Interview Motion Compensated Joint Decoding for Compressively Sampled Multiview Video Streams

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

***Abstract*—In this paper, we design a novel multiview video encoding/decoding architecture for wirelessly multiview video streaming applications, e.g., 360 degrees video, Internet of Things (IoT) multimedia sensing, among others, based on distributed video coding and compressed sensing principles. Specifically, we focus on joint decoding of independently encoded compressively sampled multiview video streams. We first propose a novel side-information (SI) generation method based on a new interview motion compensation algorithm for multiview video joint reconstruction at the decoder end. Then, we propose a technique to fuse the received measurements with resampled measurements from the generated SI to perform the final recovery. Based on the proposed joint reconstruction method, we also derive a blind video quality estimation technique that can be used to adapt online the video encoding rate at the sensors to guarantee desired quality levels in multiview video streaming. Extensive simulation results of real multiview video traces show the effectiveness of the proposed fusion reconstruction method with the assistance of SI generated by an interview motion compensation method. Moreover, they also illustrate that the blind quality estimation algorithm can accurately estimate the reconstruction quality.**

***Index Terms*—Multiview video streaming, compressed sensing (CS), Internet of Things (IoT), 360 degrees video.**

I. Introduction

TRADITIONAL multi-view video coding techniques, e.g., MVC H.264/AVC, can achieve high compression ratio by adopting intra-view and inter-view prediction, thus resulting in extremely complex encoders and relatively simple decoders. Recently, a multi-view extension of HEVC (MV-HEVC) was proposed to achieve higher coding efficiency by adopting improved flexible coding tree units (CTUs). [2]–[5] propose an efficient parallel framework based on many-core processors for coding unit partitioning tree decision, motion estimation, deblocking filter, and intra-prediction, respectively, thus achieving many fold speedups compared with current existing parallel methods. However, typical wirelessly multi-view video streaming applications emerging in recent years such as 360 degrees video, and those encountered in Internet of Thing (IoT) multimedia sensing scenarios [6]–[10] are usually composed of low-power and low-complexity mobile devices, smart sensors or wearable sensing devices. 360 degrees video enables immersive "real life", "being there" experience for users by capturing the 360 degree view of the scene of interest, thus requiring higher bitrate than conventional video because it supports a significantly wider field of view. IoT multimedia sensing also needs to simultaneously capture the same scene of interest from different viewpoints and then transmit it to a remote data warehouse, database or cloud for further processing or rendering. Therefore, they need to be based on architectures with relatively simple encoders, while there are less constraints at the decoder side. To address these challenges, so-called Distributed Video Coding (DVC) architectures have been proposed in the last two decades, where the computational complexity is shifted to the decoder side by leveraging architectures with simple encoders and complex decoder to help offload resource-constrained sensors.

Compressed Sensing (CS) is another recent advancement in signal and data processing that shows promise in shifting the computational complexity at the decoder side. CS has been proposed as a technique to enable sub-Nyquist sampling of sparse signals, and it has been successfully applied to imaging systems [11], [12] since natural imaging data can be represented as approximately sparse in a transformed domain, e.g., through discrete cosine transform (DCT) or discrete wavelet transform (DWT). As a consequence, CS-based imaging systems allow the faithful recovery of sparse signals from a relatively small number of linear combinations of the image pixels referred to as measurements. Recent CS-based video coding techniques [13]–[17] have been proposed to improve the reconstruction quality in lossy channels. Therefore, CS has been proposed as a clean-slate alternative to traditional image or video coding paradigms since it enables imaging systems that sample and compress data in a single operation, thus resulting in low-complexity encoders and more complex decoders, which can help offload the sensors and further prolong the lifetime of the mobile devices or sensors.

In this context, our objective is to develop a novel low-complexity multi-view coding/encoding architecture for wirelessly video streaming applications, e.g., 360 degrees immersive video, IoT multimedia sensing, among others, where devices or sensors are usually equipped with power-limited battery. However, current existing algorithms are mostly based on the MVC h.264/AVC or MV-HEVC architecture, which involves complex

Manuscript received May 26, 2016; revised October 18, 2016 and November 29, 2016; accepted January 2, 2017. Date of publication January 16, 2017; date of current version May 13, 2017. This work is based upon material supported in part by the U.S. National Science Foundation under Grant CNS1422874, and in part by the U.S. Office of Naval Research under Grant N00014-16-1-2213 and Grant ARMY W911NF-17-1-0034. This paper was presented in part at the Picture Coding Symposium, San Jose, CA, December 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiaoqing Zhu.

The authors are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: ncen@ece.neu.edu; zguan@ece.neu.edu; melodia@ece.neu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2653770

encoders (motion estimation, motion compensation, disparity estimation, among others) and simple decoder, and is thus not suitable to low-power multi-view video streaming applications. To address this challenge, we propose a novel mult-view encoding/decoding architecture based on compressed sensing theory, where video acquisition and compressing are implemented in one step through low-complexity and low-power compressive sampling (i.e., simple linear operations) while complex computations are shifted to the decoder side. Thus this proposed architecture is more suitable to the aforementioned multi-view scenarios compared with the conventional coding algorithm. To be specific, at the encoder end, one view is selected as a key view (K-view) and encoded at a higher measurement rate; while the other views (CS-views) are encoded at relatively lower rates. At the decoder end, the K-view is reconstructed using a traditional CS recovery algorithm, while the CS-views are jointly decoded by a novel fusion decoding algorithm based on side information generated by a new proposed inter-view motion compensation scheme. Based on the proposed architecture, we develop a blind quality estimation algorithm and apply it to perform feedback-based rate control to regulate the received video quality.

We claim the following contributions:

1) *Side information generated by inter-view motion compensation.* We design a motion compensation algorithm for inter-view prediction, based on which we propose a novel side information generation method that uses the initially reconstructed CS-view and the reconstructed K-view.
2) *CS-view fusion reconstruction.* State-of-the-art joint reconstruction methods either use side information [18] as sparsifying basis or use it as the initial point of the developed joint recovery algorithm [19]. Differently, we operate on the measurement domain and propose a novel fusion reconstruction method by padding measurements resampled from side information to the original received CS-view measurements. Then, traditional sparse signal recovery methods can be used to perform the final reconstruction of CS-view by using the resulting measurements.
3) *Blind quality estimation for compressively-sampled video.* To guarantee the CS-based multi-view streaming quality is not trivial since original pixels are not only unavailable at the encoder end but also not available at the decoder side. Therefore, how to estimate the reconstruction quality as accurate as possible plays fundamental roles on the quality-assured rate controlling. Based on the proposed reconstruction approach, we develop a blind quality estimation approach, which further can be used to effectively guide the rate adaptation at the encoder end.

The reminder of the paper is organized as follows. In Section II, related works are discussed. In Section III, we briefly review the basic concepts used in compressed imaging system. In Section IV, we introduce the overall encoding/decoding compressive multi-view video streaming framework, and in Section V, we describe the inter-view motion compensation based multi-view fusion decoder. The performance evaluations are presented in Section VI, and in Section VII we draw the main conclusions.

II. Related Work

CS-based Mono-view Video. In recent years, several mono-view video coding schemes based on compressed sensing principles have been proposed in the literature [14]–[16], [18], [20]–[22]. These works mainly focus on single view CS reconstruction by leveraging the correlation among successive frames. For example, [19] proposes a distributed compressive video sensing (DCVS) framework, where video sequences are composed of several GOPs (group of pictures), each consisting of a key frame followed by one or more non-key frames. Key frames are encoded at a higher rate than non-key frames. At the decoder end, the key frame is recovered through the GPSR (gradient projection for sparse reconstruction) algorithm [23], while the non-key frames are reconstructed by a modified GRSR where side information is used as the initial point. Based on [19], the authors further propose dynamic measurement rate allocation for block-based DCVS. In [18], the authors focus on improving the video quality by constructing better sparse representations of each video frame block, where Karhunen-Loeve bases are adaptively estimated with the assistance of implicit motion estimation. [21] and [20] consider the rate allocation and energy consumption under the above-mentioned state-of-the-art mono-view compressive video sensing frameworks. [14] and [15] improve the rate-distortion performance of CS-based codecs by jointly optimizing the sampling rate and bit-depth, and by exploiting the intra-scale and inter-scale correlation of multiscale DWT, respectively.

CS-based Multi-view Video. More recently, several proposals have appeared for CS-based multi-view video coding [24]–[27]. In [24], 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 [25] 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. The authors of [26] 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 [27] via residual recovery. The readers are referred to [26], [27] and references therein for details. Differently, in this work, we propose a novel CS-based joint decoder based on a newly-designed algorithm to construct an inter-view motion compensated side frame. With respect to existing proposals, the proposed framework considers multi-view sequences encoded at different rates and with more general sparsifying matrices. Moreover, only one reference view (not necessarily the closest one) is selected to obtain the side frame for joint decoding.

Blind Quality Estimation. Ubiquitous multi-view video streaming of visual information and the emerging applications that rely on it, e.g., multi-view video surveillance, 360 degrees video, and IoT multimedia sensing, require an effective means to assess the video quality because the compression methods and

the error-prone wireless links can introduce distortion. Peak Signal-to-Noise Ratio (PSNR) and SSIM (Structural Similarity) [28] are examples of successful image quality assessment metrics; which however require full reference image at the decoder end. In many applications such as surveillance scenarios, however, the reference signal is not available to perform the comparison. Especially, when compressed sensing is used, the reference signal may not even be available at the encoder end. Readers are referred to [29], [30] and references therein for good overviews of image quality assessment (FR-IQA) and non-reference (blind) image quality assessment (NR-IQA) for state-of-the-art video coding methods, e.g., H.264/AVC, respectively. Yet, to the best of our knowledge, we propose for the first time a NR-IQA scheme for compressive imaging systems.

III. Preliminaries

In this section, we briefly introduce the basic concepts of compressed sensing for signal acquisition and recovery as applied to compressive video streaming systems.

A. CS Acquisition

We consider the image frame signal vectorized and represented as $\mathbf{x} \in \mathbb{R}^N$, with $N = H \times W$ denoting the number of pixels in one frame, with H and W representing the dimensions of the captured scene. The element x_i of $\mathbf{x}$ represents the ith pixel in the vectorized signal representation. As mentioned above, CS-based sampling and compression are implemented in a single step. We denote the sampling matrix as $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$, with $M \ll N$. Then, the acquisition process can be expressed as

$$\mathbf{y} = \mathbf{\Phi x} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ represents the measurements and the vectorized compressed image signal.

B. CS Recovery

Most natural images can be represented as a sparse signal in some transformed domain $\mathbf{\Psi}$, e.g., DWT or DCT, expressed as

$$\mathbf{x} = \mathbf{\Psi s} \tag{2}$$

where $\mathbf{s} \in \mathbb{R}^N$ denotes the sparse representation of the image signal. Then, we can rewrite (1) as

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi s}. \tag{3}$$

If $\mathbf{s}$ has K non-zero elements, we refer to $\mathbf{x}$ as a K-sparse signal with respect to $\mathbf{\Psi}$.

In [11], the authors proved that if $\mathbf{A} \triangleq \mathbf{\Phi \Psi}$ satisfies the so-called Restricted Isometry Property (RIP) of order K

$$(1 - \delta_k)||\mathbf{s}||_{l_2}^2 \leq ||\mathbf{As}||_{l_2}^2 \leq (1 + \delta_k)||\mathbf{s}||_{l_2}^2 \tag{4}$$

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

$$\begin{aligned} \mathrm{P}_1\text{:} \quad & \underset{\mathbf{s} \in \mathbb{R}^N}{\text{Minimize}} \quad ||\mathbf{s}||_0 \\ & \text{Subject to: } \mathbf{y} = \mathbf{\Phi \Psi s} \end{aligned} \tag{5}$$

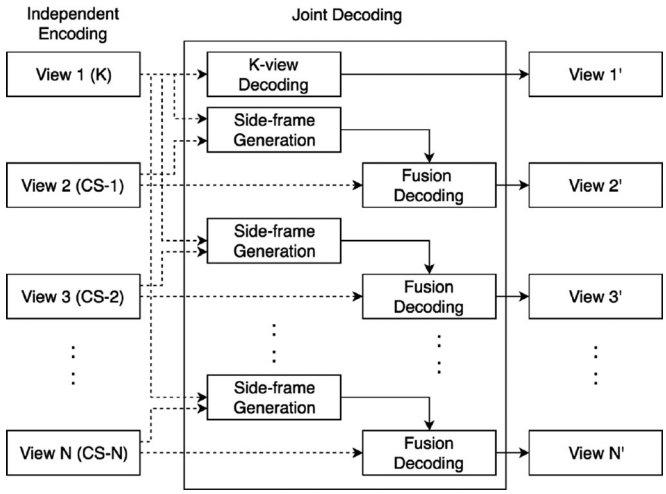


Fig. 1. Multiview encoding/decoding architecture.

by taking only

$$M = c \cdot K \log(N/K) \tag{6}$$

measurements according to the uniform uncertainty principle (UUP), where c is some predefined constant. Then, $\mathbf{x}$ can be obtained as

$$\hat{\mathbf{x}} = \mathbf{\Psi s}^*. \tag{7}$$

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 works rely on l_1 minimization with a relaxed constraint in the form of

$$\begin{aligned} \mathrm{P}_2\text{:} \quad & \underset{\mathbf{s} \in \mathbb{R}^N}{\text{Minimize}} \quad ||\mathbf{s}||_1 \\ & \text{Subject to : } ||\mathbf{y} - \mathbf{\Phi \Psi s}||_2 \leq \epsilon \end{aligned} \tag{8}$$

to recover $\mathbf{s}$. Note that P_2 is also a convex optimization problem [31]. The complexity of reconstruction is $O(M^2 N^{3/2})$ if solved by interior point methods [32]. Moreover, researchers interested in sparse signal reconstruction have developed more efficient solvers [23], [33], [34]. For measurement matrix $\mathbf{\Phi}$, there are two types, Gaussian random and deterministic. Readers are referred to [18], [35] and references therein for details about Gaussian random and deterministic measurement matrix constructions.

IV. System Architecture

We consider a multi-view video streaming system equipped with N cameras, with each camera capturing the same scene of interest from different perspectives. At the source nodes, each captured view is encoded and transmitted independently and jointly decoded at the receiver end. The proposed CS-based N-view encoding/decoding architecture is depicted in Fig. 1, with $N > 2$.

At the encoder side, we first select one of the considered views as a reference (referred to as *K-view*) for other views (referred to as *CS-views*). The frames of the K-view and of the CS-view are encoded at a measurement rate of R_k and R_{cs}, respectively. According to the asymmetric distributed video

coding principle, the reference view (i.e., K-view) is coded at a higher rate than the non-reference views (i.e., CS-views). In the following, we assume that $R_{\rm cs} \leq R_k$. The size of the scene of interest is denoted as $H \times W$ (in pixels), with the number of total pixels being $N = H \times W$. The K-view frame (denoted as $\mathbf{x}_k \in \mathbb{R}^N$) is compressively sampled into a measurement vector $\mathbf{y}_k \in \mathbb{R}^{M_k}$ with measurement rate $\frac{M_k}{N} = R_k$, and the CS-view frame $\mathbf{x}_{\rm cs} \in \mathbb{R}^N$ is sampled into $\mathbf{y}_{\rm cs} \in \mathbb{R}^{M_{\rm cs}}$ with $\frac{M_{\rm cs}}{N} = R_{\rm cs}$. Readers are referred to [36] and references therein for details of the encoding procedure.

At the decoder side, the reconstruction of K-view frames is only based on the received K-view measurements. To reconstruct a CS-view frame, we propose a novel inter-view motion compensated joint decoding method. We first generate a side frame based on the received K-view and CS-view measurements. Then, we fuse the initially received measurements of the CS-view frame with the newly sampled measurements from generated side frame through the proposed novel fusion algorithm. In the following section, we describe the joint multi-view decoder in detail.

V. Joint Multiview Decoding

In this section, we discuss the proposed joint multi-view decoding method. The frames of the K-view are first reconstructed to serve as a reference for the CS-view reconstruction procedure.

A. K-view Decoding

Denote the *received* measurement vector of any frame of the K-view video sequence as $\hat{\mathbf{y}}_k \in \mathbb{R}^{M_k}$ (i.e., a distorted version of $\mathbf{y}_k$ considering the joint effects of quantization, transmission errors, and packet drops due to playout deadline violation). Based on CS theory as discussed in Section III, the K-view frame can be simply reconstructed by solving the following convex optimization problem (sparse signal recovery)

$$\begin{aligned} \mathrm{P}_3\colon \quad & \underset{\mathbf{s}\in\mathbb{R}^N}{\text{Minimize}} \quad ||\mathbf{s}||_1 \\ & \text{Subject to}: \ ||\hat{\mathbf{y}}_k - \mathbf{\Phi}_k\mathbf{\Psi}\mathbf{s}||_2^2 \leq \epsilon \end{aligned} \tag{9}$$

and then by mapping $\hat{\mathbf{x}}_k = \mathbf{\Psi}\mathbf{s}^*$, with $\mathbf{\Phi}_k$ and $\mathbf{\Psi}$ representing the K-view sampling matrix and the sparsifying matrix, respectively. Here, ϵ denotes the predefined error tolerance, and $\mathbf{s}^*$ represents the reconstructed coefficients (i.e., the minimizer of (9)).

B. Interview Motion Compensated Side Frame

Motivated by the traditional mono-view video coding schemes, where motion estimation and compensation techniques are used to generate the prediction frame, we propose an inter-view motion estimation and compensation method for a multi-view video coding scenario. The core idea behind the proposed technique for generating the side frame is to compensate the reconstructed high-quality K-view frame $\hat{\mathbf{x}}_k$ through an estimated inter-view motion vector. To obtain a more accurate inter-view motion estimation vector, we first down-sample the received K-view measurements $\hat{\mathbf{y}}_k$ to obtain the same number of measurements as the number of received CS-view measurements. Then, we use these down-sampled K-view measurements to reconstruct a lower-quality K-view that has the equivalent level of quality as the initially reconstructed CS-view frame. Next, we compare the preliminary reconstructed CS-view with the reconstructed lower-quality K-view to obtain the side frame. Below, we elaborate on the main components of the side frame generation method as illustrated in Fig. 2.

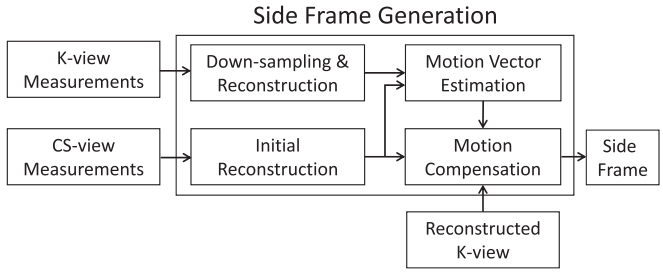


Fig. 2. Block diagram of side frame generation.

CS-view initial reconstruction. We denote $\hat{\mathbf{y}}_{\rm cs}$ and $\mathbf{\Phi}_{\rm cs}$ as the received distorted version of CS-view frame measurements and the corresponding sampling matrix, respectively. By substituting $M_{\rm cs}$ received measurements $\hat{\mathbf{y}}_{\rm cs}$, $\mathbf{\Phi}_{\rm cs}$ and $\hat{\mathbf{x}}_{\rm cs}$ into (9), a *preliminary* reconstructed CS-view frame (denoted as $\hat{\mathbf{x}}^p_{\rm cs}$) can be obtained by solving the corresponding optimization problem.

K-view down-sampling and reconstruction. As mentioned above, the reconstructed K-view frame has higher quality than the preliminary reconstructed CS-view. To achieve higher accuracy in the estimation of the inter-view motion vector, we propose to first down-sample the received K-view measurement vector $\hat{\mathbf{y}}_k$ to obtain a new K-view frame with the same (or comparable) reconstructed quality with respect to $\hat{\mathbf{x}}^p_{\rm cs}$. Experiments were conducted to validate this approach, which results in more accurate motion vector estimation than the originally reconstructed K-view frame $\hat{\mathbf{x}}_k$.

Since $R_{\rm cs} \leq R_k$ as stated in Section IV, without loss of generality, we consider the CS-view sampling matrix $\mathbf{\Phi}_{\rm cs}$ to be a sub-matrix of $\mathbf{\Phi}_k$. Then, down-sampling can be achieved by selecting from $\hat{\mathbf{y}}_k$ only measurements corresponding to $\mathbf{\Phi}_{\rm cs}$, which is equivalent, apart from transmission errors and quantization errors, to sampling the original K frame with the matrix used for sampling the CS frame. The down-sampled K-view measurement vector and the corresponding reconstructed k-view frame with lower quality are denoted as $\hat{\mathbf{y}}^d_k$ and $\hat{\mathbf{x}}^d_k$, respectively.

Inter-view motion vector estimation. With the preliminary reconstructed CS-view frame $\hat{\mathbf{x}}^p_{\rm cs}$ and the reconstructed down-sampled quality-degraded K-view frame $\hat{\mathbf{x}}^d_k$, we can then estimate the inter-view motion vector by comparing $\hat{\mathbf{x}}^p_{\rm cs}$ and $\hat{\mathbf{x}}^d_k$. The detailed inter-view vector estimation procedure is as follows. First, we divide $\hat{\mathbf{x}}^p_{\rm cs}$ into a set $\mathcal{B}^p_{\rm cs}$ of blocks with block size $B^p_{\rm cs} \times B^p_{\rm cs}$ (in pixel). For each current block $i_{\rm cs} \in \mathcal{B}^p_{\rm cs}$, within a predefined search range p in the lower-quality K-frame $\hat{\mathbf{x}}^d_k$, a set $\mathcal{B}^d_k(i_{\rm cs}, p)$ of reference blocks, each with the same block size $B^p_{\rm cs} \times B^p_{\rm cs}$, can be identified based on existing strategies [37], e.g., exhaustive search (ES), three step search (TSS), or diamond search (DS). Then, we calculate the mean of absolute difference (MAD) between block $i_{\rm cs} \in \mathcal{B}^p_{\rm cs}$ and any block $i_k \in \mathcal{B}^d_k(i_{\rm cs}, p)$,

which is defined as

$$MAD_{i_{\rm cs}i_k} = \frac{\sum_{m=1}^{B_{\rm cs}^p}\sum_{n=1}^{B_{\rm cs}^p}\left|v_{\rm cs}^p(i_{\rm cs},m,n) - v_k^d(i_k,m,n)\right|}{B_{\rm cs}^p \times B_{\rm cs}^p} \tag{10}$$

with $v_{\rm cs}^p(i_{\rm cs},m,n)$ and $v_k^d(i_k,m,n)$ denoting the value of the pixels at (m,n) in block $i_{\rm cs} \in \mathcal{B}_{\rm cs}^p$ and $i_k \in \mathcal{B}_k^d(i_{\rm cs},p)$, respectively. Next, the best matching block denoted by $i_k^* \in \mathcal{B}_k^d(i_{\rm cs},p)$ has the minimum MAD, which can be obtained by solving

$$i_k^* = \underset{i_k \in \mathcal{B}_k^d(i_{\rm cs},p)}{\arg\min}\ MAD_{i_{\rm cs}i_k} \tag{11}$$

with $MAD_{i_{\rm cs}i_k^*}$ being the corresponding minimum MAD value.

In the single view scenario [38], it is sufficient to search for the block corresponding to the minimum MAD (i.e., block i_k^*) to estimate the motion vector. However, in the multi-view case, the best matching block i_k^* is not necessarily a proper estimation of block $i_{\rm cs}$ due to the possible "hole" problem (i.e., an object that appears in a view is occluded in other views), which can be rather severe.

To address this challenge, we adopt a threshold-based policy. Let $MAD_{\rm th}$ represent the predefined MAD threshold, which can be estimated online by periodically transmitting a frame at a higher measurement rate. Denote $\Delta m(i_{\rm cs})$ and $\Delta n(i_{\rm cs})$ as the horizontal and vertical offset (aka motion vector, in pixel) of the block i_k^* relative to the current block $i_{\rm cs}$. Then, if a block $i_k^* \in \mathcal{B}_k^d(i_{\rm cs},p)$ can be found satisfying $MAD_{i_{\rm cs}i_k^*} \le MAD_{\rm th}$, then the current block $i_{\rm cs} \in \mathcal{B}_{\rm cs}^p$ is marked as *referenced* with motion vector $(\Delta m(i_{\rm cs}), \Delta n(i_{\rm cs}))$; Otherwise, the block is marked as *non-referenced*.

Inter-view motion compensation. After estimating the inter-view motion vector, the side frame $\mathbf{x}_{\rm si} \in \mathbb{R}^N$ can then be generated by compensating the initially reconstructed CS-view frame $\hat{\mathbf{x}}_{\rm cs}^p$, with above-estimated motion vector $(\Delta m(i_{\rm cs}), \Delta n(i_{\rm cs}))$ for each block in $\mathcal{B}_{\rm cs}^p$, and the reconstructed high-quality K-view frame $\hat{\mathbf{x}}_k$.[1] The detailed procedure of compensation is as follows. First, we initialize the side frame $\mathbf{x}_{\rm si}$ to $\mathbf{x}_{\rm si} = \hat{\mathbf{x}}_{\rm cs}^p$. Then, we replace each referenced block $i_{\rm cs}$ by using the corresponding block from the initially reconstructed high-quality K-view frame $\hat{\mathbf{x}}_k$ with the estimated motion vector $(\Delta m(i_{\rm cs}), \Delta n(i_{\rm cs}))$.

C. Fusion Decoding Algorithm

The side frame, aka side information, plays a very significant role in state-of-the-art CS-based joint decoding approaches, acting as the initial point [19] of the joint recovery algorithm or sparsifying basis [18]. Differently, we explore a novel joint decoding method by directly adopting the side information in the measurement domain. Specifically, we propose to fuse the received CS-view measurements $\hat{\mathbf{y}}_{\rm cs}$ and the measurements resampled from the above generated side-frame $\mathbf{x}_{\rm si}$ to obtain a new measurement vector for further reconstruction of the CS-view. The key idea is to involve more measurements with the assistance of the side frame to further improve the reconstructed quality. This is achieved by generating CS measurements by sampling $\mathbf{x}_{\rm si}$, appending the generated measurements to $\hat{\mathbf{y}}_{\rm cs}$, and then reconstructing a new CS-view frame based on the combined measurements.

To sample the side frame, we use a sampling matrix $\mathbf{\Phi}$, with $\mathbf{\Phi}_{\rm cs}$ and $\mathbf{\Phi}_k$ both being a sub-matrix of $\mathbf{\Phi}$. We then select a number $R_{\rm si} \times H \times W$ of the resulting measurements, with $R_{\rm si}$ representing the predefined measurement rate for the side frame. The value of $R_{\rm si}$ depends on the amount of CS-view measurements $\hat{\mathbf{y}}_{\rm cs}$ that have already been received. Experiments have been conducted to verify the intuitive conclusion that larger $R_{\rm cs}$ implies to smaller $R_{\rm si}$. The experiments show that if a sufficient number of CS-view measurements is received at the decoder to result in acceptable reconstruction quality, adding more measurements and combining them from the side frame will result in the introduction of more noise, ultimately reducing the video quality of the recovered frame. Based on experimental evidence, we set $R_{\rm si}$ as

$$\begin{cases} R_{\rm si} = 1 - R_{\rm cs}, & \text{if } R_{\rm cs} \le 0.5 \\ R_{\rm si} = 0.6 - R_{\rm cs}, & \text{if } 0.5 < R_{\rm cs} \le 0.6 \\ R_{\rm si} = 0, & \text{if } R_{\rm cs} > 0.6. \end{cases} \tag{12}$$

With the newly generated $R_{\rm cs} + R_{\rm si}$ measurements $\tilde{\mathbf{y}}_{\rm cs}$, following optimization problem (9), the final jointly reconstructed CS-view frame (denoted by $\hat{\mathbf{x}}_{\rm cs}$) can be obtained.

D. Blind Video Quality Estimation

A natural question for the newly designed multi-view codec is: how good is the reconstructed video quality? As stated in Section II, how to assess the reconstruction quality at the decoder end without original reference frames is substantially an open problem, especially for CS-based video coding systems where the original pixels are not available either at the transmitter or at the receiver side. To address this challenge, we propose a blind video quality estimation method within the proposed compressively-sampled multi-view coding/decoding framework described above.

Most state-of-the-art quality assessment metrics, e.g., PSNR or SSIM, are based on the comparison between a-priori-known reference frames and the reconstructed frames in the pixel domain. In this context, we propose to blindly evaluate the quality in the measurement domain by adopting an approach similar to that used to calculate PSNR. The detailed procedure is as follows. First, the reconstructed CS-view frame $\hat{\mathbf{x}}_{\rm cs}$ is resampled at the CS-view measurement rate $R_{\rm cs}$, with the same sampling matrix $\mathbf{\Phi}_{\rm cs}$, thus obtaining $M_{\rm cs}$ new measurements denoted by $\overline{\mathbf{y}}_{\rm cs}$. Then, the measurement-domain PSNR of $\hat{\mathbf{x}}_{\rm cs}$ with respect to the original frame $\mathbf{x}_{\rm cs}$ (which is not available even at the encoder side) can be estimated by comparing the measurement vector $\hat{\mathbf{y}}_{\rm cs}$ and $\overline{\mathbf{y}}_{\rm cs}$, as

$$\text{PSNR} = 10\log_{10}\frac{(2^n-1)^2}{\text{MSE}} + \Delta\text{PSNR} \tag{13}$$

where n is the number of bits per measurement, and

$$\text{MSE} = \frac{\|\hat{\mathbf{y}}_{\rm cs} - \overline{\mathbf{y}}_{\rm cs}\|_2^2}{M_{\rm cs}^2}. \tag{14}$$

[1]Note that we estimate the motion vector based on the quality-degraded K-view frame, but compensate the initially reconstructed CS-view frame using the K-view frame at the original reconstructed quality.

In (13), $\Delta\mathrm{PSNR}$ is a compensation coefficient that has been found to stay constant or vary only slowly for each view in the conducted experiments. Hence, it can be estimated online by periodically transmitting a CS-frame at a higher measurement rate.

The proposed blind estimation technique can then be used to control the encoder to dynamically adapt the encoding rate by adaptively increasing or decreasing the rate to guarantee the perceived video quality at the receiver side.

VI. Performance Evaluation

In this section, we experimentally study the performance of the proposed compressive multi-view video decoder by evaluating the perceptual quality, PSNR and SSIM. Three multi-view test sequences are used, i.e., *Vassar*, *Exit* and *Ballroom* representing scenarios with slow, moderate and fast movement characteristics, respectively. The spatial dimension for each frame is 320×240 (in pixel). All experiments are conducted only on the luminance component.

At the encoder side, the sampling matrixes $\mathbf{\Phi}_k$, $\mathbf{\Phi}_{\mathrm{cs}}$ and $\mathbf{\Phi}$ are implemented with Hadamard matrixes. At the decoder end, TSS [39] is used for motion vector estimation, with block size and search range set to $B = 16$ and $p = 32$, respectively. In the blind video quality estimation algorithm the value of Δ PSNR is set to 6 and 2.9 for *Ballroom* and *Exit*, respectively. GPSR [23] is used to solve P_3 in (9).

As stated in Section I, the inter-view motion-compensated side frame generation approach and the fusion decoding method for CS-view frames are two of the main contributions of the paper. To evaluate the effectiveness, we compare the following four approaches: i) the proposed inter-view motion compensated side frame based fusion decoding method for CS-view frame (referred to as *MC fusion*), ii) the GPSR joint decoder proposed in [19] by adopting the side frame generated by the proposed inter-view motion compensation method (referred to as *MC joint GPSR*), iii) the GPSR joint reconstruction by adopting initially reconstructed CS-view frame as side frame (referred to as *joint GPSR*)[2] and iv) independent decoding method (referred to as *Independent*) used as a baseline.

First, we evaluate the improvement of CS-view perceptual quality of the proposed *MC fusion* decoding method compared with *Independent* reconstruction approach by considering a specific frame as an example, i.e., the 5th frame of *Exit* and the 25th frame of *Vassar*. 2-view scenario is considered, where view 1 is set as K-view with measurement rate 0.6 and view 2 is CS-view. Results are illustrated in Figs. 3 and 4. We observe that the blurring effect in the independently reconstructed frame is mitigated through joint decoding. Taking the regions of the person, bookshelf and photo frame in Fig. 3(b) and 3(d), and almost the whole regions in Fig. 4(b) and 4(d) as examples, we can see that the video quality improvement is noticeable, which corresponds to an improvement in PSNR from 28.17 dB to 29.58 dB and 25.81 dB to 27.87 dB, respectively, and in an improvement in SSIM of 0.09 (from 0.75 to 0.84) and 0.14 (from 0.60 to 0.74), respectively. The block effect introduced by the block-based side frame generation method [shown in Figs. 3(c) and 4(c)] is not observed in the reconstructed frame in Figs. 3(d) and 4(d) since the proposed fusion decoding algorithm operates in the measurement domain.

[2] *Joint GPSR* is the base line for *MC joint GPSR* which is used to validate the effectiveness of the proposed interview motion compensation based side frame.

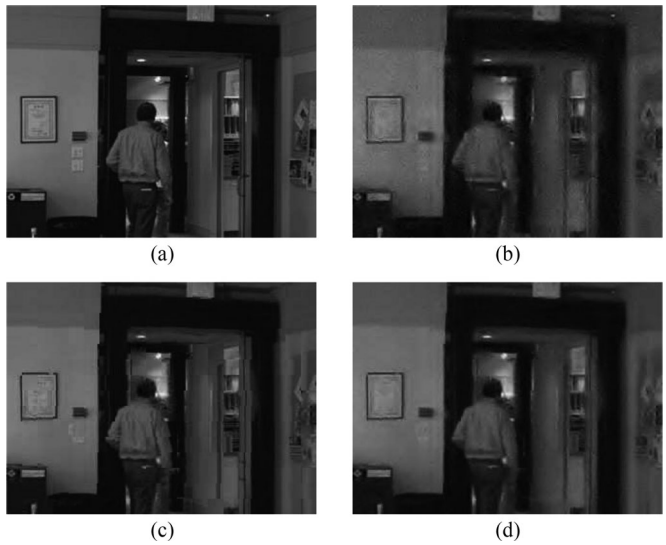


Fig. 3. (a) Original. (b) Independently reconstructed. (c) Generated side frame. (d) Fusion decoded 5th frame of *Exit*. Measurement rate is set to 0.2.

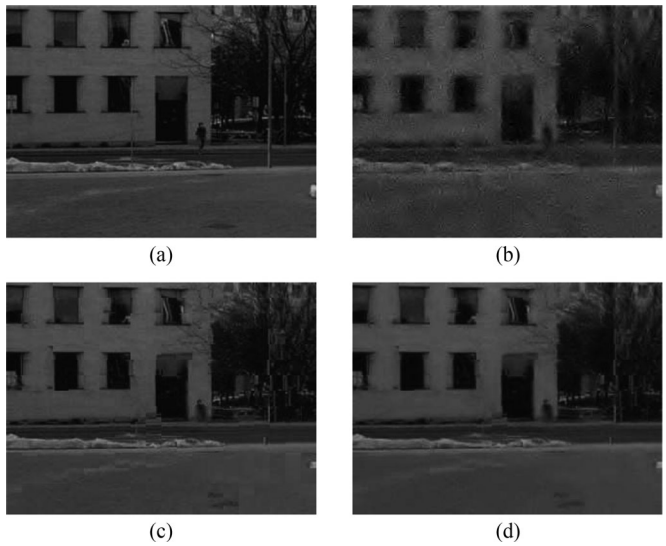


Fig. 4. (a) Original. (b) Independently reconstructed. (c) Generated side frame. (d) Fusion decoded 25th frame of *Vassar*. Measurement rate is set to 0.15.

Then, we consider the 4-view scenario, views 1, 2, 3 and 4. Without loss of the generality, view 2 is selected as K-view and the other three as CS-views. We then compare the achieved SSIM and PSNR for the first 50 frames of *Vassar*, *Exit*, *Ballroom*. We set three different CS-view measurement rates 0.3, 0.1 and 0.2 for *Vassar*, *Exit*, *Ballroom*, respectively. The results are illustrated in Figs. 5–7 with respect to PSNR and SSIM. We observe that the proposed *MC fusion* decoding method and

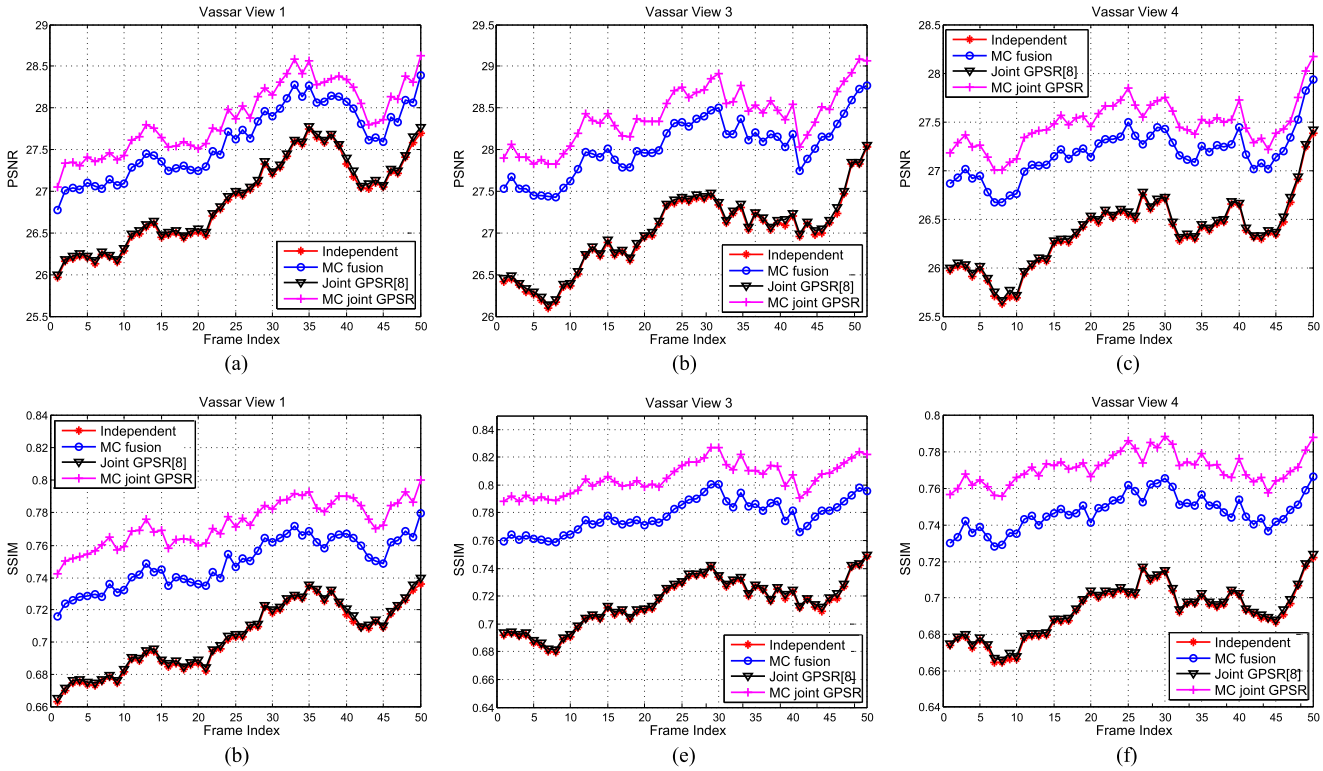


Fig. 5. PSNR comparison for CS-views: (a) view 1, (b) view 3, and (c) view 4. SSIM comparison for CS-views: (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.3 of *Vassar*.

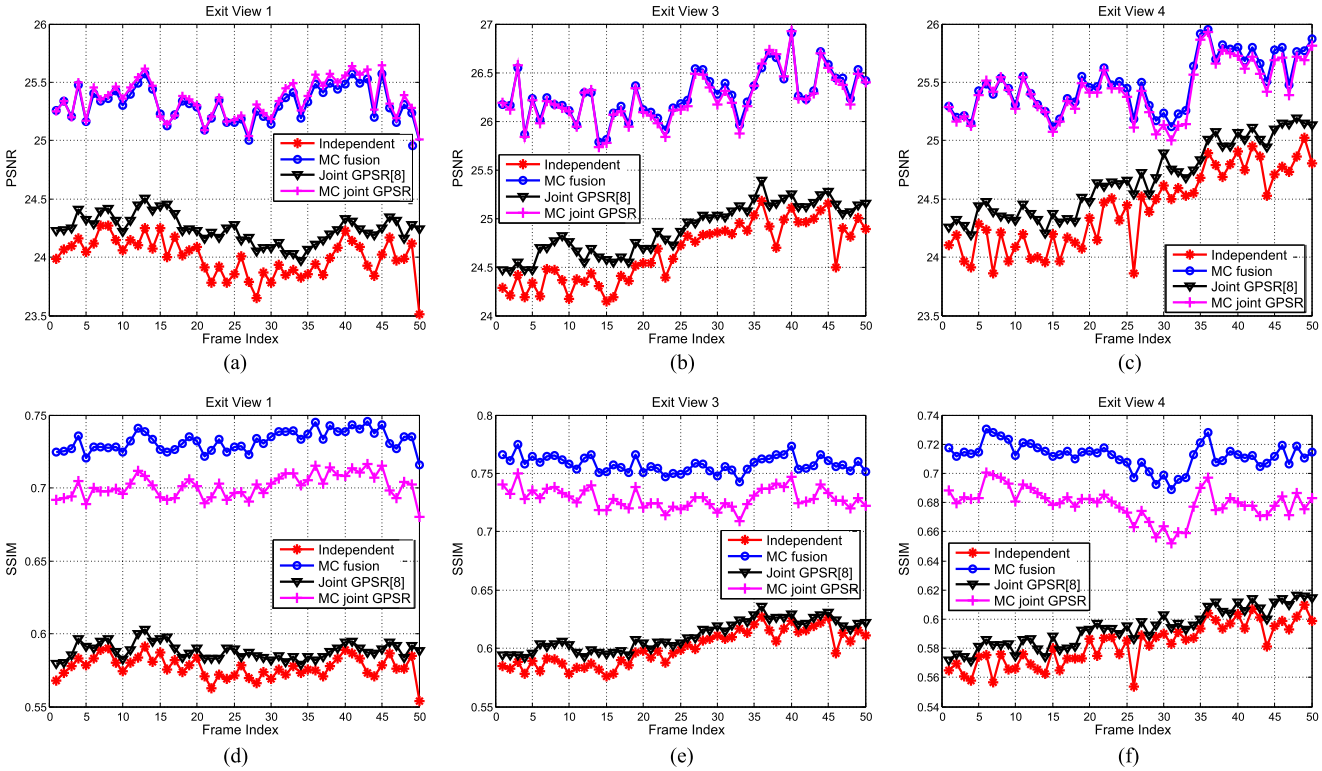


Fig. 6. PSNR comparison for CS-views: (a) view 1, (b) view 3, and (c) view 4. SSIM comparison for CS-views: (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.1 of *Exit*.

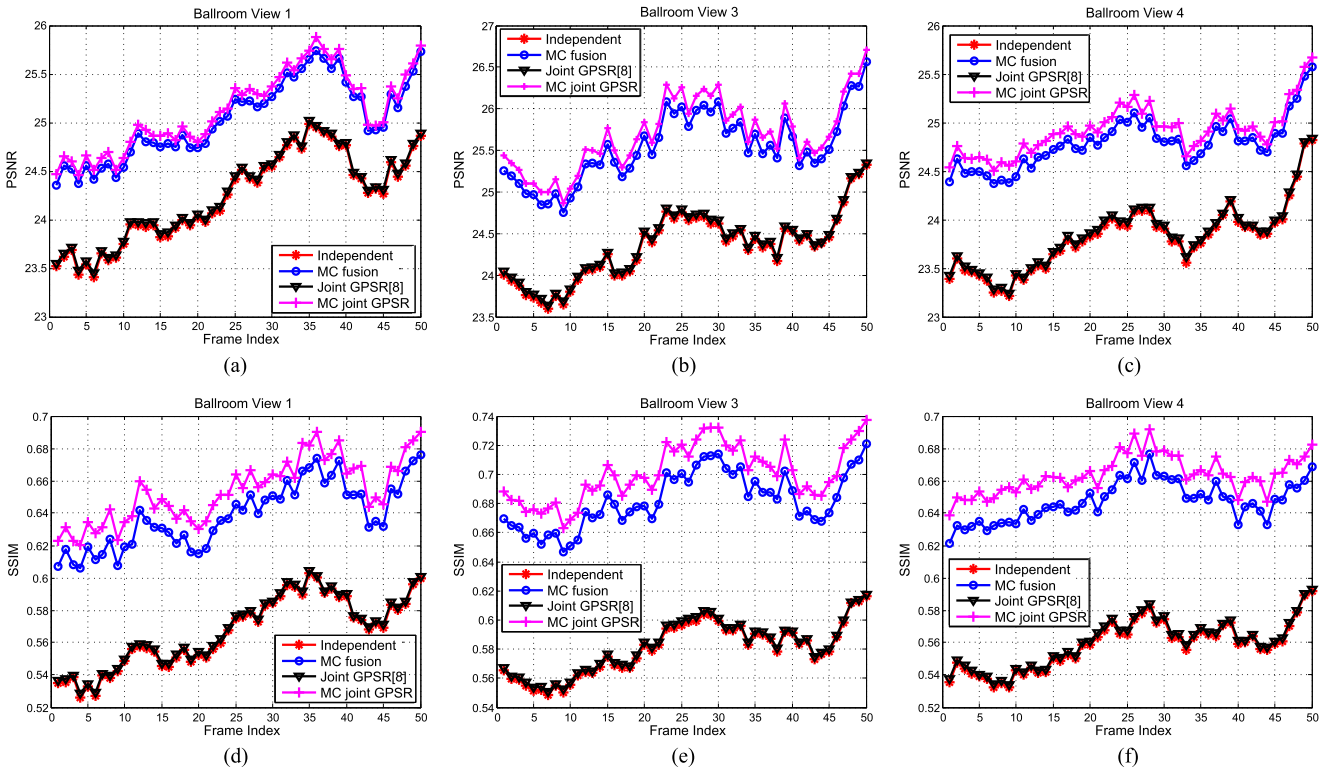


Fig. 7. PSNR comparison for CS-views: (a) view 1, (b) view 3, and (c) view 4. SSIM comparison for CS-views: (d) view 1, (e) view 3, and (f) view 4, with measurement rate 0.2 of *Ballroom*.

MC joint GPSR outperform significantly *joint GPSR* and *Independent* decoding approaches by up to 1.5 dB and 0.16 in terms of PSNR and SSIM, respectively. *MC fusion* (blue curve) and *MC joint GPSR* (pink curve) have similar performance for the tested three multi-view sequences. This observation demonstrates the effectiveness of the proposed fusion decoding method for CS-view; it also showcases the effectiveness of the side frame generated by the proposed inter-view motion compensated side frame. For the *Vassar* test sequence with CS-view encoding rate 0.3, *MC joint GPSR* is slightly better than *MC fusion* by no more than 0.3 dB and 0.03 in terms of PSNR and SSIM. Instead, for *Exit* with 0.1 encoding rate and *Ballroom* with 0.2 measurement rate sequences, *MC joint GPSR* and *MC fusion* achieve almost the same performance. We can also see that *joint GPSR* (black curve) proposed for single view video odd and even frames joint decoding just slightly outperforms *Independent* (red curve), which shows that *joint GPSR* is not suitable for the multi-view scenario and the importance of the side frame that acts as the initial point for the joint GRSR recovery algorithm.

Finally, to evaluate the proposed blind quality estimation method, we transmit the CS-view sequence over simulated time-varying channels with a randomly generated error pattern. The K-view is assumed to be correctly received and reconstructed. A setting similar to [21] is considered for CS-view transmission, i.e., the encoded CS-view measurements are first quantized and packetized. Then, parity bits are added to each packet. A packet is dropped at the receiver if detected to contain errors after a parity check. Here, we consider the *Ballroom* and *Exit* sequences as an example. The simulation result is depicted in Fig. 8, where the top figure refers to *Ballroom*, while the bottom refers to *Exit*. Different from the results in Figs. 6 and 7, where the measurement rate is set to 0.1 and 0.2, respectively, in Fig. 8, the actual received measurement rate is varying between 0.1 and 0.6 because of the randomly generated error pattern, which further results in varying PSNR. Through comparing the estimated

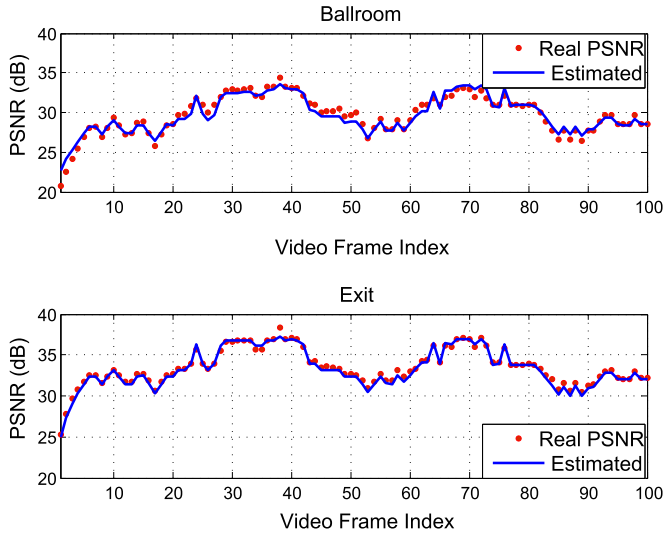


Fig. 8. Video quality estimation results for different video sequences: *Ballroom* (top) and *Exit* (bottom).

PSNR (blue line) with real PSNR (red dot) for 100 successive frames, we can conclude that the proposed blind estimation within our joint decoding of independently encoding framework is rather precise, with an estimation error of 4.32% for *Ballroom* and of 6.50% for *Exit*, respectively. With the proposed quality estimation approach, the receiver can provide precise feedback to the transmitter to guide dynamic rate adaptation.

VII. Conclusion

In this paper, we proposed an inter-view motion compensated side frame generation method for compressive multi-view video coding systems, and based on it, a novel fusion decoding approach for CS-view frame was developed. At the decoder end, a side frame is first generated and then resampled to obtain measurements and then appended after the received CS-view measurements. With the newly combined measurements, the state-of-the-art sparse signal recovery algorithm GPSR is used to obtain a final reconstructed CS-view frame. Extensive simulation results show that the proposed *MC fusion* decoder outperforms the independent CS-decoder in the case of fast-, moderate- and low-motion scenarios. The efficacy of the proposed side frame is also validated by adopting the existing *joint GPSR* with the proposed inter-view motion compensated side frame as the initial reconstruction point. Based on the proposed multi-view joint decoder, we also developed a video quality assessment metric (operating in the measurement domain) without reference frames for CS video systems. Experimental results with wireless video streaming scenario validated the accuracy of the proposed blind video quality estimation approach.

References

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

[2] C. Yan *et al.*, "A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 573–576, May 2014.

[3] C. Yan *et al.*, "Efficient parallel framework for HEVC motion estimation on many-Core processors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2077–2089, Dec. 2014.

[4] C. Yan *et al.*, "Parallel deblocking filter for HEVC on many-core processor," *Electron. Lett.*, vol. 50, no. 5, pp. 367–368, Feb. 2014.

[5] C. Yan *et al.*, "Efficient parallel HEVC intra-prediction on many-core processor," *Electron. Lett.*, vol. 50, no. 11, pp. 805–806, May 2014.

[6] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, no. 4, pp. 921–960, Mar. 2007.

[7] S. Pudlewski, N. Cen, Z. Guan, and T. Melodia, "Video transmission over lossy wireless networks: A cross-layer perspective," *IEEE J. Sel Topics Signal Process.*, vol. 9, no. 1, pp. 6–21, Jul. 2014.

[8] Z. Guan and T. Melodia, "Cloud-assisted smart camera networks for energy-efficient 3D video streaming," *IEEE Comput.*, vol. 47, no. 5, pp. 60–66, May 2014.

[9] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, Oct.–Dec. 2015.

[10] M. Budagavi *et al.*, "360 degrees video coding using region adaptive smoothing," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 750–754.

[11] 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.

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

[13] Y. Liu and D. A. Pados, "Compressed-sensed-domain L1-PCA video surveillance," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 351–363, Mar. 2016.

[14] H. Liu, B. Song, F. Tian, and H. Qin, "Joint sampling rate and bit-depth optimization in compressive video sampling," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1549–1562, Jun. 2014.

[15] C. Deng, W. Lin, B. S. Lee, and C. T. Lau, "Robust image coding based upon compressive sensing," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 278–290, Apr. 2012.

[16] M. Cossalter, G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Joint compressive video coding and analysis," *IEEE Trans. Multimedia*, vol. 12, no. 3, pp. 168–183, Apr. 2010.

[17] 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.*, Jun. 2015, pp. 137–146.

[18] 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.

[19] L.-W. Kang and C.-S. Lu, "Distributed compressive video sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1169–1172.

[20] 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.

[21] 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.

[22] 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.

[23] 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. 586–598, Dec. 2007.

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

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

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

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

[28] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[29] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[30] M. Saad, A. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, Mar. 2004.

[32] I. E. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, ser. SIAM Studies Appl. Math. Philadelphia, PA, USA: Soc. Ind. Appl. Math. 1994.

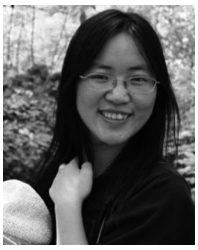
[33] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., Ser. B*, vol. 58, pp. 267–288, 1996.

[34] D. Donoho, M. Elad, and V. 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.

[35] K. Gao, S. Batalama, D. Pados, and B. Suter, "Compressive sampling with generalized polygons," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4759–4766, Oct. 2011.

[36] 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, Apr.–Jun. 2013.

[37] F. H. Jamil *et al.*, "Preliminary study of block matching algorithm (BMA) for video coding," in *Proc. Int. Conf. Mechatronics*, May 2011, pp. 1–5.
[38] A. M. Huang and T. Nguyen, "Motion vector processing using bidirectional frame difference in motion compensated frame interpolation," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2008, pp. 1–6.
[39] T. Koya, K. Lunuma, A. Hirano, Y. Lyima, and T. Ishi-guro, "Motion-compensated inter-frame coding for video conferencing," in *Proc. Nat. Telecommun. Conf.*, Nov. 1981, pp. G5.3.1–G5.3.5.



Nan Cen (S'09) received the B.S. and M.S. degrees in wireless communication engineering from the University of Shandong, Shandong, China, in 2008 and 2011, respectively, the M.S. degree in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 2014, and is currently working toward the Ph.D. degree in electrical and computer engineering at, Northeastern University, Boston, MA, USA.

She is currently working with the Wireless Networks and Embedded Systems Laboratory, Northeastern University, under the guidance of Prof. T. Melodia. Her current research interest focuses on wireless multiview video streaming based on compressed sensing.

Zhangyu Guan (S'09–M'11) received the Ph.D. degree in communication and information systems from Shandong University, Jinan, China, in 2010.

He is currently an Associate Research Scientist with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. He was previously a Visiting Ph.D. Student with the Department of Electrical Engineering, State University of New York (SUNY) at Buffalo, Buffalo, NY, USA, from 2009 to 2010. He was a Lecturer with Shandong University from 2011 to 2014. He was a Postdoctoral Research Associate with the Department of Electrical Engineering, SUNY Buffalo, from 2012 to 2015. His current research interests include cognitive and software-defined Internet of Things (IoT), wireless multimedia sensor networks, and underwater networks.

Dr. Guan has served as a TPC Member for IEEE INFOCOM 2016–2017, IEEE GLOBECOM 2015–2017, and IEEE ICNC 2012–2017, and served as a reviewer for the IEEE/ACM TRANSACTIONS ON NETWORKING and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, among others.

Tommaso Melodia (S'02–M'07–SM'16) received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2007.

He is an Associate Professor with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. He is serving as the lead PI on multiple grants from U.S. federal agencies including the National Science Foundation, the Air Force Research Laboratory, the Office of Naval Research, and the Army Research Office. His research focuses on modeling, optimization, and experimental evaluation of wireless networked systems, with applications to sensor networks and the Internet of Things, software-defined networking, and body area networks.

Prof. Melodia is an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON BIOLOGICAL, MOLECULAR, AND MULTI-SCALE COMMUNICATIONS, *Computer Networks*, and *Smart Health*. He will be the Technical Program Committee Chair for IEEE INFOCOM 2018. He is a recipient of the National Science Foundation CAREER award and of several other awards.