimization model (JOM) for compressed sensing based multi-view image reconstruction, which jointly optimizes an adaptive disparity compensated residual total variation (ARTV) and a multi-image nonlocal low-rank tensor (MNLRT). The authors of [21] propose a CS-based joint reconstruction method for multi-view images, which uses two images from the two nearest views with higher measurement rate of the current image (the right and left neighbors) to calculate a prediction frame. The authors then further improve the performance by way of a multi-stage refinement procedure [22] via residual recovery. The readers are referred to [21], [22] and references therein for details. Disparity-based joint reconstruction for multi-view video is also proposed in [23] and [24], where different reconstruction methods, i.e., residual-based and total variation based approaches are adopted, respectively. In our previous work [13], we proposed a motion-aware joint multi-view video reconstruction method based on a newly designed interview motion compensated side information generation approach. Differently, in this article, we propose a novel CS-based *independently encoding and independently decoding* architecture for multi-view video systems based on new cooperative sparsity-aware-block-level rate adaptive encoders.

Energy-Efficient CS-Enabled Video Streaming. Few articles have investigated energy-constrained compressively-sampled video streaming. [25] presents a low-complexity and energy efficient image compressive transmission scheme for camera sensor networks, where the authors use residual energy of camera sensor nodes to control the image quality to balance energy consumption of nodes. In [9], an analytical/emperical

1. DVC algorithms (aka Wyner-Ziv coding [16], [17]) exploit the source statistics at the decoder, thus shifting the complexity from the encoder side to the decoder side.

rate-energy-distortion model is developed to predict the received video quality when the overall energy available for both encoding and transmission of each frame is fixed and limited and the transmissions are affected by channel errors. The model determines the optimal allocation of encoded video rate and channel coding rate for a given available energy budget. [26] proposes a cooperative relay-assisted compressed video sensing systems that takes advantage of the error resilience of compressively-sampled video to maintain good video quality at the receiver side while significantly reducing the required SNR, thus reducing the required transmission power. Different from the previous works, which mainly aims at single-view single path CS-based video streaming, in this article, we consider CS-based *multi-view video streaming in* multi-path multi-hop wireless sensor networks.

3 Preliminaries

3.1 Compressed Sensing Basics

We first briefly review basic concepts of CS for signal acquisition and recovery, especially as applied to CS-based video streaming. We consider an image signal vectorized and then represented as $\mathbf{x} \in \mathbb{R}^N$, where $N = H \times W$ is the number of pixels in the image, and H and W represent the dimensions of the captured scene. Each element x_i denotes the ith pixel in the vectorized image signal representation. Most natural images are known to be very nearly sparse when represented using some transformation basis $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$, e.g., Discrete Wavelet Transform (DWT) or Discrete Cosine Transform (DCT), denoted as $\mathbf{x} = \mathbf{\Psi s}$, where $\mathbf{s} \in \mathbb{R}^N$ is sparse representation of $\mathbf{x}$. If $\mathbf{s}$ has at most K nonzero components, we call $\mathbf{x}$ a K-sparse signal with respect to $\mathbf{\Psi}$.

In CS-based imaging system, sampling and compression are executed simultaneously through a linear measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$, with $M \ll N$, as

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi\Psi s}, \tag{1}$$

with $\mathbf{y} \in \mathbb{R}^M$ representing the resulting sampled and compressed vector.

It was proven in [3] that if $\mathbf{A} \triangleq \mathbf{\Phi\Psi}$ satisfies the following Restricted Isometry Property (RIP) of order K

$$(1-\delta_k)||\mathbf{s}||_{\ell_2}^2 \le ||\mathbf{As}||_{\ell_2}^2 \le (1+\delta_k)||\mathbf{s}||_{\ell_2}^2, \tag{2}$$

with $0 < \delta_k < 1$ being a small "isometry" constant and ℓ_2 denoting ℓ_2 norm, then we can recover the optimal sparse representation $\mathbf{s}^*$ of $\mathbf{x}$ by solving the following optimization problem

$$\begin{aligned} \mathrm{P}_1\text{:}\quad & \text{Minimize} && ||\mathbf{s}||_0 \\ & \text{Subject to:} && \mathbf{y} = \Phi\Psi\mathbf{s}, \end{aligned} \tag{3}$$

by taking only

$$M = c \cdot K\log(N/K), \tag{4}$$

measurements, where c is some predefined constant. Afterwards, $\mathbf{x}$ can be obtained by

$$\hat{\mathbf{x}} = \Psi\mathbf{s}^*. \tag{5}$$

However, problem P_1 is NP-hard in general, and in most practical cases, measurements $\mathbf{y}$ may be corrupted by noise, e.g., channel noise or quantization noise. Then, most state-of-the-art work relies on l_1 minimization with relaxed constraints in the form

$$\begin{aligned} \mathrm{P}_2\text{:}\quad & \text{Minimize} && ||\mathbf{s}||_1 \\ & \text{Subject to :} && ||\mathbf{y} - \Phi\Psi\mathbf{s}||_2 \le \epsilon, \end{aligned} \tag{6}$$

to recover $\mathbf{s}$. Note that P_2 is a convex optimization problem. Researchers in sparse signal reconstruction have developed various solvers [27], [28], [29]. For example, the Least Absolute Shrinkage and Selection Operator (LASSO) solver [28] can solve problem P_2 with computational complexity $\mathcal{O}(M^2N)$. We consider a Gaussian random measurement matrix $\mathbf{\Phi}$ in this paper.

3.2 Rate-Distortion Model for Compressive Imaging

Throughout this paper, end-to-end video distortion is measured as mean squared error (MSE), which is a widely used performance measure in the field of signal processing, especially for objective image quality measurement where the quality of images are measured algorithmically [30]. Since Peak Signal-to-Noise Ratio (PSNR) is a more common metric in the video coding community, we use $\mathrm{PSNR} = 10\log_{10}(255^2/\mathrm{MSE})$ to illustrate simulation results. The distortion at the decoder D_{dec} in general includes two terms, i.e., D_{enc}, distortion introduced by the encoder (e.g., not enough measurements and quantization); and D_{loss}, distortion caused by packet losses due to unreliable wireless links and violating playout deadlines because of bandwidth fluctuations. Therefore

$$D_{\mathrm{dec}} = f(D_{\mathrm{enc}}, D_{\mathrm{loss}}). \tag{7}$$

To the best of our knowledge, there are only a few works [8] that have investigated rate-distortion models for compressive video streaming, but without considering losses. For example, [8] expands the distortion model in [31] to CS video transmission as

$$D(R) = D_0 + \frac{\theta}{R - R_0}, \tag{8}$$

where D_0, θ and R_0 are image- or video-dependent constants that can be determined by linear least squares fitting techniques; $R = \frac{M}{N}$ is the user-controlled measurement rate of each video frame.

4 CS-Based Multi-View Coding Architecture Design

In this section, we introduce a novel encoding/decoding architecture design for CS multi-view video streaming. The proposed framework is based on three main components: (i) cooperative sparsity-aware block-level rate adaptive encoder, (ii) independent decoder, and (iii) a centralized controller located at the decoder. As illustrated in Fig. 1, considering a two-view example, camera sensors acquire a scene of interest with adaptive block-level rates and transmit sampled measurements to the base station/controller through a multi-path multi-hop wireless sensor network. Then, the centralized controller calculates the relevant information and feeds it back to the selected R-view. The R-view then shares the limited

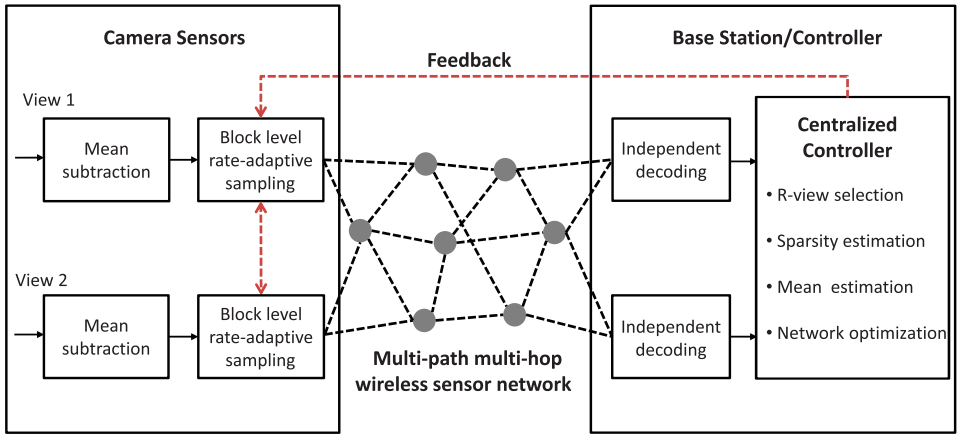


Fig. 1. Encoding/decoding architecture for multi-hop CS-based multi-view video streaming.

feedback information with the other one - NR-view. The architecture can be easily extended to $V \geq 2$ views.

Different from existing compressive encoders with equal block measurement rate [7], [8], the objective of the proposed framework is to improve the reconstruction quality by leveraging each block's sparsity as a guideline to adapt the block-level measurement rate. We next describe how to implement the proposed paradigm by discussing each component in detail.

4.1 Cooperative Block-Level Rate-Adaptive Encoder

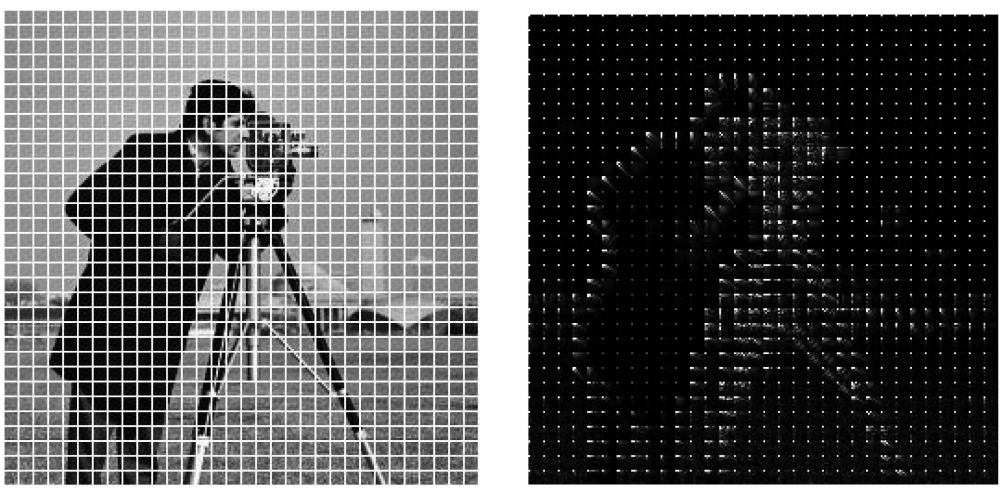
To reduce the computational burden at encoders embedded in power-constrained devices, most state-of-the-art multi-view proposals focus on developing complex joint reconstruction algorithms to improve the reconstruction quality. Differently, in our architecture we obtain improved quality only through sparsity-aware encoders.

To illustrate the idea, Fig. 2b depicts the sparse representation of Fig. 2a with respect to block-based DCT transformation. We can observe that sparsity differs among blocks, e.g., the blocks within the coat area are more sparse than others. According to basic compressed sensing theory in Section 3.1, (4) indicates that the number of required measurements is inversely proportional to the sparsity K. Therefore, we propose to adapt the measurement rate at the block level according to sparsity information, i.e., more measurements will be allocated to less-sparse blocks, and vice versa.

In our work, the number of required measurements M_{vf}^i for block i in frame f of view v, $1 \leq i \leq B$, is calculated based on the sparsity estimated at the centralized controller and sent back via a feedback channel. Here, $B = \frac{N}{N_b}$ denotes the total number of blocks in one frame with N and N_b being the total number of pixels in one frame and block, respectively. Assume that we have received $\{M_{vf}^i\}_{i=1}^B$. Then, the encoding process is similar to (1), described as

$$\mathbf{y}_{vf}^i = \mathbf{\Phi}_{vf}^i \mathbf{x}_{vf}^i, \tag{9}$$

where $\mathbf{y}_{vf}^i \in \mathbb{R}^{M_{vf}^i}$ and $\mathbf{\Phi}_{vf}^i \in \mathbb{R}^{M_{vf}^i \times N_b}$ are the measurement vector and measurement matrix for block i in frame f of view v, respectively; $\mathbf{x}_{vf}^i \in \mathbb{R}^{N_b}$ represents the original pixel vector of block i. From (9), we can see that M_{vf}^i varies among blocks from 1 to N_b, thereby implementing block-level rate adaptation. In real applications, the block-level rate adaptive encoding process can be implemented by using block-wise lensless compressive camera [32]. In Section 6, the simulation results will show that this approach can improve the quality by up to



(a) (b)

Fig. 2. Block sparsity: (a) Original image, (b) Block-based DCT coefficients of (a).

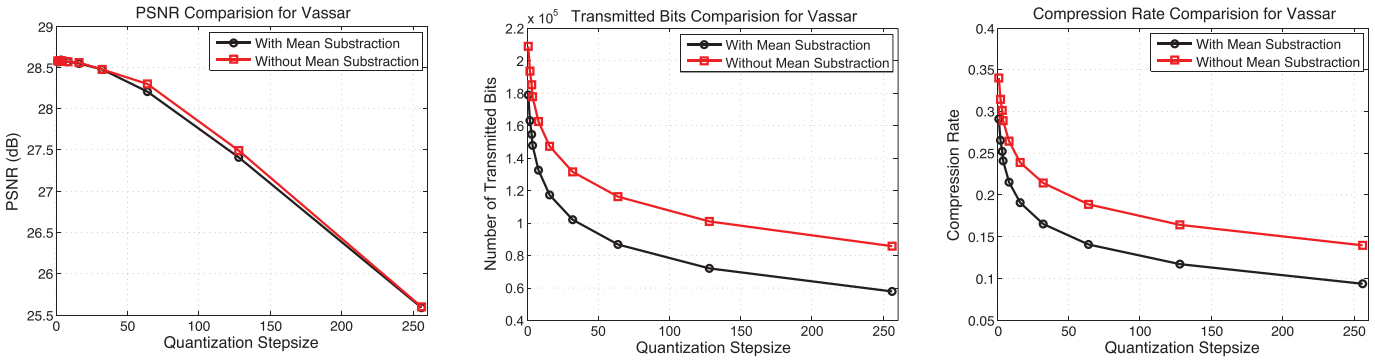


Fig. 3. Comparison of (a) PSNR, (b) the number of transmitted bits, and (c) the compression rate between approaches with and without mean subtraction.

5 dB compared with using an independent encoder and independent decoder.

Mean Value Subtraction. The CS-based imaging system acquires and compresses each frame simultaneously through simple linear operations as in (1). Therefore, it can help reduce the energy consumption compared with traditional signal acquisition and encoding approaches (e.g., H.264, MJPEG) as validated in [33].In traditional video coding techniques, frequency transform (e.g., 8×8 block of DCT transform) is one must-have component, while inter-view prediction (e.g., disparity estimation) and inter-frame prediction (e.g., motion estimation) are additions, which will all together result in higher computation complexity compared to linear operations in CS-based imaging systems, thus consuming more power. In this paper, we exploit the inter-view correlation by using the estimated sparsity from R-view feedbacked from the receiver. However, the compression rate of CS is not as high as traditional encoding schemes [9]. There is clearly an energy-consumption trade-off between the compression rate and the bit transmission rate. [9] analyzes the rate-energy-distortion for compressive video sensing encoder. To improve the compression rate, we perform *mean value subtraction*, which can further help reduce the number of transmitted bits. How to obtain the mean value $\bar{m}$ will be discussed in Section 4.3. Since the original pixels are not available at the compressive encoder, we perform the *mean value subtraction in the measurement domain*. First, we establish a mean value vector $\mathbf{m} \in \mathbb{R}^{N_b}$ with dimensions the same as $\mathbf{x}_{vf}^i$, and where each element is equal to $\bar{m}$. Then, we use the same block-level measurement matrix $\mathbf{\Phi}_{vf}^i$ to sample $\mathbf{m}$ and then subtract the result from $\mathbf{y}_{vf}^i$ as

$$\tilde{\mathbf{y}}_{vf}^i = \mathbf{y}_{vf}^i - \mathbf{\Phi}_{vf}^i \mathbf{m} = \mathbf{\Phi}_{vf}^i (\mathbf{x}_{vf}^i - \mathbf{m}). \tag{10}$$

After sampling, $\tilde{\mathbf{y}}_{vf}^i$ is transmitted to the decoder. From (10), we can see that the proposed mean value subtraction in the measurement domain is equivalent to subtraction in the pixel domain.

Next, to validate the effectiveness of mean value subtraction, we take the *Vassar* sequence as an example. We select a uniform quantization method. The forward quantization stage and the reconstruction stage can be expressed as $q = \text{sgn}(x) \cdot \lfloor \frac{|x|}{\Delta} + \frac{1}{2} \rfloor$ and $\hat{q} = \Delta \cdot q$, respectively. Here, x, q, $\hat{q}$ and Δ represent original signal, quantized signal, de-quantized signal and quantization step size, respectively. Fig. 3 shows a comparison of PSNR, the number of transmitted bits and the compression rate with and without mean subtraction, where a measurement rate 0.2 is used, and the total bits in the original frame are $320 \times 240 \times 8 = 614400$ bits. Quantization step sizes from the set {1, 2, 3, 4, 8, 16, 32, 64, 128, 256} are selected. From Fig. 3a, we can observe that mean subtraction has a negligible effect on the reconstruction quality and there is no significant quality degradation when the quantization step size is less than 32. This is because the value of measurement is up to thousand and tens of thousand compared to original pixel value with maximum 255. Figs. 3b and 3c illustrate that with mean subtraction the total number of bits transmitted for one frame is significantly reduced by up to 30 kbits compared to not using mean subtraction, which corresponds to an improvement in compression rate (defined as the ratio between the number of transmitted bits and the number of total bits in the original frame) from 0.2391 to 0.1902. Besides mean value subtraction, we can explore the temporal correlation [9] between successive frames in our future work to further improve the performance with respect to compression ratio.

Cooperation via Sparsity Pattern Sharing. Multi-view video streaming is based on reducing the redundancy among views captured by arrays of camera sensors that are assumed to be close enough to each other. Most state-of-the-art literature adopts the concept of distributed system coding architecture [16], [17], where a reference view transmits more measurements than other non-reference views and then the receiver jointly decodes by exploiting the implicit correlation among views. Instead, we allow the encoders to explicitly cooperate to a certain extent. For example, the R-view selected by the centralized controller will periodically receive feedback information, i.e., $\{M_i\}_{i=1}^B$ and $\bar{m}$, and then share it with the NR-views in the same group. Since camera sensors in the same group are assumed to be close enough to each other, the block sparsity among views will be correlated. By using the same sparsity information, we can directly exploit multi-view correlation at the encoders, thus resulting in a clean-slate compressive multi-view coding framework with simple encoders and simple decoders but with improved reconstruction quality.

4.2 Independent Decoder

As mentioned above, the proposed framework results in relatively simple decoders. At each decoder, the received $\hat{\mathbf{y}}_{vf}^i$, distorted version of $\tilde{\mathbf{y}}_{vf}^i$ because of the joint effects of

quantization, transmission errors, and packet drops, will be independently decoded. The optimal solution $\mathbf{s}_{vf}^{i,\star}$ can be obtained by solving

$$\begin{aligned} \mathrm{P}_3: \quad & \text{Minimize} \quad ||\mathbf{s}_{vf}^i||_1 \\ & \text{Subject to:} \quad ||\hat{\mathbf{y}}_{vf}^i - \mathbf{\Phi}_{\mathbf{vf}}^{\mathbf{i}} \mathbf{\Psi}_b \mathbf{s}_{vf}^i||_2 \leq \epsilon, \end{aligned} \tag{11}$$

where $\mathbf{\Psi}_b \in \mathbb{R}^{N_b \times N_b}$ represents the sparsifying matrix (2-D DCT in this work). We then use (5) to obtain the reconstructed block-level image $\hat{\mathbf{x}}_{vf}^i$, by solving $\hat{\mathbf{x}}_{vf}^i = \mathbf{\Psi}_b \mathbf{s}_{vf}^{i,\star}$. Afterward, $\{\hat{\mathbf{x}}_{vf}^i\}_{i=1}^B$ can be simply reorganized to obtain the reconstructed frame $\hat{\mathbf{x}}_{vf}$.

4.3 Centralized Controller

The centralized controller is the key component at the receiver, which is mainly in charge of selecting the R-view and estimating sparsity and mean value required to be sent back to the transmitter via a assumed delay-negligible [34], [35] and error-free feedback link.[2] How to implement a fast feedback channel in practical scenarios is important and feasible, which can help further improve the performance of the proposed framework and is currently beyond the scope of the paper. Additionally, the controller is also responsible for implementing the power-efficient multi-path rate allocation algorithm discussed in Section 5. Next, we introduce the three key functions executed at the controller in sequence, i.e., *R-view selection, sparsity estimation, and mean value estimation*.

R-View Selection. The controller selects a view to be used as reference view (R-view) among views in the same group and then sends feedback information to the selected R-view. For this purpose, the controller first calculates the Pearson correlation coefficient among the measurement vectors of any two views as

$$\rho_{mn} = \mathrm{corr}(\hat{\mathbf{y}}_{mf}, \hat{\mathbf{y}}_{nf}), \ \forall m \neq n, \ m, n = 1, \ldots, V, \tag{12}$$

where $\hat{\mathbf{y}}_{mf}$ is the simple cascaded version of all $\hat{\mathbf{y}}_{mf}^i$ and $\mathrm{corr}(\hat{\mathbf{y}}_{mf}, \hat{\mathbf{y}}_{nf}) \triangleq \frac{\mathrm{COV}(\hat{\mathbf{y}}_{mf}, \hat{\mathbf{y}}_{nf})}{\sigma_{mf}\sigma_{nf}}$. Then, view $m^\star$, referred to as R-view, is selected by solving

$$m^\star = \underset{m=1,\ldots,V}{\mathrm{argmax}} \, \tilde{\rho}_m, \tag{13}$$

where $\tilde{\rho}_m \triangleq \frac{1}{V-1} \sum_{n \neq m} \rho_{mn}$ denotes the average Pearson coefficient for view m. From (13), we can see that the view with the highest average Pearson coefficient is selected as R-view.[3] The reconstructed frame $\hat{\mathbf{x}}_{vf}$ of the R-view is then used to estimate the block sparsity K^i and the frame mean value $\bar{m}$ for block i.

Table 1 shows the calculated $\tilde{\rho}_m$ for *Vassar*, *Exit* and *Ballroom* 5-view sequences with lower resolution (i.e., 320×240, represented as L) and higher resolution (i.e., 640×480, represented as H), respectively. We can see that the average Pearson correlation coefficient of view 3 is the largest for all scenarios while the correlation degree decreases as resolution increases. Therefore, view 3 is selected as R-view. Moreover, we take the *Vassar* 5-view sequences as an example to elaborate how much quality gain we can obtain if the other views except view 3 are selected as R-view with respect to lower resolution and higher resolution, respectively, as shown in Table 2. We can observe that the improved average PSNR is proportional to $\tilde{\rho_m}$, where selecting view 3 as R-view results in the highest improved average PSNR gain, i.e., 1.6674 and 1.3132 dB for lower and higher resolution scenarios, respectively. We can also see that the quality gain slightly decreases for higher resolution *Vassar* sequences because of the decreased correlation degree. For this case, because the *Vassar* multi-view sequences used here is captured by parallel-deployed cameras with equal spacing, we obtain the same result, i.e., view 3 as R-view, as if we were to choose simply the most central sensor. However, for scenarios with cameras that are not parallel-deployed with unequal spacing, selecting the most central sensor is not necessarily a good choice.

TABLE 1
Average Pearson Correlation Coefficient $\tilde{\rho}_m$ for *Vassar*, *Exit* and *Ballroom* Five Views

	View 1	View 2	View 3	View 4	View 5
Vassar-L	0.8184	0.8988	0.9243	0.8973	0.8435
Vassar-H	0.7655	0.8464	0.8815	0.8551	0.7920
Exit-L	0.5703	0.6787	0.7038	0.6838	0.5078
Exit-H	0.5358	0.6281	0.6643	0.6315	0.4745
Ballroom-L	0.8366	0.8713	0.8812	0.8627	0.8135
Ballroom-H	0.7574	0.7961	0.8099	0.7922	0.7484

TABLE 2
Improved Average PSNR (dB) When Selecting Different *Vassar* Views as R-View

R-view	View 1	View 2	View 3	View 4	View 5
Vassar-L	1.2312	1.6241	1.6674	1.6167	1.3833
Vassar-H	0.6686	0.8865	1.3132	1.2138	0.9019

Sparsity Estimation. Most natural images are characterized by large smooth or textured regions and relatively few sharp edges. Signals with this structure are known to be very nearly sparse when represented using DWT or DCT domain [38], where lowest frequency components provide a coarse scale approximation of the image, while the higher frequency components fill in the detail and resolve edges. Moreover, most DWT or DCT coefficients are very small. Hence, we can obtain a good approximation of the signal by setting the small coefficients to zero, or thresholding the coefficients, to obtain k-sparse representation. Moreover, in CS-based imaging system, the original frame in the pixel domain is not available, therefore, we propose to estimate sparsity based on the reconstructed frame $\hat{\mathbf{x}}_{vf}$ as follows. By solving the optimization problem P_3 in (11), we can obtain the block sparse representation $\mathbf{s}_{vf}^{i,\star}$ and then reorganize $\{\mathbf{s}_{vf}^{i,\star}\}_{i=1}^B$ to get the frame sparse representation $\mathbf{s}_{vf}^\star$ periodically. The sparsity coefficient K^i is defined as the number of non-zero entries of $\mathbf{s}_{vf}^\star$. However, natural pictures in general are not exactly sparse in the transform domain. Hence, we

2. Since we mainly consider a slow block-fading environment, the feedback delay is significantly less than the coherence time of the fading channels concerned. Thus, the effect of feedback delay can be negligible. Moreover, with a small data rate, efficient error control coding techniques over feedback link can be used to achieve error-free feedback [36], [37].

3. The adopted Pearson correlation just considers the linear relation among views. We believe that more advanced correlation algorithms which also consider the features in the correlation will result in more accurate R-view selection and better performance gain based on our proposed CS-based multi-view coding/decoding architecture.

introduce a predefined percentile p_s,[4] and assume that the frame can be perfectly recovered with $N \cdot p_s$ measurements. Based on this, one can adaptively find a threshold T above which transform-domain coefficients are considered as non-zero entries. The threshold can be found by solving

$$\frac{||\max(|s^{\star}_{vf}| - T, 0)||_0}{N} = p_s, \tag{14}$$

which is a ℓ_0 counting norm problem. Since the sample space of the above-mentioned problem is small and limited, we employ an exhaustive search approach to solve it. Then, we apply T to each block i to estimate the block sparsity K^i as

$$K^i = ||\max(|s^{i,\star}_{vf}| - T, 0)||_0. \tag{15}$$

According to (4) and given the frame measurement rate R, M^i_{vf} can then be obtained as

$$M^i_{vf} = \frac{K^i \log_{10}(\frac{N_b}{K_i})}{\sum_{i=1}^{B} K^i \log_{10}(\frac{N_b}{K_i})} NR. \tag{16}$$

Mean Value Estimation. Finally, the mean value $\bar{m}$ can be estimated from $\hat{\mathbf{x}}_{vf}$ as

$$\bar{m} = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{x}}_{vf}(i). \tag{17}$$

With limited feedback and lightweight information sharing, implementing block-level rate adaptation at the encoder without adding computational complexity can improve the reconstruction performance of our proposed encoding/decod-ing paradigm. This claim will be validated in Section 6 in terms of Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity (SSIM) [39].

5 Network Modeling Framework

We consider compressive wireless video streaming over multi-path multi-hop WMSNs. We first formulate a video-quality-assured power minimization problem, and then solve the resulting nonlinear nonconvex optimization problem by proposing an online solution algorithm with low computational complexity.

Network Model. In the considered WMSN there are a set $\mathcal{V}$ of camera sensors at the transmitter side, with each camera capturing a video sequence of the same scene of interest, and then sending the sequence to the server side through a set $\mathcal{Z}$ of pre-established multi-hop paths. Denote $\mathcal{L}^z$ as the set of hops belonging to path $z \in \mathcal{Z}$, with $d^{z,l}$ being the hop distance of the lth hop in $\mathcal{L}^z$. Let $V = |\mathcal{V}|$, $Z = |\mathcal{Z}|$, and $L^z = |\mathcal{L}^z|$ represent cardinality of sets $\mathcal{V}$, $\mathcal{Z}$ and $\mathcal{L}^z$, respectively. The following three assumptions are considered:

- *Pre-established routing*, i.e., the set of multi-hop paths $\mathcal{Z}$ is established in advance through a given routing protocol (e.g., AODV [40]) and does not change during the video streaming session.
- *Orthogonal channel access*, i.e., there exists a pre-established orthogonal channel access, e.g., based on TDMA, FDMA, or CDMA, and hence concurrent transmissions do not interfere with each other [41].
- *Time division duplexing*, i.e., each node cannot transmit and receive simultaneously, implying that only half of the total air-time is used for transmission or reception.

At the receiver side, the video server concurrently and independently decodes each view of the received video sequences, and based on the reconstructed video sequences it then computes the rate control information and sends the information back to camera sensors for actual rate control. For this purpose, we define two types of video frames, Reference Frame (referred to as *R-frame*) and Non-Reference Frame (referred to as *NR-frame*). A R-frame is only periodically transmitted by the R-view, which is encoded at fixed block-level rate and used for sparsity and mean value estimation in the centralized controller. All other frames sent out by the R-view and all frames transmitted by the NR-views are categorized as NR-frames, which are encoded at adaptive block-level rate based on the estimated sparsity. Compared to an NR-frame, an R-frame is encoded with equal or higher sampling rate and then sent to the receiver side with much lower transmission delay. Hence, an R-frame can be reconstructed with equal or higher video quality and used to estimate sparsity pattern information, which is then fed back to video cameras for rate control in encoding the following NR-frames. For the R-view, we consider a periodic frame pattern, meaning that the R-view camera encodes its captured video frames as R-frames periodically, e.g., one every 30 consecutive frames.

In the above setting, our objective is to minimize the average power consumption of all cameras and communication sensors in the network with guaranteed reconstructed video quality for each view, by jointly controlling video encoding rate and allocating the rate among candidate paths.

To handle CS-based multi-view video streaming with guaranteed quality, a rate-distortion model to measure the end-to-end distortion that jointly captures the effects of encoder distortion and transmission distortion as stated in (7) is needed. To this end, we modify the R-D model (8) proposed in [8] by adding a packet loss term to jointly account for compression loss and packet loss[5] in compressive wireless video streaming systems, described as

$$D_{\text{dec}} = f(D_{\text{enc}}, D_{\text{loss}}) = D_0 - \frac{\theta}{R - \kappa p_{\text{loss}} - R_0}. \tag{18}$$

The parameters D_0, θ, and R_0 can be estimated from empirical rate-distortion curves via a linear least squared curve fitting [42]. Next, we describe how to derive them in compressive multi-view streaming systems.

Since the original pixel values are not available at the receiver end and even not available at the transmitter side in compressive

4. p_s represents the number of the largest original coefficients that are kept for video reconstruction. We can set it to an empirical well-performing value, e.g., 15 percent, and then slightly tune it during the video streaming. If the estimated sparsity is too small and cannot result in good reconstruction quality, then we can gradually increase p_s till we reach a satisfied point.

5. Different from traditional predictive-encoding based imaging systems, each packet in CS-based imaging systems has the same importance, i.e., it contributes equally to the reconstruction quality. Therefore, the packet loss probability can be converted into a measurement rate reduction through a conversion parameter and considered into the rate-distortion performance.

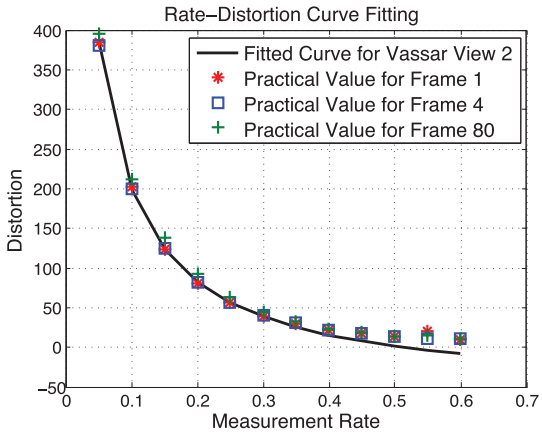


Fig. 4. Rate-distortion curve fitting for *Vassar* view 2 sequence.

multi-view streaming systems, we let the R-view periodically transmit a frame at a higher measurement rate, e.g., 60 percent.[6] In this way, after reconstruction at the decoder side, the reconstructed frame is considered as the original image in the pixel domain. We then resample it at different measurement rates and perform the reconstruction procedure to obtain several rate-distortion sample pairs which are then used to complete the linear least squared curve fitting to obtain. Fig. 4 illustrates the effectiveness of the above-mentioned online rate-distortion estimation approach, where *Vassar* view 2 sequence is used. We can observe that the fitted rate-distortion curve (depicted in black solid line) matches well the ground-truth distortion values (depicted in red pentagrams, blue squares and green pluses for frames 1, 4 and 80, respectively).

Besides determining the fitting parameters in (18), next we derive the packet loss probability p_{loss} and packet loss rate to measurement rate reduction converter κ in (18) to formalize the network optimization problem.

Packet Loss Probability. According to the proposed modified R-D model (18), packet losses affect the video reconstruction quality because they introduce an effective measurement rate reduction. Therefore, effective estimation of packet loss probability at the receiver side has significant impact on frame-level measurement rate control.

In real-time wireless video streaming systems, a video packet can be lost primarily for two reasons: i) the packet fails to pass a parity check due to transmission errors introduced by unreliable wireless links, and ii) it takes too long for the packet to arrive at the receiver side, hence violating the maximum playout delay constraint. Denoting the corresponding packet loss probability as p_{per} and p_{dly}, respectively, the total packet loss rate p_{loss} can then be written as

$$p_{\text{loss}} = p_{\text{per}} + p_{\text{dly}}. \tag{19}$$

In the case of multi-path routing as considered above, p_{per} and p_{dly} in (19) can be further expressed as

$$p_{\text{per}} = \sum_{z\in\mathcal{Z}} \frac{b^z}{b} p^z_{\text{per}}, \tag{20}$$

$$p_{\text{dly}} = \sum_{z\in\mathcal{Z}} \frac{b^z}{b} p^z_{\text{dly}}, \tag{21}$$

where p^z_{per} and p^z_{dly} represent the packet loss rate for path $z \in \mathcal{Z}$ due to transmission error and delay constraint violation, respectively; b and b^z represent total video rate and the rate allocated to path $z \in \mathcal{Z}$, respectively.

Since each path $z \in \mathcal{Z}$ may have one or multiple hops, to derive the expressions for p^z_{per} and p^z_{dly} in (20) and (21), we need to derive the resulting packet error rate and delay violation probability at each hop l of path $z \in \mathcal{Z}$, denoted as $p^{z,l}_{\text{per}}$ and $p^{z,l}_{\text{dly}}$, respectively. For this purpose, we first express the feasible transmission rate achievable at each hop. For each hop $l \in \mathcal{L}^z$ along path $z \in \mathcal{Z}$, let $G^{z,l}$ and $N^{z,l}$ represent the channel gain that accounts for both path loss and fading, and the additive white Gaussian noise (AWGN) power currently measured by hop l, respectively. Denoting $P^{z,l}$ as the transmission power of the sender of hop l, then the attainable transmission rate for the hop, denoted by $C^{z,l}(P^{z,l})$, can be expressed as [43]

$$C^{z,l}(P^{z,l}) = \frac{W}{2}\log_2\left(1 + K\frac{P^{z,l}G^{z,l}}{N^{z,l}}\right), \tag{22}$$

where W is channel bandwidth in Hz, calibration factor K is defined as

$$K = \frac{-\phi_1}{\log(\phi_2 p_{\text{ber}})}, \tag{23}$$

with ϕ_1, ϕ_2 being constants depending on available set of channel coding and modulation schemes, and p_{ber} is the predefined maximum residual bit error rate (BER). Then, if path $z \in \mathcal{Z}$ is allocated video rate b^z, for each hop $l \in \mathcal{L}^z$, the average attainable transmission rate should be equal to or higher than b^z, i.e.,

$$\mathrm{E}[C^{z,l}(P^{z,l})] \geq b^z, \tag{24}$$

with $\mathrm{E}[C^{z,l}(P^{z,l})]$ defined by averaging $C^{z,l}(P^{z,l})$ over all possible channel gains $G^{z,l}$ in (22).

Based on the above setting, we can now express the single hop packet error rate $p^{z,l}_{\text{per}}$ for each hop $l \in \mathcal{L}^z$ of path $z \in \mathcal{Z}$ as

$$p^{z,l}_{\text{per}} = 1 - (1 - p_{\text{ber}})^L, \tag{25}$$

where L is the predefined packet length in bits. Further, we characterize the queueing behavior at each wireless hop as in [44] using a M/M/1 model to capture the effects of channel-state-dependent transmission rate (22) single-hop queueing delay. Denoting $T^{z,l}$ as the delay budget tolerable at each hop $l \in \mathcal{L}^z$ of path $z \in \mathcal{Z}$, the resulting packet drop rate due to delay constraint violation can then be given as [45]

$$p^{z,l}_{\text{dly}} = e^{-(\mathrm{E}[C^{z,l}(P^{z,l})]-b^z)\frac{T^{z,l}}{L}}, \tag{26}$$

with $\mathrm{E}[C^{z,l}(P^{z,l})]$ defined in (24). For each path $z \in \mathcal{Z}$, the maximum tolerable end-to-end delay $T^{\max}$ can be assigned to each hop in different ways, e.g., equal assignment or distance-proportional assignment [46]. We adopt the same delay budget assignment scheme as in [46].

6. Based on CS theory [38], image reconstructed by using 60 percent measurement can result in basically the original image, i.e., the differences between the reconstructed image and the original image cannot be perceived by human eyes.

Finally, given $p_{\text{per}}^{z,l}$ and $p_{\text{dly}}^{z,l}$ in (25) and (26), we can express the end-to-end packet error rate p_{per}^{z} and delay violation probability p_{dly}^{z} in (20) and (21) as, for each path $z \in \mathcal{Z}$

$$p_{\text{per}}^{z} = \sum_{l \in \mathcal{L}^z} p_{\text{per}}^{z,l}, \ \forall z \in \mathcal{Z}, \tag{27}$$

$$p_{\text{dly}}^{z} = \sum_{l \in \mathcal{L}^z} p_{\text{dly}}^{z,l}, \ \forall z \in \mathcal{Z}, \tag{28}$$

by neglecting the second and higher order product of $p_{\text{per}}^{z,l}$ and of $p_{\text{dly}}^{z,l}$. The resulting p_{per}^{z} and p_{dly}^{z} provide an upper bound on the real end-to-end packet error rate and delay constraint violation probability. The approximation error is negligible if packet loss rate at each wireless hop is low or moderate. Note that it is also possible to derive a lower bound on the end-to-end packet loss rate, e.g., by applying the Chernoff Bound [47].

Packet Loss to Measurement Rate. After having modeled p_{loss}, we now concentrate on determining κ to convert p_{loss} to measurement rate reduction (referred to as $R_d = \kappa \cdot p_{\text{loss}}$). First, parameter $\tau = \frac{1}{QN}$ is defined to convert the amount of transmitted bits of each frame to its measurement rate R used in the (18), with Q being the bit-depth for each measurement. We assume that b is equally distributed among F frames within 1 second for all V views, i.e., the transmitted bits for each frame is $b/F/V$. Thus, measurement rate R for each frame of each view is equal and defined as $R = \tau b/F/V$. Then, we can define κ as

$$\kappa = \tau L \lceil \frac{b/F/V}{L} \rceil, \tag{29}$$

and rewrite (18) as

$$D_{\text{dec}} = D_0 - \frac{\theta}{\tau b/F/V - \kappa p_{\text{loss}} - R_0}. \tag{30}$$

Problem Formulation. Based on (30), we formulate, as an example of applicability of the proposed framework, the problem of power consumption minimization for quality-assured compressive multi-view video streaming over multi-hop wireless sensor networks, by jointly determining the optimal frame-level encoding rate and allocating transmission rate among multiple paths, i.e.,

$$\text{P}_4: \underset{P^{z,l}, b^z, l \in \mathcal{L}^z, \forall z \in \mathcal{Z}}{\text{Minimize}} \sum_{z \in \mathcal{Z}} \sum_{l \in \mathcal{L}^z} P^{z,l} \tag{31}$$

$$\text{Subject to:} b = \sum_{z \in \mathcal{Z}} b^z \tag{32}$$

$$D_{\text{dec}} \leq D_{\text{t}} \tag{33}$$

$$0 < \tau b/F/V - \kappa p_{\text{loss}} \leq 1 \tag{34}$$

$$0 \leq P^{z,l} \leq P_{\max}, \ \forall l \in \mathcal{L}^z, z \in \mathcal{Z}, \tag{35}$$

where D_{t} and $P_{\max}$ represent the constraints upon distortion and power consumption, respectively. Here, (33) and (34) are the constraints for required video quality level and total measurement rate not lower than 0 and higher than 1, respectively. In fact, the optimization problem P_4 is non-convex because the distortion constraint is non-convex. Solving it directly will be computationally expensive due to the large space of b. Therefore, in the following, we design a solution algorithm to find the solution to the problem in real time.

Solution Algorithm. The core idea of the solution algorithm is to iteratively control video encoding and transmission strategies at two levels, i.e., adjusting video encoding rate for each frame (*frame level*) and allocating the resulting video data rate among different paths (*path level*). In each iteration, the algorithm first determines at the frame level the minimum video encoding rate required to achieve predefined reconstructed video quality, i.e., b in (33); and then determines at the path level the optimal transmission rate strategy with minimal power consumption, i.e., b^z for each path $z \in \mathcal{Z}$.

At the frame level, given the current total video encoding rate b and assigned rate b^z for each path $z \in \mathcal{Z}$, the algorithm estimates the video construction distortion D_{dec} based on (19)-(30). Then, if the video quality constraint in optimization problem P_4 can be strictly satisfied, i.e., the inequality holds in (33), it means that power consumption can be further reduced by reducing the total video encoding rate b, e.g., by a predefined step Δb, while keeping the distortion constraint (33) still satisfied. Otherwise, if constraint (33) is violated, we need to reduce reconstructed video D_{dec} by increasing the video encoding rate b hence transmission power. Whenever there are changes with the total encoding rate b, it triggers at the path level rate allocation among different paths. For example, if b is increased by Δb, the increased amount of video data rate is allocated to the path that results in minimum increase of power consumption, and vice versa.

As the above procedure goes on, the resulting video distortion D_{dec} is maintained fluctuating around, ideally equal to, the predefined maximum tolerable distortion $D_{\max}$. Hence, we approximately solve the optimization problem P4 formulated in (31)-(35), and the resulting power consumption provides an *upper bound* on the real minimum required total power. Next, in Section 6 we validate the effectiveness of the proposed solution algorithm through extensive simulation results.

6 Performance Evaluation

The topology includes a certain number V camera sensors and pre-established paths with random number of hops between camera sensors and the receiver. The frame rate is $F = 30$ fps, and the R-view periodically sends the R-frame every second. At the sparsity-aware CS independent encoder side, each frame is partitioned into 16×16 non-overlapped blocks implying $N_d = 256$. A measurement matrix $\mathbf{\Phi}_{vf}^{i}$ with elements drawn from independent and identically distributed (i.i.d) Gaussian random variables is considered, where the random seed is fixed for all experiments to make sure that $\mathbf{\Phi}_{vf}^{i}$ is drawn from the same matrix. The elements of the measurement vector $\tilde{\mathbf{y}}_{vf}^{i}$ are quantized individually by an 8-bit uniform scalar quantizer and then transmitted to the decoder. At the independent decoder end, we use $\mathbf{\Psi}_b$ composed of DCT transform basis as sparsifying matrix and choose the LASSO algorithm for reconstruction motivated by its low-complexity and excellent recovery performance characteristics. We consider four test multi-view sequences, *Vassar*, *Exit*, *Ballroom*, and *Balloons*, which are made publicly available [48], [49] and represent scenarios with slow, moderate and fast movement characteristics,

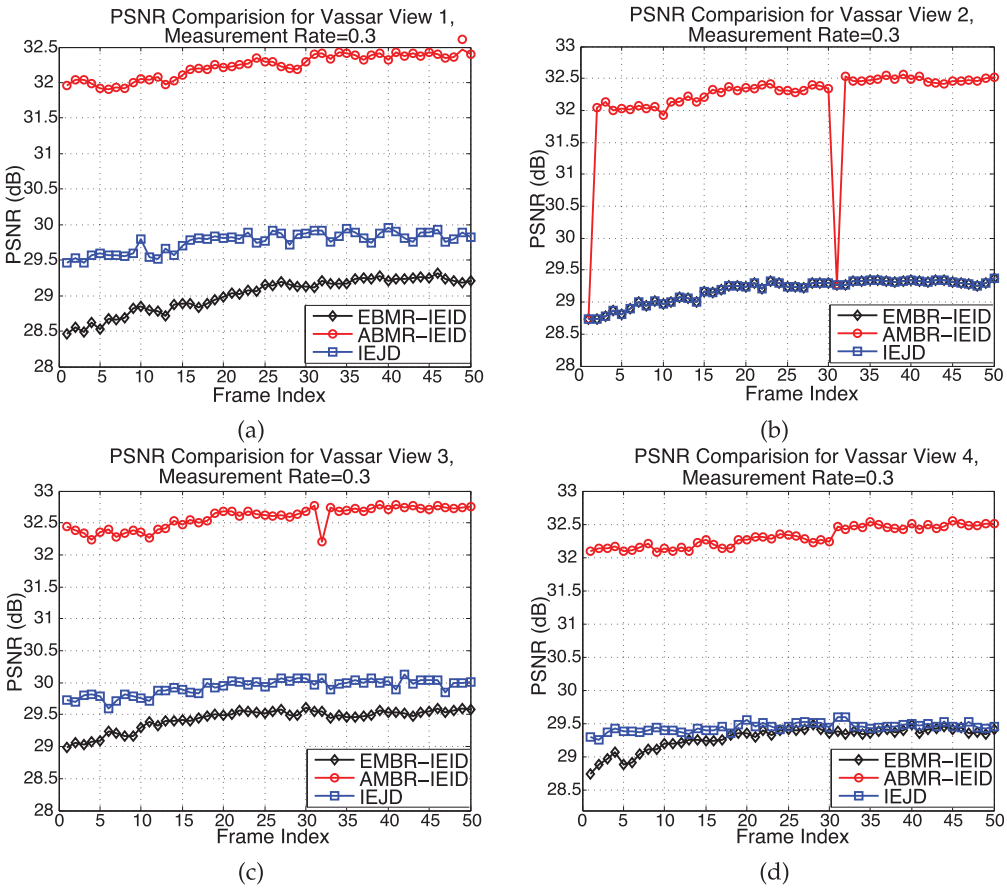


Fig. 5. PSNR against frame index for (a) view 1, (b) view 2 (R-view), (c) view 3, and (d) view 4 of sequence *Vassar*.

respectively. In the sequences considered, the optical axis of each camera is parallel to the ground, and each camera is 19.5 cm away from its left and right neighbors. There spatial resolutions of $(H = 240) \times (W = 320)$, $(H = 480) \times (W = 640)$, and $(H = 768) \times (W = 1024)$ (in pixel) are considered.

6.1 Evaluation of CS-Based Multi-View Encoding/ Decoding Architecture

We first experimentally study the performance of the proposed CS-based multi-view encoding/decoding architecture by evaluating the PSNR of the reconstructed video sequences. Experiments are carried out only on the luminance component. Next, we discuss a performance comparison among (i) traditional Equal-Block-Measurement-Rate Independent Encoding and Independent Decoding approach (referred to as EBMR-IEID), (ii) the proposed sparsity-aware Adaptive-Block-Measurement-Rate Independent Encoding and Independent Decoding approach (referred to as ABMR-IEID) and (iii) Independent Encoding and Joint Decoding (referred to as IEJD) proposed in [12] which selects one view as reference view reconstructed by traditional CS recovery method, while other views are jointly reconstructed by using a reference frame.

Figs. 5 and 6 show the PSNR comparisons of 50 frames for views 1, 2, 3 and 4 of *Vassar* and *Exit* multi-view sequences, where a 0.3 measurement rate for each view of ABMR-IEID[7] and EBMR-IEID is selected. To assure fair comparison, the measurement rate of each view in IEJD is also set to 0.3. Besides, according to the R-view selection algorithm, view 2 is chosen as the R-view for this scenario. Since the R-view transmits the R-frame periodically which is not encoded based on sparsity pattern at the encoder, therefore we can observe drops occurred periodically in Figs. 5b and 6b. For the *Vassar* sequences, as illustrated in Fig. 5, we can see that the proposed method ABMR-IEID outperforms the traditional approach EBMR-IEID and IEJD by up to 3.5 and 2.5 dB in terms of PSNR, respectively. For the *Exit* sequences, Fig. 6 shows improvement in the reconstruction quality of ABMR-IEID compared with EBMR-IEID and IEJD fluctuates more than that of *Vassar* video, with increased PSNR varying from 5 to 2 dB and from 4 to 1 dB, respectively. This phenomenon occurs because of the video-based features, i.e., the texture of *Exit* changes faster than in *Vassar*. In other words, the proposed scheme is more robust in surveillance scenarios where the changes of texture are less severe. However, we can eliminate this phenomenon by transmitting R-frames more frequently. Figs. 5 and 6 also depict performance improvement on NR-views (views 1, 3 and 4 here), i.e., by sharing the sparsity information between R-view and NR-views, correlation among views is implicitly exploited to improve the reconstruction quality.

We then illustrate the rate-distortion characteristics of ABMR-IEID, EBMR-IEID and IEJD. Figs. 7, 8, and 9 show the comparisons of 3-view *Ballroom*, 4-view *Vassar* and 4-view *Exit* scenarios with resolution 240×320, where the 3rd frame of *Ballroom*, 75th frame of *Vassar* and 9th frame of *Exit* are taken as example, respectively. Evidently, ABMR-

7. Here, we consider a worst case for ABMR-IEID, i.e., the measurement rate of R-frame in R-view is set to 0.3 as the same as NR-frames. Higher measurement rate of R-view will result in higher performance gain because the estimated sparsity pattern will be more accurate.

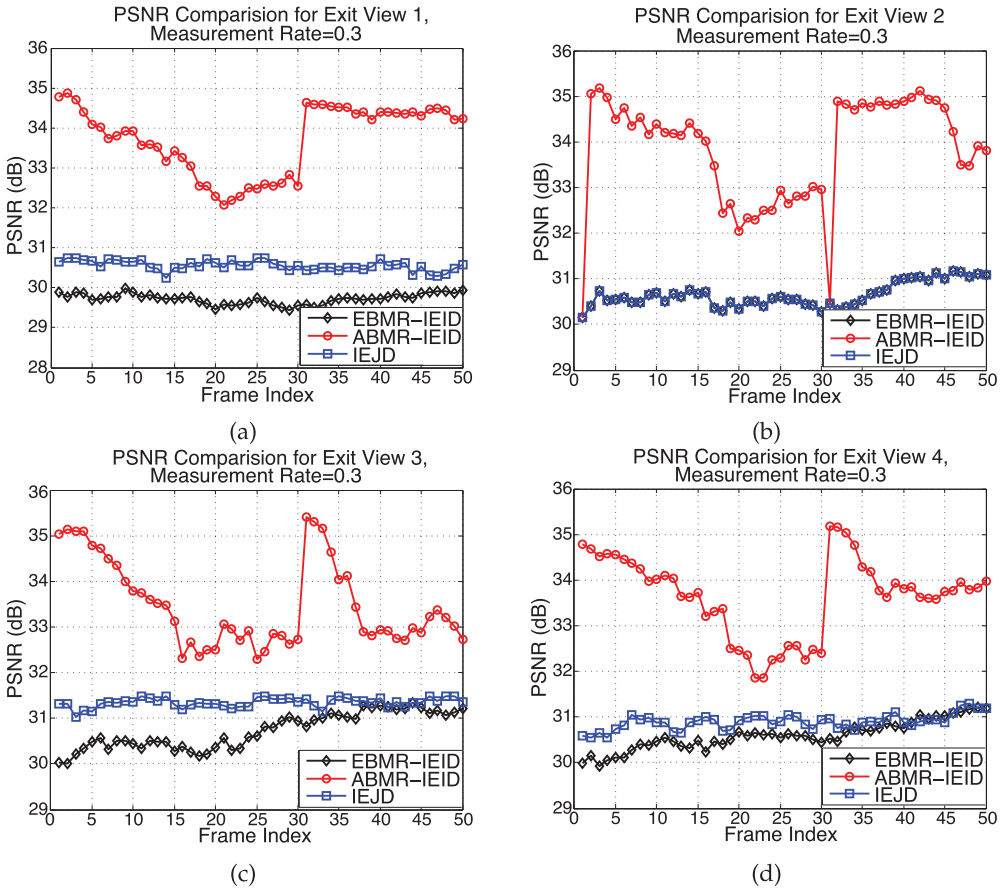


Fig. 6. PSNR against frame index for (a) view 1, (b) view 2 (R-view), (c) view 3, and (d) view 4 of sequence *exit*.

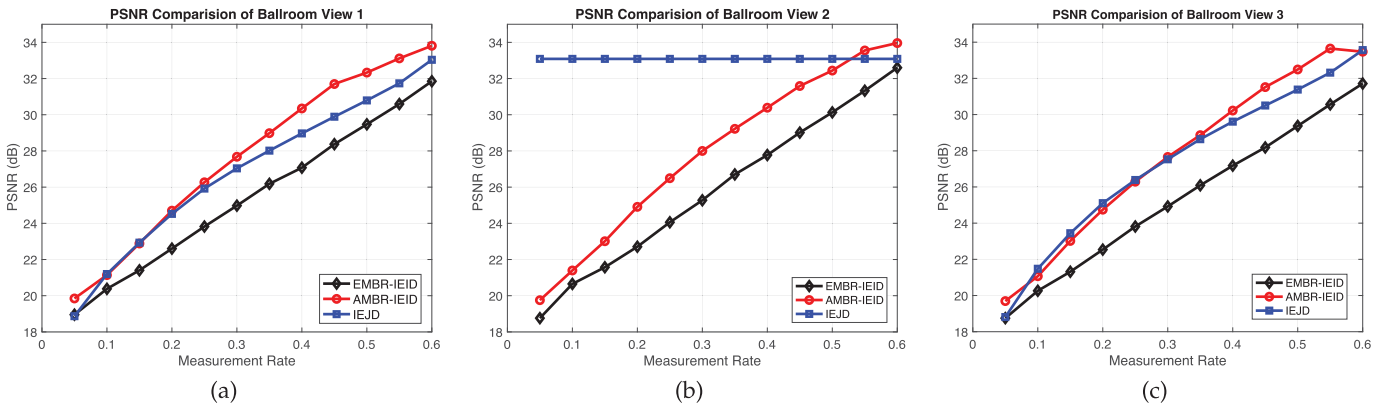


Fig. 7. Rate-distrotion comparison of frame 3 of *Ballroom* sequences: (a) view 1, (b) view 2, and (c) view 3.

IEID outperforms significantly EBMR-IEID and IEJD, especially as the number of measurements increases. Since view 2 is selected as reference view, aka K-view for IEJD, we set a fixed measurement rate 0.6 for the K-view [12], therefore, a platform is observed in view 2 for IEJD method. We can observe that at measurement rate 0.4, ABMR-IEID can improve PSNR by up to 3.5, 4.4 and 2.4 dB, not only on R-view but also on NR-views for all video sequences. To further evaluate the impact of resolution, we take 3-view *Ballroom* and*Balloons* with resolution 480×640 (denoted as Ballroom-H), and 768×1024, respectively as examples since the textures of both sequences change more frequently compared to*Vassar* and*Exit*. As illustrated in Fig. 10, we can see that the proposed ABMR-IEID also outperforms the other two methods by up to 4 dB for *Ballroom* although the correlation slightly decreases as the resolution increases as shown in Table 1. In Fig. 11, we also observe upto 5 dB PSNR improvement obtained by ABMR-IEID compared to the other two methods for *Balloons*.

Next, we extend the scenario to 8 views on *Vassar*, where view 4 is selected as R-view, and the measurement rate is set to 0.35 for all views. Fig. 12 shows the specific reconstructed image comparison, where the left column illustrates the reconstructed frame 25 of view 3 and view 7 by ABMR-IEID, respectively. The milldle column shows the reconstructed images by EBMR-IEID, and the left columns shows the results obtained by using IEJD. We can observe that the quality of images located in the left column is much

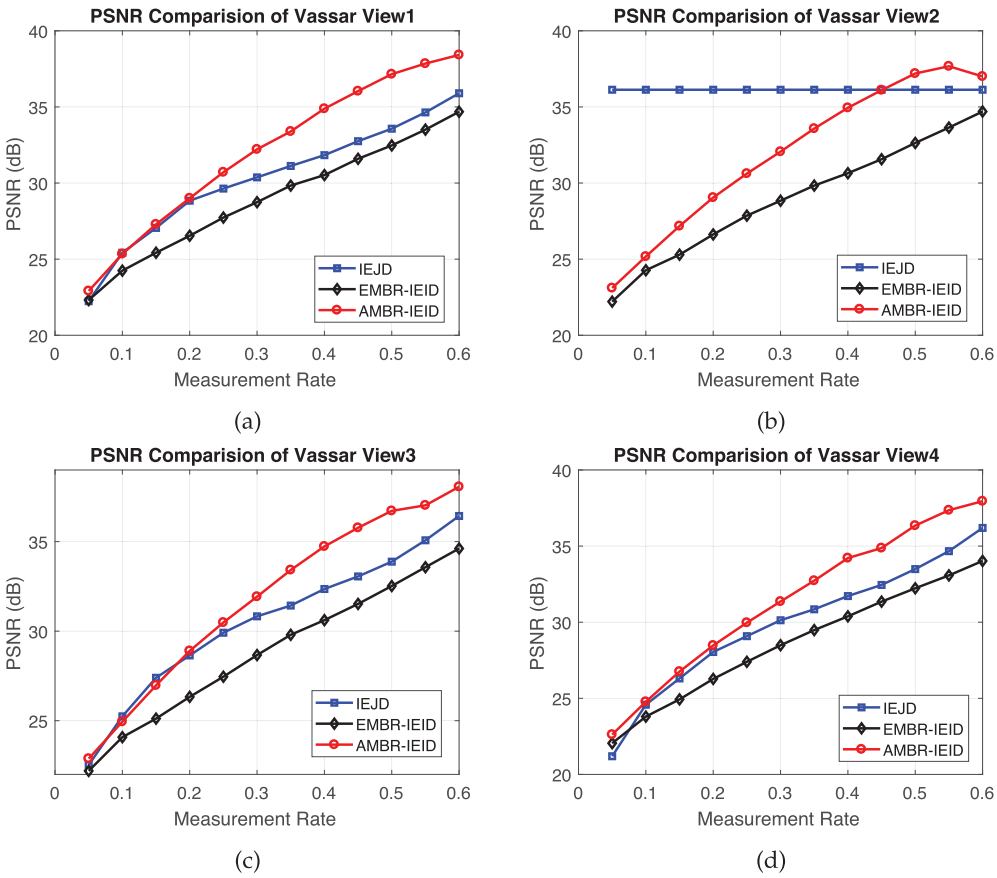


Fig. 8. Rate-distortion comparison for frame 75 of *Vassar* sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4.

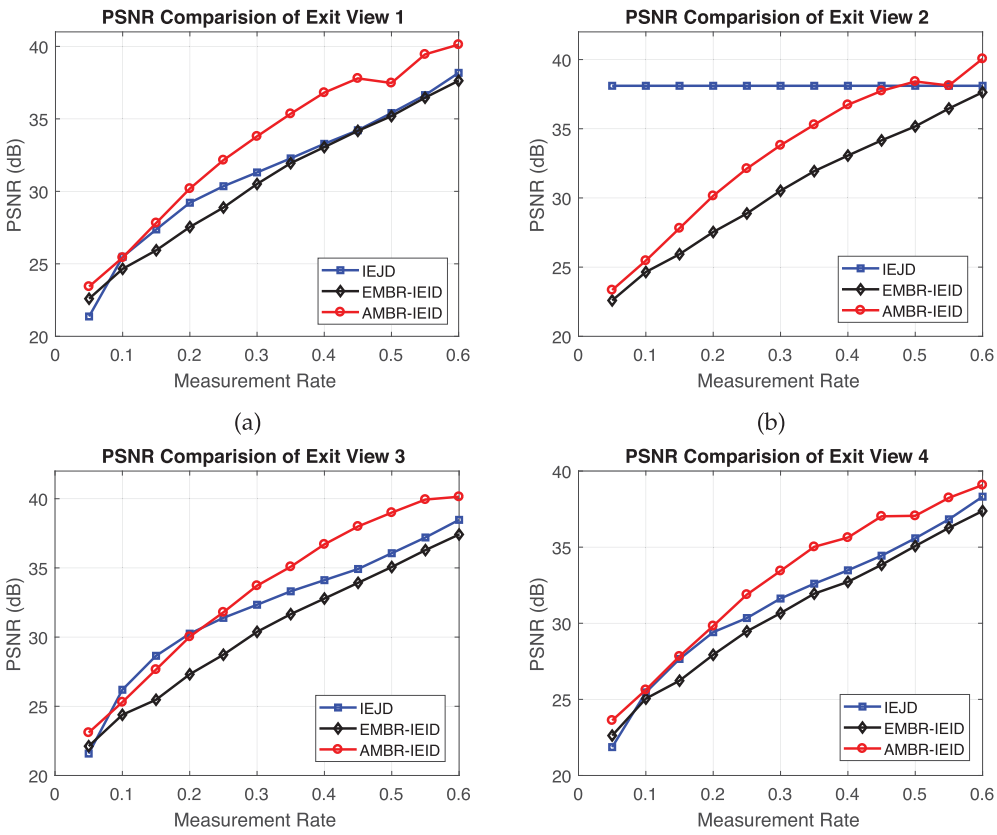


Fig. 9. Rate-distortion comparison for frame 9 of *Exit* sequences: (a) view 1, (b) view 2, (c) view 3, and (d) view 4.

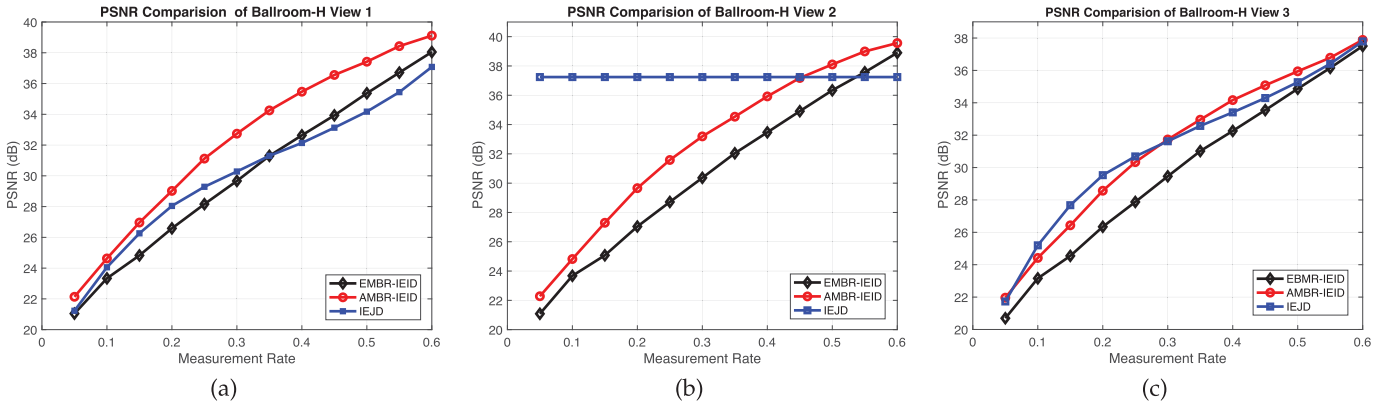


Fig. 10. Rate-distortion comparison of frame 3 of higher resolution *Ballroom* sequences: (a) view 1, (b) view 2, and (c) view 3.

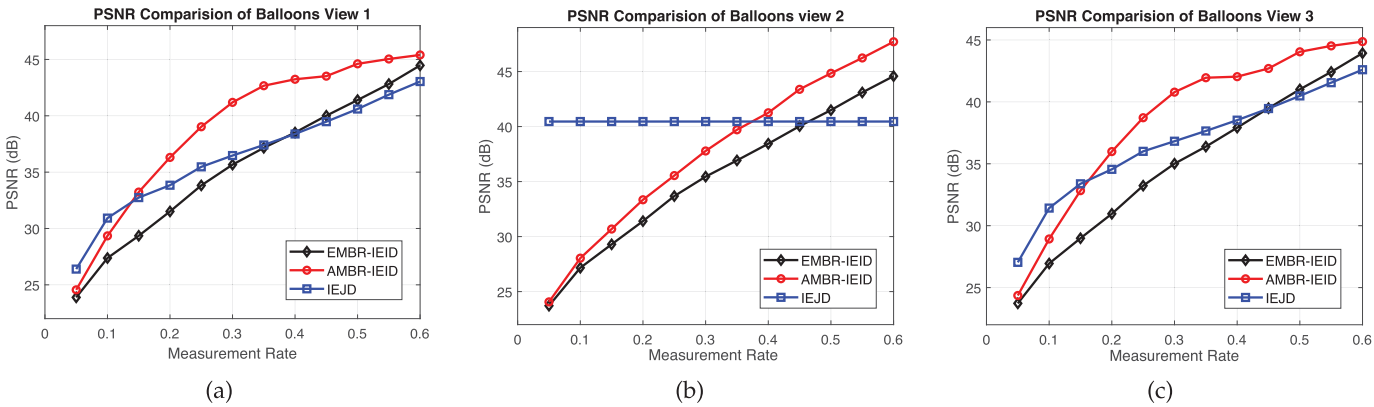


Fig. 11. Rate-distortion comparison of frame 25 of higher resolution *Balloons* sequences: (a) view 1, (b) view 2, and (c) view 3.

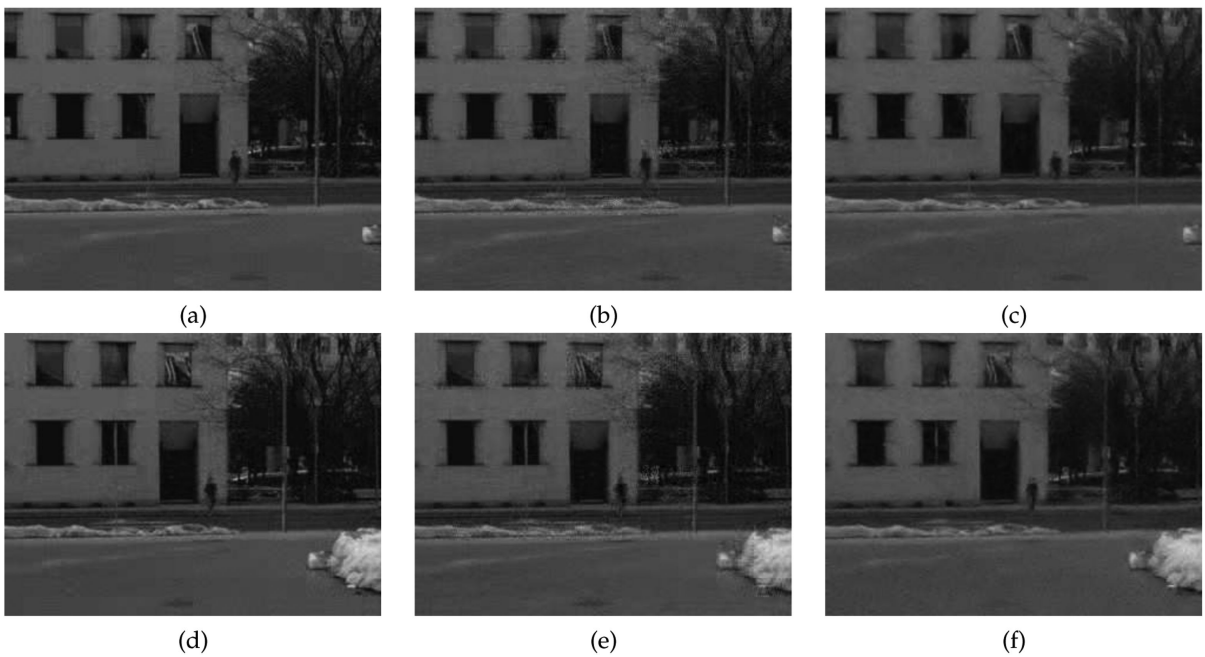


Fig. 12. Reconstructed frame 25 of view 3 by (a) ABMR-IEID, (b) EBMR-IEID, (c) IEJD, and reconstructed frame 25 of view 7 by (d) ABMR-IEID, (e) EBMR-IEID, and (f) IEJD.

better than that in the right two columns (e.g., the curtain in the 2nd floor and person in the scene, and etc.). Furthermore, Table 3 shows the detailed PSNR and SSIM value comparison between ABMR-IEID and EBMR-IEID and IEJD for frame 25 of 8 views. From Fig. 12 and Table 3, we can see that ABMR-IEID also works well on 8 views compared to ABMR-IEID and EBMR-IEID, with PSNR and SSIM improvement up to 3.5 dB and 0.05, respectively. However, the IEJD method proposed in [12] does not perform well on 8 views, where the gain is almost negligible.

TABLE 3
PSNR and SSIM Comparison for *Vassar* Eight Views

View #	ABMR-IEID		EBMR-IEID		IEJD	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
1	33.6675	0.8648	30.0883	0.8215	30.2717	0.7887
2	33.7768	0.8686	30.3459	0.8262	30.3355	0.7902
3	34.1934	0.8771	30.6265	0.8323	30.9214	0.8106
4	33.5766	0.8696	30.4168	0.8294	30.4168	0.8294
5	33.3030	0.8624	30.1011	0.8169	30.3641	0.7909
6	34.2191	0.8846	30.6803	0.8382	30.7265	0.8059
7	32.9924	0.8575	29.8250	0.8162	29.6648	0.7772
8	32.3376	0.8472	29.3713	0.8054	29.5466	0.7742

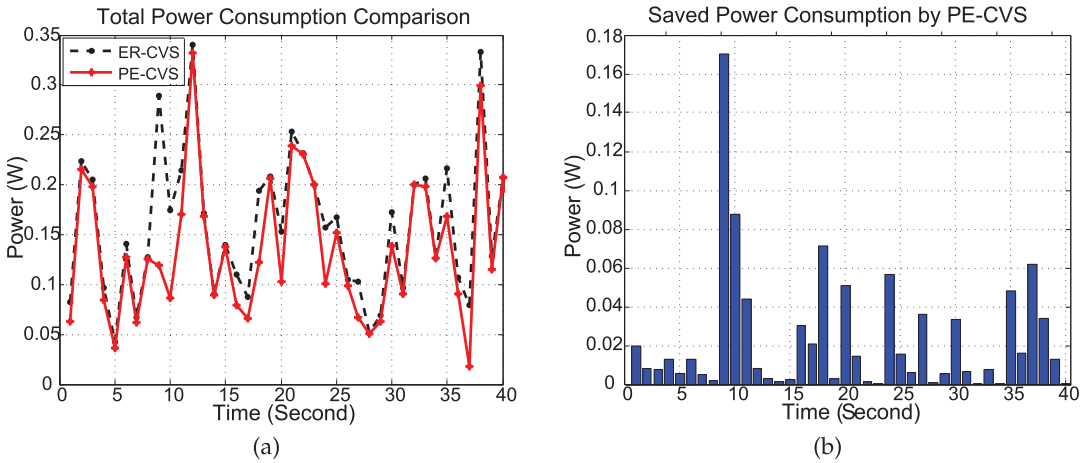


Fig. 13. 2-path Scenario: (a) Total power consumption comparison, (b) Saved power consumption by PE-CVS compared to ER-CVS.

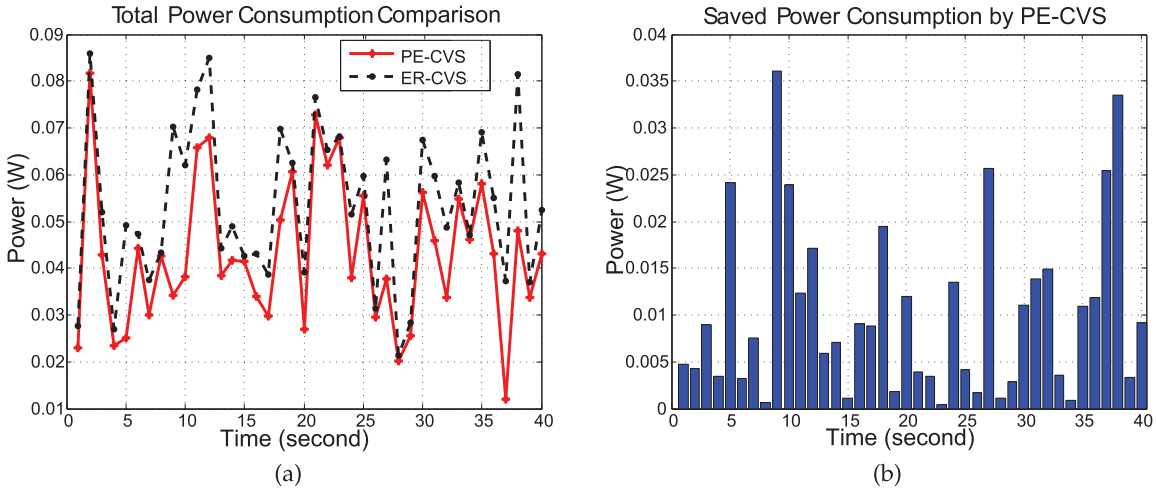


Fig. 14. 3-path Scenario: (a) Total power consumption comparison, (b) Saved power consumption by PE-CVS compared to ER-CVS.

6.2 Evaluation of Power-Efficient Compressive Video Streaming

The following network topologies are considered: i) 2-path scenario with 2-hop path 1 and 1-hop path 2; ii) 3-path scenario with 2-hop path 1, 1-hop path 2 and 2-hop path 3. We assume bandwidth $W = 1$ MHz for each channel. The maximum transmission power at each node is set to 1W and the target distortion in MSE is 50. We also assume the maximum end-to-end delay is $T^{\max} = 0.5$s assigned to each hop proportional to the hop distance. To evaluate PE-CVS (referred to as the proposed power-efficient compressive video streaming algorithm proposed in Section 5), we compare it with an algorithm (referred to as ER-CVS) that equally splits the frame-level rate calculated by PE-CVS onto different paths.

Figs. 13 and 14 illustrate the total power consumption comparison between PE-CVS and ER-CVS and the saved power by PE-CVS compared to ER-CVS for 2-path and 3-path topologies, respectively. From Figs. 13a and 14a, we see that PE-CVS (depicted in red line) results in less power consumption than ER-CVS (black dash line) for both cases. At some points, the total power consumption of PE-CVS

and ER-CVS is almost the same. This occurs because the path-level bit rates calculated by PE-CVS are equal to each other. Since ER-CVS uses frame-level rate obtained from PE-CVS and equally allocates it to each path, thereby resulting in the same power consumption. As shown in Figs. 13b and 14b, the histograms clearly show that PE-CVS saves more power than ER-CVS, up to $170\ \mathrm{mW}$.

7 Conclusion

We addressed the problem of compressive multi-view coding and power-efficient streaming in multi-hop WMSNs. We first proposed a novel compressed sensing based multi-view video coding/decoding architecture, composed of cooperative sparsity-aware independent encoder and independent decoder. We also introduced a central controller to do the sparsity pattern estimation, R-view selection, mean value estimation and implement network optimization algorithms. By introducing limited channel feedback and enabling lightweight sparsity information sharing between R-view and NR-views, the encoders independently encode the video sequences with sparsity awareness and exploit multi-view correlation to improve the reconstruction quality of NR-views. Based on the proposed encoding/decoding architecture, we developed a modeling framework to minimize the multi-view video transmission power but with guaranteed video quality for a multi-hop multi-path sensor network. Extensive simulation results showed that the designed compressive multi-view framework can considerably improve the video reconstruction quality with minimal power consumption.

Acknowledgments

This work was supported by the Office of Naval Research (ONR) under Grant N00014-17-1-2046. A preliminary shorter version of this paper [1] appeared in the ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc).

References

[1] N. Cen, Z. Guan, and T. Melodia, "Multi-view wireless video streaming based on compressed sensing: Architecture and network optimization," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 137–146.

[2] A. E. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Multi-view coding and routing of local features in visual sensor networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.

[3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] S. Pudlewski and T. Melodia, "A tutorial on encoding and wireless transmission of compressively sampled videos," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 754–767, Second Quarter 2013.

[6] H. W. Chen, L. W. Kang, and C. S. Lu, "Dynamic measurement rate allocation for distributed compressive video sensing," *Vis. Commun. Image Process.*, vol. 7744, pp. 1–10, Jul. 2010.

[7] Y. Liu, M. Li, and D. A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 3, pp. 438–444, Mar. 2013.

[8] S. Pudlewski, T. Melodia, and A. Prasanna, "Compressed-sensing enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2012.

[9] S. Pudlewski and T. Melodia, "Compressive video streaming: Design and rate-energy-distortion analysis," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2072–2086, Dec. 2013.

[10] X. Chen and P. Frossard, "Joint reconstruction of compressed multi-view images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 1005–1008.

[11] M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed sensing of multiview images using disparity compensation," in *Proc. Int. Conf. Image Process.*, 2010, pp. 3345–3348.

[12] N. Cen, Z. Guan, and T. Melodia, "Joint decoding of independently encoded compressive multi-view video streams," in *Proc. Picture Coding Symp.*, 2013, pp. 341–344.

[13] N. Cen, Z. Guan, and T. Melodia, "Inter-view motion compensated joint decoding of compressive-sampled multi-view video streaming," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1117–1126, Jun. 2017.

[14] C. Li, D. Wu, and H. Xiong, "Delay—Power-rate-distortion model for wireless video communication under delay and energy constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 7, pp. 1170–1183, Jul. 2014.

[15] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 645–658, May 2005.

[16] A. 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.

[17] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[18] X. Fei, L. Li, H. Cao, J. Miao, and R. Yu, "View's dependency and low-rank background-guided compressed sensing for multi-view image joint reconstruction," *IET Image Process.*, vol. 13, no. 12, pp. 2294–2303, 2019.

[19] J. Zhu, J. Wang, and Q. Zhu, "Compressively sensed multi-view image reconstruction using joint optimization modeling," in *Proc. IEEE Vis. Commun. Image Process.*, 2018, pp. 1–4.

[20] V. Thirumalai and P. Frossard, "Correlation estimation from compressed images," *J. Vis. Commun. Image Representation*, vol. 24, no. 6, pp. 649–660, 2013.

[21] M. Trocan, T. Maugey, J. E. Fowler, and B. Pesquet-Popescu, "Disparity-compensated compressed-sensing reconstruction for multiview images," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2010, pp. 1225–1229.

[22] M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Multistage compressed-sensing reconstruction of multiview images," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2010, pp. 111–115.

[23] S. Elsayed, M. Elsabrouty, O. Muta, and H. Furukawa, "Distributed perceptual compressed sensing framework for multiview images," *Electron. Lett.*, vol. 52, no. 10, pp. 821–823, Dec. 2016.

[24] Y. Liu, C. Zhang, and J. Kim, "Disparity-compensated total-variation minimization for compressed-sensed multiview image reconstruction," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 1458–1462.

[25] Y. Wang, D. Wang, X. Zhang, J. Chen, and Y. Li, "Energy-efficient image compressive transmission for wireless camera networks," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3875–3886, May 2016.

[26] S. Pudlewski and T. Melodia, "Cooperating to stream compressively sampled videos," in *Proc. IEEE Int. Conf. Commun.*, 2013, pp. 1821–1826.

[27] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 267–288, 1996.

[29] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[30] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *Majlesi J. Elect. Eng.*, vol. 9, no. 1, pp. 55–83, Jun. 2014.

[31] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.

[32] X. Yuan, G. Huang, H. Jiang, and P. A. Wilford, "Block-wise lensless compressive camera," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 31–35.

[33] S. Pudlewski and T. Melodia, "A rate-energy-distortion analysis for compressed-sensing-enabled wireless video streaming on multimedia sensors," in *Proc. IEEE Global Telecommun. Conf.*, 2011, pp. 1–6.
[34] F. Mager, D. Baumann, R. Jacob, L. Thiele, S. Trimpe, and M. Zimmerling, "Feedback control goes wireless: Guaranteed stability over low-power multi-hop networks," in *Proc. ACM/IEEE Int. Conf. Cyber-Phys. Syst.*, 2019, pp. 97–108.
[35] D. Baumann, F. Mager, R. Jacob, L. Thiele, M. Zimmerling, and S. Trimpe, "Fast feedback control over multi-hop wireless networks with mode changes and stability guarantees," *ACM Trans. Cyber-Phys. Syst.*, vol. 4, no. 2, pp. 18:1–18:32, Nov. 2019.
[36] X. Guo, Y. He, S. Atapattu, S. Dey, and J. S. Evans, "Power allocation for distributed detection systems in wireless sensor networks with limited fusion center feedback," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4753–4766, Oct. 2018.
[37] Y. He and S. Dey, "Power allocation for secondary outage minimization in spectrum sharing networks with limited feedback," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2648–2663, Jul. 2013.
[38] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge University Press, 2012.
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[40] C. E. Perkins and E. M. Royer, "Ad hoc on-demand distance vector (AODV) routing," in *Proc. IEEE Workshop Mobile Comput. Syst. Appl.*, 1999, pp. 90–100.
[41] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2008, pp. 3–12.
[42] H. Prautzsch and K. Šalkauskas, "Curve and surface fitting: An introduction," *SIAM Rev.*, vol. 31, no. 1, pp. 155–157, 1989.
[43] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
[44] X. Zhu, E. Setton, and B. Girod, "Congestion-distortion optimized video transmission over ad hoc networks," *EURASIP J. Signal Process., Image Commun.*, vol. 20, pp. 773–783, 2005.
[45] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ, USA: Prentice Hall, 2000.
[46] T. Melodia and I. F. Akyildiz, "Cross-layer quality of service support for UWB wireless multimedia sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2008, pp. 2038–2046.
[47] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, 1952.
[48] Mitsubishi Electric Research Laboratories, "MERL multi-view video sequences," 2005. [Online]. Available: ftp://ftp.merl.com/pub/avetro/mvc-testseq/
[49] Accessed: Jan. 20, 2021. [Online]. Available: http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/

Nan Cen (Member, IEEE) received the PhD degree in electrical engineering from Northeastern University, Boston, Massachusetts, in 2019. She is currently an assistant professor with the Department of Computer Science, Missouri University of Science and Technology, Rolla, Missouri. She directs the Wireless Networks and Intelligent Systems (WNIS) Lab, MST. Her research interests include system modeling, control and prototyping for next-generation intelligent wireless networks.

Zhangyu Guan (Senior Member, IEEE) received the PhD degree in communication and information systems from Shandong University, Jinan, China, in 2010. He is currently an assistant professor with the Department of Electrical Engineering, State University of New York at Buffalo, where he directs the Wireless Intelligent Networking and Security (WINGS) Lab. His research interests include network design automation, new spectrum technologies, and wireless network security. He has served as an area editor of the *Elsevier Journal of Computer Networks* since July 2019. He has served as TPC chair for IEEE INFOCOM Workshop on Wireless Communications and Networking in Extreme Environments (WCNEE) 2020, Student Travel grants chair for IEEE Sensor, Mesh and Ad Hoc Communications and Networks (SECON) 2019-2020, Information System (EDAS) chair for IEEE Consumer Communications Networking Conference (CCNC) 2021. He has also served as TPC member for IEEE INFOCOM 2016- 2020, IEEE GLOBECOM 2015-2020, IEEE MASS 2017-2019, IEEE IPCCC 2015-2019, among others.

Tommaso Melodia (Fellow, IEEE) received the PhD degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, Georgia, in 2007. He is the William Lincoln Smith chair professor with the Department of Electrical and Computer Engineering, Northeastern University. He is also the founding director of the Institute for the Wireless Internet of Things and the director of Research for the PAWR Project Office. He is a recipient of the National Science Foundation CAREER Award. He has served as an associate editor of the *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Multimedia*, and is the editor in chief of the *Computer Networks*. He has served as Technical Program Committee chair for IEEE Infocom 2018, general chair for IEEE SECON 2019 and ACM MobiHoc 2020. He is the director of Research for the Platforms for Advanced Wireless Research (PAWR) Project Office, a $100M public-private partnership to establish four city-scale platforms for wireless research to advance the US wireless ecosystem in years to come. His research on modeling, optimization, and experimental evaluation of Internet-of-Things and wireless networked systems has been funded by the National Science Foundation, the Air Force Research Laboratory, Office of Naval Research, DARPA, and Army Research Laboratory. He is a senior member of the ACM.