Reconstruction of Compressed-Sensed Multiview Video With Disparity- and Motion-Compensated Total Variation Minimization

Ying Liu, *Member, IEEE*, Dimitris A. Pados, *Senior Member, IEEE*, Joohee Kim, *Member, IEEE*, and Chen Zhang

***Abstract*— Compressed sensing (CS) is the theory and practice of sub-Nyquist sampling of sparse signals of interest. Exact reconstruction may then be possible with much fewer than the Nyquist-required number of data. In this paper, we consider a multiview video system in which multiple cameras at different locations perform independent CS to simultaneously capture different views of a scene. At the decoder, we propose a disparity- and motion-compensated total variation minimization algorithm to jointly reconstruct the multiview video sequence. The experimental results show that the proposed joint reconstruction algorithm successfully exploits simultaneously intra-frame, inter-frame, and inter-view sparsity and significantly outperforms existing independent-view reconstruction, residue-view reconstruction, and motion-adaptive reconstruction algorithms.**

***Index Terms*— 360° video, compressed sensing (CS), disparity compensation, immersive video, motion compensation, multiview video, sparse representation, total variation (TV) minimization.**

I. Introduction

MULTIVIEW video refers to video sequences of the same dynamic 3D scene captured simultaneously by multiple cameras from different viewpoints [1]. Together with new display technologies such as autostereoscopic displays and free viewpoint television [2], [3], the audience can experience a 3D depth impression of the observed scene or freely navigate within real-world visual scenes.

Compared with traditional monoview video systems, multiview video systems impose much higher storage and transmission cost due to the large amount of generated raw data. While joint predictive coding of data across camera nodes can be adopted [4]–[6], this approach is not suitable for prevalent wireless multimedia sensor network (WMSN)-enabled applications [7] (for example, security surveillance, object tracking, environmental monitoring, and industrial process control). In a typical WMSN, battery-powered miniature video cameras acquire video data in a distributed manner and forward to a multimedia sink (gateway) in a multihop fashion. Due to energy and memory constraints, the computational task at each camera sensor is expected to be simple. At the same time, performing joint-view predictive coding at the remote multimedia sink requires that all camera sensors in the network transmit wirelessly the captured raw uncompressed video data to the sink, which is arguably infeasible. Furthermore, predictive video encoding at the sink increases the impact of channel errors [7], [8] on the sink-to-consumer wireless link. In particular, when predictive coding is in use for scalable multiview video streaming [9], such as video conferencing, distance learning, remote presentation, and video-on-demand, video data are encoded in ordered layers. Loss of a lower layer in transmission renders higher layer data useless.

Recently, inspired by distributed source coding principles [10], [11] and compressed sensing (CS) [12]–[15], CS-based multiview video systems were proposed [16], [17], in which each camera independently captures and compresses one view of the same scene by taking a small number of (random [14] or deterministic [18]) linear measurements. High-quality reconstruction of the multiview video is achievable by sparsity-aware joint-view decoding at the receiver. The CS scheme integrates video acquisition and compression into a single step in the form of "compressive samples" [15] or measurements, which means practically zero computation at each camera node. At the same time, video source information is distributed among CS measurements of equal significance, and therefore, no single measurement is more important than others. Reconstruction of video requires a certain number of measurements to be available, but it does not require the availability of any particular measurement. In this sense, a measurement lost in transmission can simply be replaced by any other measurement. Furthermore, if the measurements of the video are transmitted by broadcast or multicast, a receiver in a higher capacity channel will have more measurements available and hence can reconstruct video of higher quality than a receiver in a lower capacity channel. These properties motivate multiview video coding using compressive sampling as a technology that is inherently energy efficient, error resilient, and scalable in channel capacity.

In this paper, we consider a distributed CS multiview video encoder and aim at developing an effective sparsity-aware decoder to recover the multiview video data from the CS measurements. As the video data originate from the same

Manuscript received July 5, 2016; revised October 18, 2016 and December 10, 2016; accepted January 15, 2017. Date of publication January 23, 2017; date of current version June 4, 2018. This work was supported in part by the National Science Foundation under Grant CNS-1422874. This paper was recommended by Associate Editor J. Yuan.

Y. Liu and D. A. Pados are with the Department of Electrical Engineering, University at Buffalo, SUNY, Buffalo, NY 14260 USA (e-mail: yl72@buffalo.edu; pados@buffalo.edu).

J. Kim and C. Zhang are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: joohee@ece.iit.edu; czhang57@hawk.iit.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/TCSVT.2017.2656920

scene, the inherent correlation of the multiview imagery can be exploited for sparse representation from three aspects: intra-frame correlation between adjacent pixels, inter-frame correlation between temporally successive images, and inter-view correlation between adjacent camera views. For the first time in the literature, we propose a *joint* decoder that exploits these three types of correlation *simultaneously*.

In our approach, all frames of all views are jointly recovered from independent compressive samples via disparity- and motion-compensated (DMC) total variation (TV) minimization. Initially, each frame of each view is reconstructed individually via two-dimensional TV (2D-TV) minimization. A group of disparity maps is then calculated for all the views at each time slot from the initial reconstructions and a DC prediction model is created in which each view is predicted by its neighbor view(s). Similarly, a set of motion fields is estimated for all frames at each view captured by a single camera and an MC prediction model is established in which each frame is predicted by its temporally adjacent frame(s). In the joint reconstruction stage, all frames of all views are recovered *simultaneously* using three sparsity penalty terms: 1) the original pixel-domain 2D-TV; 2) 2D-TV of the DC inter-view residue; and 3) 2D-TV of the MC temporal residue. The disparity- and motion-adaptation stage then iterates between disparity-estimation (DE), motion estimation (ME), and DMC TV minimization until no significant performance improvement can be further achieved. The contributions of our work lie in three aspects. First, we consider for the first time in the literature the joint reconstruction of compressed-sensed 4D multiview video data by simultaneously exploiting its intrinsic sparsity along all different dimensions. Second, we provide a practical algorithm to solve the formulated optimization problem. Specifically, we formulate the problem as minimizing an augmented Lagrangian function and solve via the alternating direction method of multipliers (ADMM) [19]–[21]. Finally, we successfully apply the new algorithm to benchmark multiview video sequence recovery. Extensive experimental results presented in this paper demonstrate the superiority of the proposed algorithm to state-of-the-art approaches.

The remainder of this paper is organized as follows. Section II is an overview of the related works on compressed-sensed image and video reconstruction, including multiview scenarios. In Section III, our simple distributed CS multiview video encoder system is described. In Section IV, three types of CS multiview video sparsity are introduced as the basis of the proposed decoder. In Section V, the proposed DMC joint TV minimization decoder is developed with a detailed description of the augmented Lagrangian method (ALM) with ADMM for solving the formulated optimization problem. The experimental results and performance analysis are presented in Section VI. Finally, a few conclusions are drawn in Section VII.

II. Related Works

Reconstruction of compressed-sensed image and video signals has been studied extensively in the literature. For instance, the intra-frame sparse representation is adopted for CS single image reconstruction, such as the 2D discrete cosine transform (2D-DCT) [22], the 2D discrete wavelet transform (2D-DWT), and TV minimization [23]–[25]. Besides, more advanced block-based sparse transforms are developed for single image CS reconstruction such as the following: 1) the collaborative sparsity in [26] which simultaneously enforces local 2D sparsity in the form of 2D-TV and nonlocal 3D sparsity in the form of a fixed 3D-DWT transform and 2) the group-based sparse representation developed in [27] that sparsely represents natural images in the group domain, which enforces intrinsic local sparsity and nonlocal self-similarity of images simultaneously, in a unified framework. Although promising, the adaptive sparse transform for each group of image patches therein [27] is pursued by dictionary learning, which highly increases the complexity of the algorithm.

While [26] and [27] offer state-of-the-art CS single image recovery and outperform fixed basis sparse transforms, they only exploit intra-frame sparsity. In the last decade, in pursue of CS monoview video recovery, inter-frame correlations are also well exploited to increase signal sparsity level in the form of 3D-DWT [28], 2D-DWT for frame difference data [29], and 3D-TV minimization [30] that extends 2D-TV minimization with an extra sparse temporal gradient constraint. Minimizing TV of the temporal 1D-DCT coefficients was demonstrated to be very effective for highly correlated video frames [30]. Independent CS acquisition with inter-frame TV minimization decoders was developed in [31] and [32].

Recently, motion-adaptive methods were developed to further exploit temporal sparsity in CS monoview video systems, including the block-adaptive Karhunen–Loève transform (KLT) basis [33] and the motion-adaptive spatiotemporal sparse regularization method [34], [35], which minimizes the transform-domain ℓ_1-norm of both video frames and the MC frame residue. The transform domain therein can be either 2D-DWT or TV domain. Nonetheless, these methods were developed only for monoview videos and using 2D-DWT as the spatial sparse transform is inefficient.

For multiview video sequences, correspondence between neighboring views captured by different cameras can be established and utilized for sparse representation. Such inter-view correlation has been extensively exploited for CS multiview still images [36]–[45]. These methods can be classified into three categories. The first category is to jointly recover all views by imposing a spatial joint-view sparsity with the aid of multi-camera geometrical transformations such as translation, rotation, and scaling [36], [37]. Algorithms in the second category recover view-difference data [38], [39] in which each view is initially recovered independently, followed by inter-view DE and DC predictions. The residue between each view and its DC prediction is then recovered using ℓ_1 minimization in a proper sparse transform domain such as 2D-DCT or 2D-TV and is added back to the prediction. The whole algorithm then iterates between DE and DC-residue reconstruction, until the performance converges or the maximum number of iterations is reached. In the third category, rather than recovering the DC-residue separately as approaches in the second category do, the

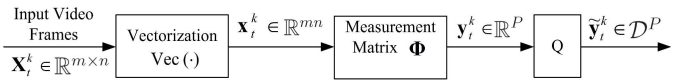


Fig. 1. Distributed CS of multiview video frames/views with quantization alphabet $\mathcal{D}$ (frame index t, view index k).

reconstruction algorithms simultaneously penalize the spatial sparsity and the DC-residue sparsity of all images. For instance, [40], [41] model DC-residue data as noise terms and impose ℓ_2-norm constraints on them, while [42]–[45] model DC-residue data as sparse outliers and impose ℓ_1-norm constraints. In particular, [45] offers state-of-the-art performance of this kind via joint TV minimization over all views and their DC-residue.

While [36]–[45] are for CS multiview still images only, in the problem of reconstructing CS multiview video signals, superior sparsity can be attained if inter-view similarity is exploited simultaneously with intra- and inter-frame similarities. In the most recent work [17], which is a direct extension of the CS multiview still image recovery algorithms in [38] and [39], the receiver uses intra-frame(view) similarity alone for an initial reconstruction. Then, disparity among views and motion among frames are estimated from the initial reconstruction and each view (frame) is predicted by the average of DC neighbor view(s) and MC adjacent frame(s). Finally, the prediction residue is recovered by TV minimization and is added back to the prediction. Although the approach exploits both inter-frame and inter-view correlations, the intra-frame(view) sparsity and inter-frame(view) correlations are considered serially rather than in parallel, and therefore, joint sparsity is not exploited and performance is not satisfactory for multiview video sequences with fast motion and large inter-view differences.

III. Distributed CS Acquisition of Multiview Video Sequences

In this paper, we propose a practical CS multiview video acquisition system that performs pure direct framewise encoding. In the encoding block diagram shown in Fig. 1, each frame/view $\mathbf{X}_t^k$ of size $m\times n$ is viewed as a vectorized column $\mathbf{x}_t^k \in \mathbb{R}^{mn}$, where t is the frame index and k is the view index. Compressive sampling is performed by projecting $\mathbf{x}_t^k$ onto a $P \times mn$ measurement matrix $\mathbf{\Phi}$, $P < mn$,

$$\mathbf{y}_t^k = \mathbf{\Phi}\mathbf{x}_t^k, \quad \mathbf{y}_t^k \in \mathbb{R}^P, \tag{1}$$

where $\mathbf{\Phi}$ is generated by arbitrarily permuting the columns of an order-s, $s \geq mn$, and multiple of four Walsh–Hadamard (WH) matrix followed by arbitrary selection of P rows (if $s > mn$, only mn arbitrary columns are utilized). This class of WH measurement matrices has the advantage of easy implementation (antipodal ± 1 entries), fast transformation, and satisfactory reconstruction performance, as we will see in the sequel. A richer class of matrices can be found in [46] and [47]. For practical implementation, $\mathbf{\Phi}$ is generated once and fixed for all frames and views in the video sequence. Each element of the resulting measurement vector $\mathbf{y}_t^k \in \mathbb{R}^P$ is quantized by an 8-bit uniform scalar quantization (block $\mathbf{Q}$ in Fig. 1). The resulting quantized values $\widetilde{\mathbf{y}}_t^k$ are then indexed and transmitted to the decoder.

IV. CS Multiview Video Sparsity

In this section, we introduce the concepts of TV-induced intra-frame sparsity, inter-view sparsity in terms of DC view residue, and inter-frame sparsity in terms of MC frame residue, which form the basis of our proposed reconstruction algorithm.

A. TV-Induced Intra-Frame Sparsity

Under the assumption that images are mostly piecewise smooth in the horizontal and vertical pixel directions, it is natural to consider utilizing the sparsity of the spatial gradient of an $m \times n$ image $\mathbf{X}_t^k$ (vectorized as $\mathbf{x}_t^k$) for CS image reconstruction [25], [48], [49]. If $x_{i,j}$[1] denotes the pixel in the ith row and jth column of $\mathbf{X}_t^k$, the discrete horizontal and vertical gradients at $x_{i,j}$ are defined, respectively, as

$$G_{h;ij}[\mathbf{x}_t^k] \triangleq \begin{cases} x_{i,j+1} - x_{i,j}, & j < n, \\ 0, & j = n, \end{cases}$$

and

$$G_{v;ij}[\mathbf{x}_t^k] \triangleq \begin{cases} x_{i+1,j} - x_{i,j}, & i < m, \\ 0, & i = m. \end{cases}$$

The spatial gradient of $\mathbf{x}_t^k$ at pixel $x_{i,j}$ can then be interpreted as the 2D vector

$$\mathbf{G}_{ij}[\mathbf{x}_t^k] = \begin{pmatrix} G_{h;ij}[\mathbf{x}_t^k] \\ G_{v;ij}[\mathbf{x}_t^k] \end{pmatrix}. \tag{2}$$

If we define now a matrix $\mathbf{G} \in \{+1, -1, 0\}^{2mn\times mn}$ such that the linear operation (matrix multiplication) $\mathbf{G}\mathbf{x}_t^k \in \mathbb{R}^{2mn}$ generates the vector concatenation of $\mathbf{G}_{ij}[\mathbf{x}_t^k]$ at all pixel locations (i, j), then the 2D anisotropic TV (ATV) of $\mathbf{x}_t^k$ is simply the ℓ_1-norm of $\mathbf{G}\mathbf{x}_t^k$, that is

$$\mathrm{ATV}(\mathbf{x}_t^k) \triangleq \|\mathbf{G}\mathbf{x}_t^k\|_1. \tag{3}$$

The 2D isotropic TV (ITV), in contrast, is defined as

$$\mathrm{ITV}(\mathbf{x}_t^k) \triangleq \sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{G}_{ij}\mathbf{x}_t^k\|_2. \tag{4}$$

Both ATV and ITV can serve as the sparse regularizer of an image. In the sequel, we adopt ATV for algorithmic development. Note that for any $m \times n$ image, the so-called "gradient operator" $\mathbf{G}$ is a fixed $2mn \times mn$ matrix that performs neighbor-pixel subtractions. Since $\mathbf{\Phi}$ and $\mathbf{G}$ are known, to reconstruct $\mathbf{x}_t^k$ from its dequantized CS measurements vector $\widehat{\mathbf{y}}_t^k$, we can solve the 2D-ATV minimization problem [24], [25]

$$\widehat{\mathbf{x}}_t^k = \arg\min_{\mathbf{x}_t^k} \left(\frac{1}{2}\|\widehat{\mathbf{y}}_t^k - \mathbf{\Phi}\mathbf{x}_t^k\|_2^2 + \mu\|\mathbf{G}\mathbf{x}_t^k\|_1 \right). \tag{5}$$

Now, let $\mathbf{x} \in \mathbb{R}^{pqmn}$ be the vector concatenation of p successive frames and q views $\mathbf{x}_t^k$, $t = 1, \ldots, p$, $k = 1, \ldots, q$. Then, the pq frames can be jointly recovered by solving the following convex optimization problem:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\frac{1}{2}\|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 + \mu\|\widetilde{\mathbf{G}}\mathbf{x}\|_1 \right) \tag{6}$$

[1]For simplicity in notation, the frame index t and view index k are dropped at this time.

where $\widetilde{\mathbf{G}}$ is the block diagonal matrix with the gradient operator $\mathbf{G}$ as the pq diagonal elements

$$\widetilde{\mathbf{G}} = \mathrm{diag}\{\mathbf{G} \ldots \mathbf{G}\}, \tag{7}$$

$\widehat{\mathbf{y}}$ is the concatenation of the dequantized measurement vectors $\widehat{\mathbf{y}}_t^k$, $t = 1, \ldots, p$, $k = 1, \ldots, q$, and $\widetilde{\mathbf{\Phi}}$ is the block diagonal matrix with pq diagonal elements

$$\widetilde{\mathbf{\Phi}} = \mathrm{diag}\{\mathbf{\Phi} \ldots \mathbf{\Phi}\}. \tag{8}$$

In short, the intra-frame sparsity penalty takes the form of the ℓ_1-norm of the spatial gradients over all frames and all views.

B. Disparity-Compensation-Induced Inter-View Sparsity

To model the inter-view correlations, we represent the multiview video frames captured by all cameras at a fixed time slot t in the form of a linear prediction system where each view is related to its immediate "left" and "right" neighbors via inter-view disparity compensation.

Define the kth view $\mathbf{X}_t^k$ as the base view, $1 < k \leq q$. Then, the geometry relation between $\mathbf{X}_t^k$ and its left neighbor $\mathbf{X}_t^{k-1}$ can be described with a horizontal disparity map $\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1} \in \mathbb{R}^{m\times n}$ and a vertical disparity map $\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1} \in \mathbb{R}^{m\times n}$, where the subscript k represents the base view index and the superscript $k-1$ represents the reference view index. With the aid of $\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}$ and $\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1}$, the pixel at position (i,j) in $\mathbf{X}_t^k$ can be predicted by a matching pixel at position $(i+\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1}(i,j),$ $j+\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}(i,j))$ in $\mathbf{X}_t^{k-1}$ in the form

$$\begin{aligned}&\mathbf{X}_t^k(i,j)_{\mathrm{left-predict}}\\ &\quad= \mathbf{X}_t^{k-1}\big(i+\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1}(i,j),\quad j+\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}(i,j)\big).\end{aligned} \tag{9}$$

Similarly, the geometry relation between the base view $\mathbf{X}_t^k$ and its right neighbor view $\mathbf{X}_t^{k+1}$ can be described with disparity maps $\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1} \in \mathbb{R}^{m\times n}$ such that the pixel at position (i,j) in the base view $\mathbf{X}_t^k$ can be predicted by a matching pixel at position $(i+\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1}(i,j), j+\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}(i,j))$ in its right neighbor $\mathbf{X}_t^{k+1}$ in the form of

$$\begin{aligned}&\mathbf{X}_t^k(i,j)_{\mathrm{right-predict}}\\ &\quad= \mathbf{X}_t^{k+1}\big(i+\boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1}(i,j),\quad j+\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}(i,j)\big).\end{aligned} \tag{10}$$

The following linear system gives a combined model for the linear measurements and inter-view relationship:

$$\widehat{\mathbf{y}}_t^k = \mathbf{\Phi}\mathbf{x}_t^k + \mathbf{e}_t^k, \tag{11}$$

$$\mathbf{x}_t^k = \frac{1}{2}\mathbf{D}_{t,k}^{k-1}\mathbf{x}_t^{k-1} + \frac{1}{2}\mathbf{D}_{t,k}^{k+1}\mathbf{x}_t^{k+1} + \mathbf{f}_t^k \tag{12}$$

where $\mathbf{D}_{t,k}^{k-1}$ and $\mathbf{D}_{t,k}^{k+1}$, which are the functions of $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1})$ and $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1})$, respectively, denote the left and right disparity compensation operators for the kth view at the tth time slot, leading to expressions (9) and (10) (see Fig. 2), and $\mathbf{f}_t^k$ denotes the DC prediction residue.

Consider now a group of q views, $\mathbf{x}_t = [\mathbf{x}_t^{1\mathrm{T}}, \ldots, \mathbf{x}_t^{q\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{qmn}$ (T is the transpose operator). The frame residue after bidirectional disparity compensation is

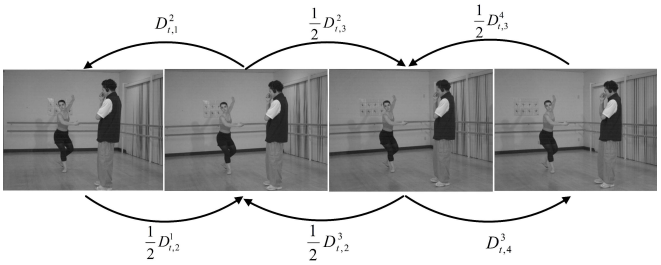


Fig. 2. Illustration of disparity compensation with $q = 4$ views.

$$\underbrace{\begin{bmatrix}\mathbf{f}_t^1\\ \mathbf{f}_t^2\\ \mathbf{f}_t^3\\ \vdots\\ \mathbf{f}_t^q\end{bmatrix}}_{\mathbf{f}_t} = \underbrace{\begin{bmatrix}\mathbf{I} & -\mathbf{D}_{t,1}^2 & & & \\ -\frac{1}{2}\mathbf{D}_{t,2}^1 & \mathbf{I} & -\frac{1}{2}\mathbf{D}_{t,2}^3 & & \\ & -\frac{1}{2}\mathbf{D}_{t,3}^2 & \mathbf{I} & -\frac{1}{2}\mathbf{D}_{t,3}^4 & \\ & & \cdots & & \\ & & & -\mathbf{D}_{t,q}^{q-1} & \mathbf{I}\end{bmatrix}}_{\mathbf{D}_t} \underbrace{\begin{bmatrix}\mathbf{x}_t^1\\ \mathbf{x}_t^2\\ \mathbf{x}_t^3\\ \vdots\\ \mathbf{x}_t^q\end{bmatrix}}_{\mathbf{x}_t} \tag{13}$$

where $\mathbf{D}_t$ is defined as the disparity compensation operator for all the views at time slot t. If DE is accurate enough, the inter-view prediction residue $\mathbf{f}_t^k$, $k = 1, \ldots, q$, shall also have small horizontal and vertical gradients, thereby its 2D-ATV can be utilized as a sparse penalty for CS multiview video reconstruction.

C. Motion-Compensation-Induced Inter-Frame Sparsity

Similar to inter-camera correlations, to model the temporal variations, we represent p successive frames captured by the kth camera (i.e., kth view) in the form of a linear prediction system where each frame is related to its "previous" and "next" adjacent frames via inter-frame motion compensation.

Define the tth frame $\mathbf{X}_t^k$ as the target frame, $1 < t \leq p$. Then, the relation between $\mathbf{X}_t^k$ and its preceding frame $\mathbf{X}_{t-1}^k$ can be described with a horizontal motion field $\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1} \in \mathbb{R}^{m\times n}$ and a vertical motion field $\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1} \in \mathbb{R}^{m\times n}$, where the subscript t represents the target frame index and the superscript $t-1$ represents the reference frame index. With the aid of $\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}$ and $\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1}$, the pixel at position (i,j) in $\mathbf{X}_t^k$ can be predicted forward by a matching pixel at position $(i+\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1}(i,j), j+\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}(i,j))$ in $\mathbf{X}_{t-1}^k$ in the form

$$\begin{aligned}&\mathbf{X}_t^k(i,j)_{\mathrm{forward-predict}}\\ &\quad= \mathbf{X}_{t-1}^k(i+\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1}(i,j),\quad j+\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}(i,j)).\end{aligned} \tag{14}$$

Similarly, the relation between the target frame $\mathbf{X}_t^k$ and its subsequent frame $\mathbf{X}_{t+1}^k$ can be described with horizontal and vertical motion fields $\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1} \in \mathbb{R}^{m\times n}$ such that the pixel at position (i,j) in $\mathbf{X}_t^k$ can be predicted backward by a matching pixel at position $(i+\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1}(i,j), j+\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}(i,j))$ in $\mathbf{X}_{t+1}^k$ in the form of

$$\begin{aligned}&\mathbf{X}_t^k(i,j)_{\mathrm{backward-predict}}\\ &\quad= \mathbf{X}_{t+1}^k\big(i+\boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1}(i,j),\quad j+\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}(i,j)\big).\end{aligned} \tag{15}$$

The following linear system gives a combined model for the linear measurements and inter-frame relationship:

$$\widehat{\mathbf{y}}_t^k = \mathbf{\Phi}\mathbf{x}_t^k + \mathbf{e}_t^k, \tag{16}$$

$$\mathbf{x}_t^k = \frac{1}{2}\mathbf{M}_t^{k,t-1}\mathbf{x}_{t-1}^k + \frac{1}{2}\mathbf{M}_t^{k,t+1}\mathbf{x}_{t+1}^k + \mathbf{g}_t^k \tag{17}$$

where $\mathbf{M}_t^{k,t-1}$ and $\mathbf{M}_t^{k,t+1}$, which are the functions of $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1})$ and $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1})$, respectively, denote the forward and backward motion compensation operators for the tth time slot at the kth camera, leading to expressions (14) and (15), and $\mathbf{g}_t^k$ denotes the MC prediction residue.

Consider a group of p successive frames at the kth view, $\mathbf{x}^k = [\mathbf{x}_1^{k\mathrm{T}}, \ldots, \mathbf{x}_p^{k\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{pmn}$. The prediction residue after bidirectional motion compensation is

$$\underbrace{\begin{bmatrix}\mathbf{g}_1^k\\ \mathbf{g}_2^k\\ \mathbf{g}_3^k\\ \vdots\\ \mathbf{g}_p^k\end{bmatrix}}_{\mathbf{g}^k} = \underbrace{\begin{bmatrix} \mathbf{I} & -\mathbf{M}_1^{k,2} & & & \\ -\frac{1}{2}\mathbf{M}_2^{k,1} & \mathbf{I} & -\frac{1}{2}\mathbf{M}_2^{k,3} & & \\ & -\frac{1}{2}\mathbf{M}_3^{k,2} & \mathbf{I} & -\frac{1}{2}\mathbf{M}_3^{k,4} & \\ & & \cdots & & \\ & & & -\mathbf{M}_p^{k,p-1} & \mathbf{I}\end{bmatrix}}_{\mathbf{M}^k} \underbrace{\begin{bmatrix}\mathbf{x}_1^k\\ \mathbf{x}_2^k\\ \mathbf{x}_3^k\\ \vdots\\ \mathbf{x}_p^k\end{bmatrix}}_{\mathbf{x}^k} \tag{18}$$

where $\mathbf{M}^k$ is the motion-compensation operator for the kth view. If motion estimation is accurate enough, the MC prediction residue $\mathbf{g}_t^k$, $t = 1, \ldots, p$, shall also have small horizontal and vertical gradients, thereby its 2D-ATV can be utilized as a sparse penalty for CS multiview video reconstruction.

V. Proposed Joint Reconstruction Algorithm

In this section, we describe and develop our proposed DMC ATV minimization decoder that jointly reconstructs all frames of all views. The proposed decoder consists of an initialization stage and a joint reconstruction stage. In particular, the joint reconstruction stage iterates between DE, ME, and DMC joint recovery.

Initially, each frame of each view is recovered independently from its CS measurements using 2D-ATV minimization as in (5). With the initially recovered frames $\widehat{\mathbf{x}}_{t,\mathrm{ini}}^k$, $t = 1, \ldots, p$, $k = 1, \ldots, q$, the disparity fields $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1})$ between the k and $(k-1)$ views and $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1})$ between the k and $(k+1)$ views, and the motion fields $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1})$ between the t and $(t-1)$ frames and $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1})$ between the t and $(t+1)$ frames can be estimated by the conventional variational optical flow (OF) method [50], [51].[2] As discussed in Sections IV-B and IV-C, the left and right disparity compensation operators, $\mathbf{D}_{t,k}^{k-1}$ and $\mathbf{D}_{t,k}^{k+1}$, can be defined as a function of $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1})$ and $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1})$, respectively, while the forward and backward motion compensation operators, $\mathbf{M}_t^{k,t-1}$ and $\mathbf{M}_t^{k,t+1}$, can be defined as a function of $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1})$ and $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1})$, respectively.

We consider next the joint reconstruction of a group of q views. Each view contains p frames and each frame has $m \times n$ pixels. The 4D multiview data is vectorized as $\mathbf{x} \in \mathbb{R}^{qpmn}$. We define disparity compensation of all images in the group as

$$\widetilde{\mathbf{D}}\mathbf{x} = \mathrm{diag}\{\mathbf{D}_1 \cdots \mathbf{D}_p\}\left[\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_p^{\mathrm{T}}\right]^{\mathrm{T}} \tag{19}$$

where $\mathbf{D}_t$ and $\mathbf{x}_t$, $t = 1, \ldots, p$ are defined in (13). Specifically, the elements in matrix $\mathbf{D}_t$ involve the precalculated $\mathbf{D}_{t,k}^{k-1}$ and $\mathbf{D}_{t,k}^{k+1}$, $k = 1, \ldots, q$, operators. Similarly, we define motion compensation of all images in the group as

$$\widetilde{\mathbf{M}}\mathbf{x} = \mathrm{diag}\{\mathbf{M}^1 \cdots \mathbf{M}^q\}[\mathbf{x}^{1\mathrm{T}}, \ldots, \mathbf{x}^{q\mathrm{T}}]^{\mathrm{T}} \tag{20}$$

where $\mathbf{M}^k$ and $\mathbf{x}^k$, $k = 1, \ldots, q$, are defined in (18). Again, the elements in matrix $\mathbf{M}^k$ involve the precalculated $\mathbf{M}_t^{k,t-1}$ and $\mathbf{M}_t^{k,t+1}$, $t = 1, \ldots, p$, operators. Then, we formulate our reconstruction problem as

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^{qpmn}} \Big(\frac{1}{2}\|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 + \mu_1\|\widetilde{\mathbf{G}}\mathbf{x}\|_1 + \mu_2\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}\|_1 + \mu_3\|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}\|_1\Big). \tag{21}$$

In the objective function in (21), the first term $\frac{1}{2}\|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2$ imposes data consistency. The three ℓ_1 norms, $\mu_1\|\widetilde{\mathbf{G}}\mathbf{x}\|_1$, $\mu_2\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}\|_1$, and $\mu_3\|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}\|_1$, promote intra-frame, inter-view, and inter-frame sparsity, respectively, with positive weights μ_i, $i = 1, 2, 3$. In particular, intra-frame sparse representation adopts 2D-ATV with gradient operator $\widetilde{\mathbf{G}}$ as in (6). Inter-view sparse representation is defined by applying the DC operation $\widetilde{\mathbf{D}}$ to all views, followed by calculating the 2D-ATV of the DC-residue via the gradient operator $\widetilde{\mathbf{G}}$. The reason for using 2D-ATV rather than using direct ℓ_1-norm of the DC-residue for sparse representation is that for small DC-residue it is very likely that 2D-ATV can significantly increase the degree of sparseness. Similarly, inter-frame sparse representation is defined by applying the MC operation $\widetilde{\mathbf{M}}$ to all frames followed by the 2D-ATV operation.

To solve the core minimization problem in (21), we follow the approach of TV minimization by augmented Lagrangian and alternating direction algorithms (TVAL3) [19], [55]–[57]. In this work, we first convert the problem in (21) into an equivalent variant through variable splitting by introducing the auxiliary variables $\mathbf{w}$, $\mathbf{u}$, and $\mathbf{v}$

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^{qpmn}} \frac{1}{2}\|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 + \mu_1\|\mathbf{w}\|_1 + \mu_2\|\mathbf{u}\|_1 + \mu_3\|\mathbf{v}\|_1 \quad \text{s.t. } \widetilde{\mathbf{G}}\mathbf{x} = \mathbf{w}, \widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x} = \mathbf{u}, \widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x} = \mathbf{v}. \tag{22}$$

The corresponding augmented Lagrangian function of (22) can

[2] Other ME (DE) approaches may also be adopted, such as the basic block matching method, the large displacement OF (LDOF) method [52], and the DeepFlow approach [53], which incorporates a descriptor matching scheme [54] in the LDOF algorithm and allows boosting performance in fast motions.

be formed as

$$\mathcal{L}_{\mathcal{A}}(\mathbf{w},\mathbf{u},\mathbf{v},\mathbf{x}) \tag{23}$$
$$= \mu_1\|\mathbf{w}\|_1 - \boldsymbol{\lambda}_1^T(\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}) + \frac{\beta_1}{2}\|\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}\|_2^2 \tag{23.a}$$
$$+ \mu_2\|\mathbf{u}\|_1 - \boldsymbol{\lambda}_2^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}) + \frac{\beta_2}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}\|_2^2 \tag{23.b}$$
$$+ \mu_3\|\mathbf{v}\|_1 - \boldsymbol{\lambda}_3^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}) + \frac{\beta_3}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}\|_2^2 \tag{23.c}$$
$$+ \frac{1}{2}\|\widehat{\mathbf{y}}-\widetilde{\boldsymbol{\Phi}}\mathbf{x}\|_2^2 \tag{23.d}$$

where β_1, β_2, and β_3 are tuning parameters.

The basic idea behind minimizing the augmented Lagrangian function is to seek a saddle point of $\mathcal{L}_{\mathcal{A}}(\mathbf{w},\mathbf{u},\mathbf{v},\mathbf{x})$, which is also a solution to problem (22). We minimize $\mathcal{L}_{\mathcal{A}}(\mathbf{w},\mathbf{u},\mathbf{v},\mathbf{x})$ by iteratively, $\ell = 0,1,\ldots$, solving for the variables

$$(\mathbf{w}^{\ell+1},\mathbf{u}^{\ell+1},\mathbf{v}^{\ell+1},\mathbf{x}^{\ell+1}) = \arg\min_{\mathbf{w},\mathbf{u},\mathbf{v},\mathbf{x}} \mathcal{L}_{\mathcal{A}}(\mathbf{w},\mathbf{u},\mathbf{v},\mathbf{x}) \tag{24}$$

and the Lagrangian multipliers

$$\boldsymbol{\lambda}_1^{\ell+1} = \boldsymbol{\lambda}_1^{\ell} - \beta_1(\widetilde{\mathbf{G}}\mathbf{x}^{\ell+1} - \mathbf{w}^{\ell+1}), \tag{25}$$
$$\boldsymbol{\lambda}_2^{\ell+1} = \boldsymbol{\lambda}_2^{\ell} - \beta_2(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}^{\ell+1} - \mathbf{u}^{\ell+1}), \tag{26}$$
$$\boldsymbol{\lambda}_3^{\ell+1} = \boldsymbol{\lambda}_3^{\ell} - \beta_3(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}^{\ell+1} - \mathbf{v}^{\ell+1}). \tag{27}$$

Following ADMM principles, (24) is decomposed into four subproblems associated with $\mathbf{w}^{\ell+1}$, $\mathbf{u}^{\ell+1}$, $\mathbf{v}^{\ell+1}$, and $\mathbf{x}^{\ell+1}$, respectively. Each subproblem solves for one variable assuming the other three variables are known. For simplicity in notation, the iteration index ℓ is omitted.

A. $\mathbf{w}$ *Subproblem,* $\mathbf{u}$ *Subproblem, and* $\mathbf{v}$ *Subproblem*

The subproblems associated with $\mathbf{w}$, $\mathbf{u}$, and $\mathbf{v}$ can be simplified as

$$\min_{\mathbf{w}} \frac{\beta_1}{2}\|\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}\|_2^2 - \boldsymbol{\lambda}_1^T(\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}) + \mu_1\|\mathbf{w}\|_1, \tag{28}$$
$$\min_{\mathbf{u}} \frac{\beta_2}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}\|_2^2 - \boldsymbol{\lambda}_2^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}) + \mu_2\|\mathbf{u}\|_1, \tag{29}$$

and

$$\min_{\mathbf{v}} \frac{\beta_3}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}\|_2^2 - \boldsymbol{\lambda}_3^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}) + \mu_3\|\mathbf{v}\|_1. \tag{30}$$

Since (28)–(30) are of the same form, by soft thresholding [21] their closed-form solutions are given by

$$\widetilde{\mathbf{w}} = \max\left\{\left|\widetilde{\mathbf{G}}\mathbf{x} - \frac{\boldsymbol{\lambda}_1}{\beta_1}\right| - \frac{\mu_1}{\beta_1}, \mathbf{0}\right\} \cdot \operatorname{sgn}\left(\widetilde{\mathbf{G}}\mathbf{x} - \frac{\boldsymbol{\lambda}_1}{\beta_1}\right), \tag{31}$$

$$\widetilde{\mathbf{u}} = \max\left\{\left|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x} - \frac{\boldsymbol{\lambda}_2}{\beta_2}\right| - \frac{\mu_2}{\beta_2}, \mathbf{0}\right\} \cdot \operatorname{sgn}\left(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x} - \frac{\boldsymbol{\lambda}_2}{\beta_2}\right), \tag{32}$$

and

$$\widetilde{\mathbf{v}} = \max\left\{\left|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x} - \frac{\boldsymbol{\lambda}_3}{\beta_3}\right| - \frac{\mu_3}{\beta_3}, \mathbf{0}\right\} \cdot \operatorname{sgn}\left(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x} - \frac{\boldsymbol{\lambda}_3}{\beta_3}\right). \tag{33}$$

B. $\mathbf{x}$ *Subproblem*

With the aid of $\mathbf{w}$, $\mathbf{u}$, $\mathbf{v}$, the $\mathbf{x}$ subproblem is equivalent to

$$\begin{aligned}\min_{\mathbf{x}} &- \boldsymbol{\lambda}_1^T(\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}) + \frac{\beta_1}{2}\|\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}\|_2^2 \\ &- \boldsymbol{\lambda}_2^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}) + \frac{\beta_2}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}\|_2^2 \\ &- \boldsymbol{\lambda}_3^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}) + \frac{\beta_3}{2}\|\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}\|_2^2 \\ &+ \frac{1}{2}\|\widehat{\mathbf{y}}-\widetilde{\boldsymbol{\Phi}}\mathbf{x}\|_2^2. \end{aligned} \tag{34}$$

Clearly, (34) is a quadratic function with gradient

$$\begin{aligned}\mathbf{d} = &-\widetilde{\mathbf{G}}^T\boldsymbol{\lambda}_1 + \beta_1\widetilde{\mathbf{G}}^T(\widetilde{\mathbf{G}}\mathbf{x}-\mathbf{w}) \\ &-(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}})^T\boldsymbol{\lambda}_2 + \beta_2(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}})^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}-\mathbf{u}) \\ &-(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}})^T\boldsymbol{\lambda}_3 + \beta_3(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}})^T(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}}\mathbf{x}-\mathbf{v}) \\ &+\widetilde{\boldsymbol{\Phi}}^T(\widetilde{\boldsymbol{\Phi}}\mathbf{x}-\widehat{\mathbf{y}}) \end{aligned} \tag{35}$$

and Hessian

$$\begin{aligned}\mathbf{H} = &\beta_1\widetilde{\mathbf{G}}^T\widetilde{\mathbf{G}} + \beta_2(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}})^T\widetilde{\mathbf{G}}\widetilde{\mathbf{D}} \\ &+ \beta_3(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}})^T\widetilde{\mathbf{G}}\widetilde{\mathbf{M}} + \widetilde{\boldsymbol{\Phi}}^T\widetilde{\boldsymbol{\Phi}}. \end{aligned} \tag{36}$$

Thus, setting $\mathbf{d} = \mathbf{0}$ can give us the exact minimizer of problem (34) [21], which is

$$\begin{aligned}\mathbf{x} = \mathbf{H}^{-1}(&\widetilde{\mathbf{G}}^T\boldsymbol{\lambda}_1 + (\widetilde{\mathbf{G}}\widetilde{\mathbf{D}})^T\boldsymbol{\lambda}_2 + (\widetilde{\mathbf{G}}\widetilde{\mathbf{M}})^T\boldsymbol{\lambda}_3 + \beta_1\widetilde{\mathbf{G}}^T\mathbf{w} \\ &+ \beta_2(\widetilde{\mathbf{G}}\widetilde{\mathbf{D}})^T\mathbf{u} + \beta_3(\widetilde{\mathbf{G}}\widetilde{\mathbf{M}})^T\mathbf{v} + \widetilde{\boldsymbol{\Phi}}^T\widehat{\mathbf{y}}). \end{aligned} \tag{37}$$

Since computing $\mathbf{H}^{-1}$ at each iteration is too costly to implement, an iterative method is highly desirable. In our presented work, the steepest descent method with optimal step is used to solve (34) iteratively by applying

$$\widetilde{\mathbf{x}} = \mathbf{x} - \eta\mathbf{d} \tag{38}$$

where the optimal step size is

$$\eta = \left|\frac{\mathbf{d}^T\mathbf{d}}{\mathbf{d}^T\mathbf{H}\mathbf{d}}\right|. \tag{39}$$

Thus, solving for $\mathbf{x}$ involves only computing (38) from iteration to iteration.

Assume now that the solution $\widehat{\mathbf{x}}$ of (21) that contains p frames from q views is obtained. Since joint decoding relies on accurate DC and MC operations $\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{M}}$, which are functions of the estimated disparity fields $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1})$, $(\boldsymbol{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}, \boldsymbol{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1})$ and motion fields $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1})$, $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1})$, we can use the jointly decoded high-quality frames to recalculate the disparity and motion fields, again using the OF method, and update the corresponding DC and MC operators. With the updated DC and MC operators $\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{M}}$, we rerun the joint decoding algorithm in (21). The decoding algorithm then iterates between DE (ME) with current recovered frames and DMC ATV minimization joint reconstruction. The overall algorithm is summarized in Table I.

TABLE I
PROPOSED (DMC-ATV2D) RECONSTRUCTION ALGORITHM

Input: $\widehat{\mathbf{y}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$, $\widetilde{\mathbf{\Phi}}$.
Initialization: Solve (5) to obtain the initial frame estimates $\widehat{\mathbf{x}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$.
Iterative Joint Decoding:
1: **for** iteration $= 1$ to max iteration
2: Estimate disparity fields $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k-1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k-1})$ and $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k+1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k+1})$ from $\widehat{\mathbf{x}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$, (for example, by optical flow method [50], [51]).
3: Estimate motion fields $(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ and $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$ from $\widehat{\mathbf{x}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$.
4: Update DC operator $\widetilde{\mathbf{D}}$ with new $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k-1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k-1})$ and $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k+1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k+1})$.
5: Update MC operator $\widetilde{\mathbf{M}}$ with new $(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ and $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$.
6: Solve (21) with new $\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{M}}$ to update $\widehat{\mathbf{x}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$.
7: **end for**
Output: $\widehat{\mathbf{x}}_t^k$, $t = 1, ..., p$, $k = 1, ..., q$.

VI. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

A. Rate–Distortion Performance on Five Data Sets

In this section, we study experimentally the performance of the proposed CS multiview video decoders by evaluating the peak signal-to-noise ratio (PSNR) (as well as the perceptual quality) of reconstructed video sequences. Five data sets, *Bookarrival*, *Balloons*, *Ballet*, *Breakdancer*, and *Kendo*, each with resolution of 768×1024 pixels, are used. These data sets are frequently utilized in the literature to evaluate the performance of multiview video coding schemes [58]–[66]. *Balloons* and *Kendo* [58]–[63] are two test sequences from the HEVC reference software. *Ballet* and *Breakdancer* [64]–[66] are provided by Microsoft Research [67]. *Bookarrival* is released by the Heinrich Hertz Institute (HHI) [68]. Each of the sequences consists of $q = 6$ views captured by different cameras and each view has a total number of $T = 60$ frames. For *Bookarrival* and *Balloons*, the difference between adjacent views is relatively small and the motion along the temporal direction is relatively slow. *Kendo* has moderate motion and small inter-view differences. On the other hand, *Ballet* and *Breakdancer* have relatively large inter-view differences and fast motion. Processing is carried out only on the luminance component.

At our independent distributed CS encoder side, each frame of each view is handled as a vectorized column of length $N = 768 \times 1024 = 786432$ multiplied by a $P \times N$ randomized partial WH matrix $\mathbf{\Phi}$, where P is the resulting number of CS measurements per frame per view. The sensing matrix $\mathbf{\Phi}$ is generated only once to encode all frames in the video sequence. In our experiments, the CS ratios are set at $\frac{P}{N} = 0.0469, 0.0625, 0.125, 0.25$, and 0.375. The elements of the captured P-dimensional measurement vector of each frame are quantized by an 8-bit uniform scalar quantizer. Hence, the resulting number of encoding bits is $\frac{P}{N} \times 8 = 0.3752, 0.5, 1, 2$, and 3 bits per pixel (bpp), correspondingly. To save storage space at both the encoder and decoder, we avoid using different sensing matrices for different frames of different views. As will be demonstrated later in all experiments, the adopted fixed sensing matrix produces satisfactory reconstruction quality and suffices to illustrate the superiority of the proposed joint reconstruction algorithm. We emphasize that our encoder has purposefully lowest complexity for energy and cost savings in WMSN applications. Therefore, we do not intend to compare its coding efficiency with the state-of-the-art high efficiency video coding (HEVC) codec [69], [70]. While HEVC achieves high compression ratio, its high encoder complexity makes it unsuitable for power-constrained applications in WMSNs.

At the decoder side, for each independent 4D data cube of size $768 \times 1024 \times (p = 6 \text{ frames}) \times (q = 6 \text{ views})$, we collect the corresponding CS measurements and reconstruct the video sequence by the proposed joint reconstruction algorithm. For each test sequence, we independently reconstruct $\frac{T=60 \text{ frames}}{p=6 \text{ frames}} = 10$ such 4D data cubes. In our experimental studies, seven reconstruction algorithms are examined for the same low-complexity CS encoding/acquisition system:

1) the proposed DMC 2D ATV minimization (DMC-ATV2D) joint decoder;
2) the DMC 2D ITV minimization (DMC-ITV2D) joint decoder[3];
3) the DMC residue (DMC-residue) decoder of [17];
4) the MC ATV (MC-ATV2D) minimization decoder [34];
5) the intra-frame ATV minimization (Intra-ATV2D) decoder [25];
6) the MC 2D-DWT (MC-DWT) decoder [34];
7) the intra-frame group sparse representation (Intra-GSR) decoder [27].

Fig. 3 shows the enlarged decodings of *Bookarrival* (3rd frame of the 3rd view) [Fig. 3(a)] at a bit rate of 0.5 bpp produced by the proposed DMC-ATV2D decoder [Fig. 3(b)], DMC-ITV2D [Fig. 3(c)], DMC-residue [Fig. 3(d)] [17], MC-ATV2D [Fig. 3(e)] [34], Intra-ATV2D [Fig. 3(f)] [25], MC-DWT [Fig. 3(g)] [34], and Intra-GSR [Fig. 3(h)] [27]. It can be observed that the texture details of the scene are best preserved by the proposed DMC-ATV2D decoder. Meanwhile, the DMC-residue and MC-DWT decodings have noticeable noise artifacts, MC-ATV2D and Intra-ATV2D decoders have blurry recovery, and Intra-GSR decoding has severe distortions on texture details. Compared with DMC-ITV2D, DMC-ATV2D presents clearer edges while slightly introducing stair artifact;

[3]To create DMC-ITV2D, we replace the ATV in the objective function of (21) by the ITV.

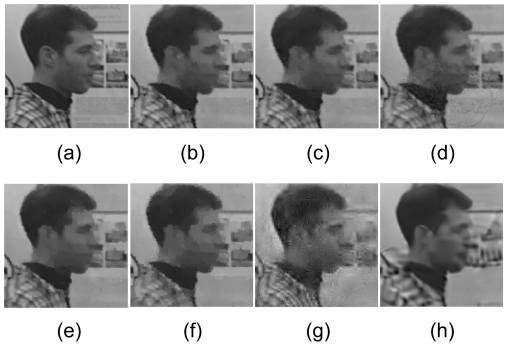


Fig. 3. Enlarged decodings of the *Bookarrival* data set (0.5 bpp): (a) Original 3rd frame of the 3rd view; (b) proposed DMC-ATV2D decoder (PSNR = 36.01 dB); (c) DMC-ITV2D (PSNR = 35.46 dB); (d) DMC-residue [17] (PSNR = 34.44 dB); (e) MC-ATV2D [34] (PSNR = 34.14 dB); (f) ATV2D [25] (PSNR = 33.05 dB); (g) MC-DWT [34] (PSNR = 29.20 dB); and (h) Intra-GSR [27] (PSNR = 32.89 dB).

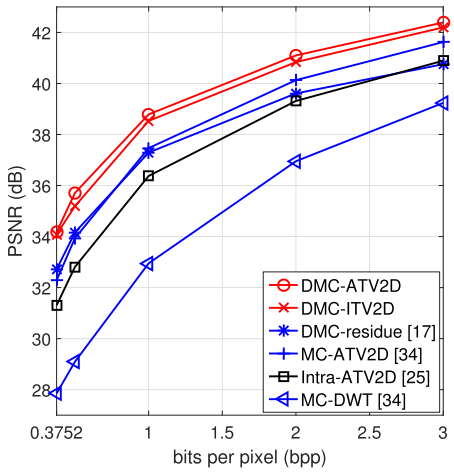


Fig. 4. Rate–distortion comparison for the *Bookarrival* data set.

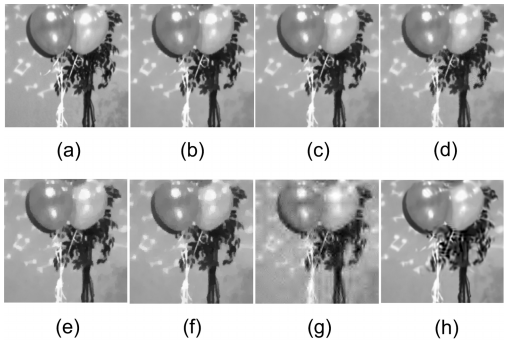


Fig. 5. Enlarged decodings of the *Balloons* data set (0.5 bpp): (a) Original 3rd frame of the 3rd view; (b) proposed DMC-ATV2D decoder (PSNR = 37.85 dB); (c) DMC-ITV2D (PSNR = 37.41 dB); (d) DMC-residue [17] (PSNR = 36.45 dB); (e) MC-ATV2D [34] (PSNR = 34.97 dB); (f) ATV2D [25] (PSNR = 33.89 dB); (g) MC-DWT [34] (PSNR = 29.50 dB); and (h) Intra-GSR [27] (PSNR = 35.00 dB).

DMC-ITV2D avoids the stair artifact while slightly increases blurriness. To quantify the reconstruction performance, the PSNR values are given in the caption of Fig. 3. The proposed DMC-ATV2D decoder has the highest PSNR of 36.01 dB (DMC-ITV2D has a PSNR of 35.46 dB).

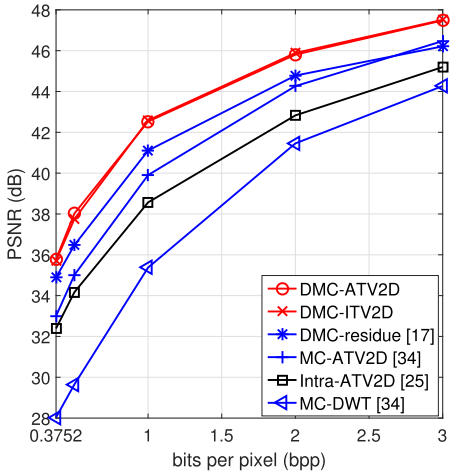


Fig. 6. Rate–distortion comparison for the *Balloons* data set.

Fig. 4 presents the rate–distortion characteristics of the six decoders for the *Bookarrival* data set. The PSNR values (dB) are averaged over all frames of all views. We note that the PSNR curve of Intra-GSR [27] is not provided in Fig. 4, due to the high complexity (time consumption) of the dictionary training process in this algorithm. From Fig. 3(h), we have already demonstrated that the Intra-GSR decoded image has a significantly lower PSNR (32.89 dB) compared with the proposed DMC-ATV2D decoder (36.01 dB), as well as unsatisfactory perceptual quality in view of severe distortion around object edges. From Fig. 4, we observe that the DMC-residue [17] and MC-ATV2D [34] decoders outperform Intra-ATV2D [25] as they both exploit joint sparsity to some extent. Nevertheless, by utilizing the sparsity of the MC and DC-residue and the pixel-domain 2D ATV simultaneously, our proposed DMC-ATV2D decoder further improves the reconstruction PSNR by about 1 to 1.8 dB compared with the DMC-residue and MC-ATV2D decoders. Further, we observe from Fig. 4 that by modifying the proposed DMC-ATV2D to adopt the ITV instead of the ATV, the obtained DMC-ITV2D scheme has a slightly lower PSNR than DMC-ATV2D. The very same experiments of Figs. 3 and 4 are repeated in Figs. 5 and 6 for the *Balloons* sequence. Similar conclusions can be drawn.

For increased credibility of our study, we also experiment with the *Ballet* [Figs. 7 and 10(a)], *Breakdancer* [Figs. 8 and 10(b)], and *Kendo* [Figs. 9 and 10(c)] sequences. Once again, the proposed DMC-ATV2D (DMC-ITV2D) decoder offers consistently better reconstruction quality than the other five decoders. Since *Ballet* and *Breakdancer* sequences have faster motion and large inter-view differences, the DMC-residue [17] decoder that utilizes limited amount of signal sparsity offers better performance than the Intra-ATV2D decoder [25] at low bit rate (bpp) and deteriorates at medium-to-high bit rates (bpp). In contrast, the proposed DMC-ATV2D (DMC-ITV2D) decoder is superior to all decoders at all bit rates (bpp). Specifically, we demonstrate the effectiveness of the proposed DMC-ATV2D decoder for low bit rate scenarios

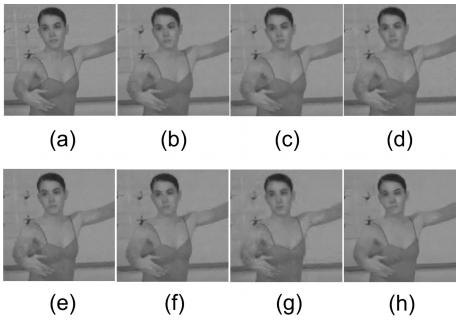


Fig. 7. Enlarged decodings of the *Ballet* data set (1 bpp): (a) Original 3rd frame of the 3rd view; (b) proposed DMC-ATV2D decoder (PSNR = 39.58 dB); (c) DMC-ITV2D (PSNR = 39.2527 dB); (d) DMC-residue [17] (PSNR = 38.72 dB); (e) MC-ATV2D [34] (PSNR = 38.77 dB); (f) ATV2D [25] (PSNR = 38.35 dB); (g) MC-DWT [34] (PSNR = 36.39 dB); and (h) Intra-GSR [27] (PSNR = 39.32 dB).

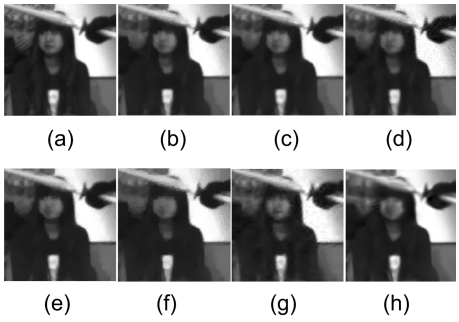


Fig. 9. Enlarged decodings of the *Kendo* data set (0.5 bpp): (a) Original 3rd frame of the 3rd view; (b) proposed DMC-ATV2D decoder (PSNR = 40.49 dB); (c) DMC-ITV2D (PSNR = 41.13 dB); (d) DMC-residue [17] (PSNR = 39.61 dB); (e) MC-ATV2D [34] (PSNR = 39.82 dB); (f) ATV2D [25] (PSNR = 36.88 dB); (g) MC-DWT [34] (PSNR = 37.06 dB); and (h) Intra-GSR [27] (PSNR = 37.65 dB).

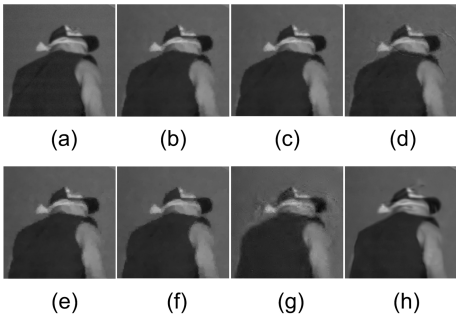


Fig. 8. Enlarged decodings of the *Breakdancer* data set (0.5 bpp): (a) Original 3rd frame of the 3rd view; (b) proposed DMC-ATV2D decoder (PSNR = 37.01 dB); (c) DMC-ITV2D (PSNR = 36.99 dB); (d) DMC-residue [17] (PSNR = 36.39 dB); (e) MC-ATV2D [34] (PSNR = 36.46 dB); (f) ATV2D [25] (PSNR = 35.43 dB); (g) MC-DWT [34] (PSNR = 33.63 dB); and (h) Intra-GSR [27] (PSNR = 36.07 dB).

in Fig. 11 on the *Ballet* data set. We observe that for low bit rates (0.188–0.35 bpp), the proposed DMC-ATV2D decoder again offers the highest reconstruction PSNR, similar to higher bit rate scenarios [$0.3752 \leq \text{bpp} \leq 3$ in Fig. 10(a)].

To further study the proposed iterative scheme, we plot the PSNR versus iteration step curves for the proposed DMC-ATV2D method, DMC-residue [17], and MC-ATV2D [34] methods in Figs. 12 and 13 for the *Ballet* data set at 0.25 bpp and the *Kendo* data set at 1 bpp, respectively. This is indeed a fair comparison study because all three decoders initialize the decoding procedure with the same intra-frame 2D-ATV minimization reconstruction, followed by their own iterative decoding scheme. We observe that all three decoders have improved reconstruction quality measured in PSNR (dB) as the number of iteration increases, among which the proposed DMC-ATV2D decoder converges to the highest PSNR value.

To evaluate the subjective video quality of the proposed DMC-ATV2D decoder and compare it with that of other algorithms, in addition to the visual results shown in Figs. 3, 5, and 7–9, we also adopt the three-component structural similarity (3SSIM) index proposed in [71]. SSIM is an effective and computationally efficient perceptual image/video quality assessment metric [72]. 3SSIM is a variant of SSIM that takes into account the fact that the human eye is more sensitive to differences in textured or edge regions than smooth regions. The overall metric value is calculated as a weighted average of the SSIM of three categories of regions: edges, textures, and smooth regions. 3SSIM is considered to give results that are highly consistent with human subjective perception. Fig. 14 shows the 3SSIM curves on *Bookarrival* and *Ballet* as representative of slow-motion and fast-motion sequences, respectively. We observe that the proposed DMC-ATV2D (DMC-ITV2D) decoder achieves the highest 3SSIM index, especially for small bpp values.

B. Robustness to Occlusions or Corruptions

We stress that our joint decoding algorithm does not require the knowledge of the geometry of the camera setup. Indeed, as demonstrated in Figs. 3–14, the proposed joint decoding algorithm can handle not only collinear multi-camera arrangement in which the captured videos from different viewpoints merely have horizontal parallax, such as *Balloons* and *Bookarrival*, but also camera arrangements in which videos captured from different viewpoints have severe horizontal and vertical parallax, such as *Ballet* and *Breakdancer*. In particular, the correlations between neighbor views are automatically exploited by the OF algorithm that estimates pixel-level disparities. Nevertheless, occlusions or corruptions may occur due to acquisition failure of a few individual cameras in the setup. Since each frame (view) is predicted by multiple reference images, including adjacent frame(s) and neighbor view(s), our joint decoding algorithm offers certain degree of robustness to possible occlusions/corruptions. In Figs. 15 and 16, we carry out a set of experiments on the *Ballet* and *Bookarrival* data sets, as representative of fast-motion and slow-motion video sequences, respectively, to illustrate these findings. For each group of $768 \times 1024 \times (p = 6 \text{ frames}) \times (q = 6 \text{ views})$ that is jointly decoded, the 1st, 3rd, 5th, 2nd, 4th, and 6th frame of the 1st, 2nd, 3rd, 4th, 5th, and 6th view, respectively, has occluded/corrupted patches (6 out of 36 frames). An occlusion/corruption example pattern is shown in Fig. 15. We divide the frame into non-overlapping blocks of size 32×32 and replace randomly half of the blocks by dark patches. Next, we perform exactly the same

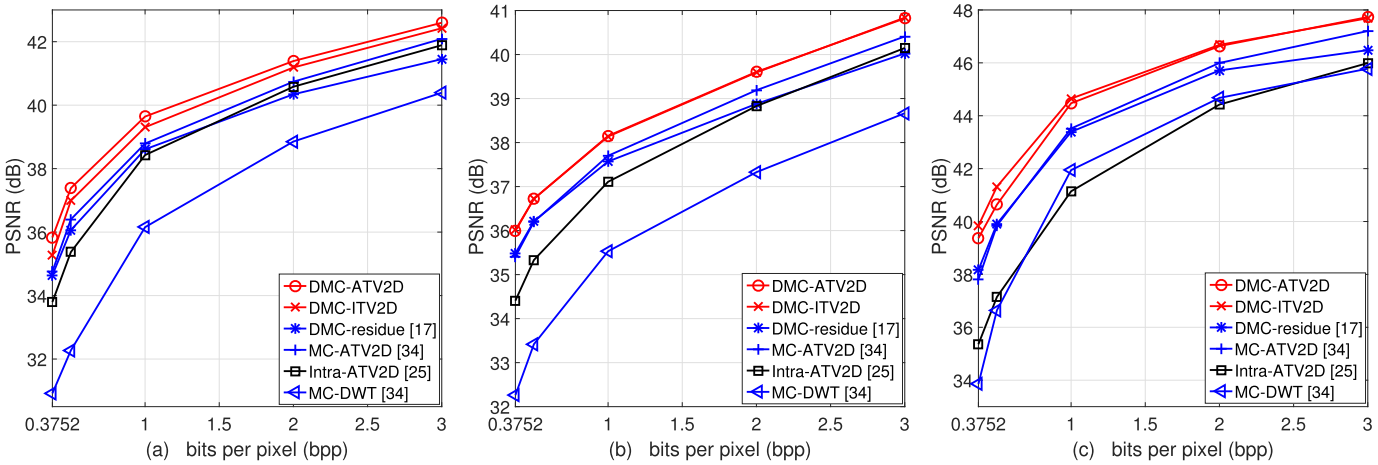


Fig. 10. Rate–distortion comparisons for the (a) *Ballet*, (b) *Breakdancer*, and (c) *Kendo* data sets.

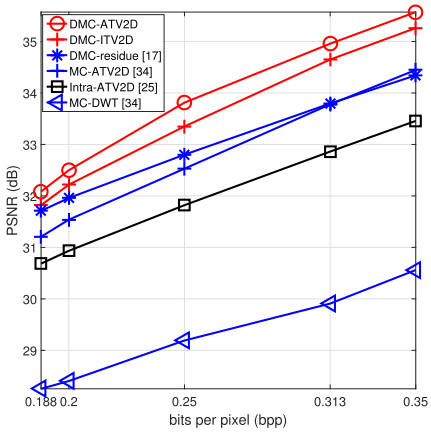


Fig. 11. Rate–distortion performance on *Ballet* in a low bit rate range.

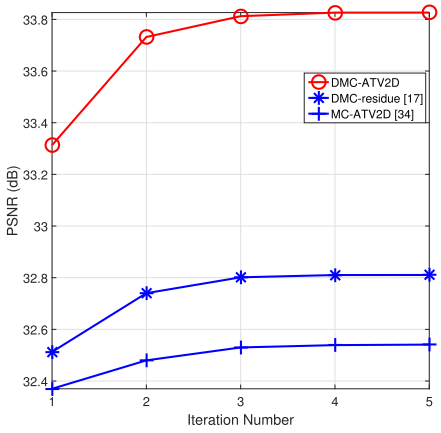


Fig. 12. PSNR versus number of iterations for the *Ballet* data set (0.25 bpp).

independent CS acquisition and proposed DMC-ATV2D joint reconstruction on the group of 36 partially occluded/corrupted frames and compute the average reconstruction PSNR of the 30 clean frames. From the rate–distortion performances in Fig. 16, we observe that using our proposed joint reconstruction scheme, the reconstruction quality of the clean frames only slightly deteriorates compared with the scenario where there are no occluded/corrupted frames. This demonstrates that our proposed decoder tolerates well certain degree of occlusions/corruptions (50% occlusion/corruption of 6 out of 36 frames).

C. Parameter Setting

In our implementation, the parameters μ_i and β_i, $i = 1, 2, 3$, in the augmented Lagrangian function (23.a)–(23.d) are set empirically. Empirically selecting regularization parameters is a common approach for CS image or video signal reconstruction [26], [27], [30], [40], [41], [44]. In the early work of ATV minimization for compressed-sensed signal recovery via ALM and ADMM [19], the sparse constraint is imposed on the spatial gradient only. Therein, the augmented Lagrangian function includes only two terms [see (23.a) and (23.d)], in which a default pair of parameters $\mu_1 = 1$ and $\beta_1 = 0.0625$ is used. Studies on more advanced ATV minimization for CS video recovery [31], [32] adopt the same default parameter values and generate superior reconstruction quality. Following the same parameter setting, in this work, we adopt parameter values $\mu_1 = 1$ and $\beta_1 = 0.0625$ for

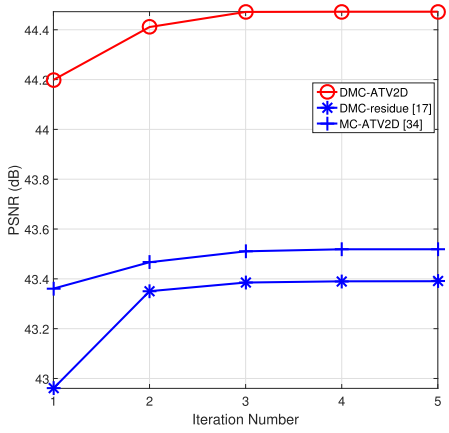


Fig. 13. PSNR versus number of iterations for the *Kendo* data set (1 bpp).

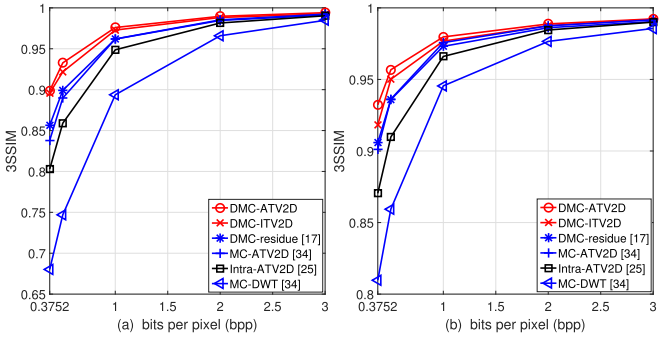


Fig. 14. Structural similarity image index comparison for the (a) *Bookarrival* and (b) *Ballet* data sets.

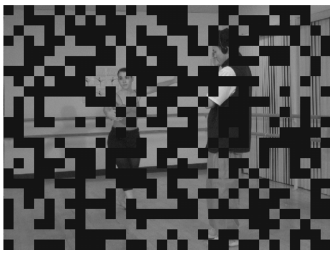


Fig. 15. Illustration of an occluded/corrupted frame of *Ballet*.

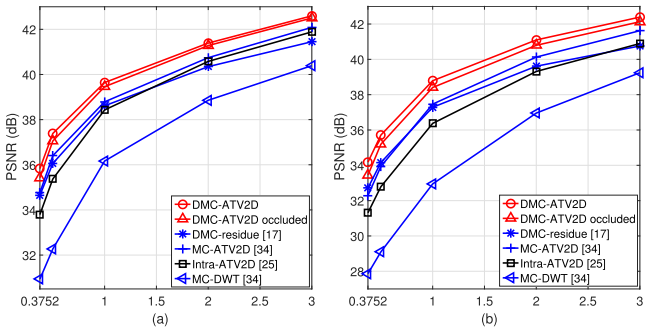


Fig. 16. Rate–distortion comparison for the partially occluded (a) *Ballet* and (b) *Bookarrival* data sets.

intra-frame 2D-ATV regularization. Then, for DC prediction residue ATV and MC prediction residue ATV, we set β_i at 0.0625, $i = 2, 3$, and tested four different sets of μ_i values, $i = 2, 3$, to determine the proper range for these parameters for general multiview video sequences. The testing data sets are *Ballet* and *Bookarrival* as representative of fast and slow motion sequences, respectively. From Tables II and III, it can be observed that the best rate–distortion performance is achieved at $\mu_i = 0.5$, $i = 2, 3$, for both sequences. Hence, we conclude that $\mu_1 = 1$, $\mu_i = 0.5$, $i = 2, 3$, and $\beta_i = 0.0625$, $i = 1, 2, 3$, are a proper set of parameters to be adopted in the proposed DMC-ATV2D joint decoder for general multiview video sequences. This set of values is used for all five data sets in this present work. Future studies may focus on determining the regularization parameters automatically for the proposed joint decoding algorithm.

TABLE II

PSNR VALUES FOR *Bookarrival*

	$r = 0.0469$	$r = 0.0625$	$r = 0.125$	$r = 0.25$	$r = 0.375$
$\mu_2 = \mu_3 = 0.25$	33.34 dB	34.68 dB	38.19 dB	40.61 dB	42.05 dB
$\mu_2 = \mu_3 = 0.5$	34.18 dB	35.71 dB	38.79 dB	41.10 dB	42.39 dB
$\mu_2 = \mu_3 = 1$	33.22 dB	34.86 dB	38.83 dB	41.29 dB	42.50 dB
$\mu_2 = \mu_3 = 2$	20.69 dB	28.13 dB	37.11 dB	40.92 dB	42.52 dB

TABLE III

PSNR VALUES FOR *Ballet*

	$r = 0.0469$	$r = 0.0625$	$r = 0.125$	$r = 0.25$	$r = 0.375$
$\mu_2 = \mu_3 = 0.25$	35.50 dB	37.08 dB	39.42 dB	41.25 dB	42.57 dB
$\mu_2 = \mu_3 = 0.5$	35.83 dB	37.39 dB	39.64 dB	41.39 dB	42.60 dB
$\mu_2 = \mu_3 = 1$	34.97 dB	36.73 dB	39.54 dB	41.31 dB	42.43 dB
$\mu_2 = \mu_3 = 2$	30.60 dB	34.67 dB	38.81 dB	41.04 dB	42.20 dB

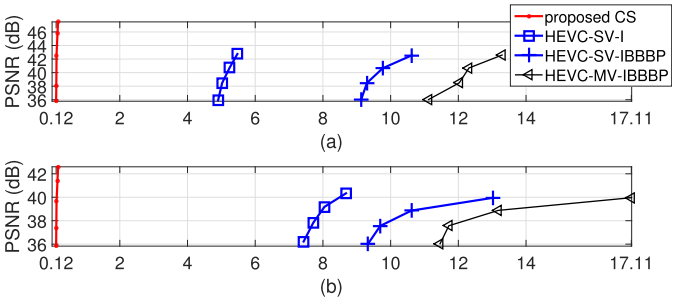


Fig. 17. Encoding time (sec/image) versus reconstruction PSNR of the proposed DMC-ATV2D method and HEVC codec for (a) *Balloons* and (b) *Ballet* data sets.

D. Encoding Time

To illustrate the extremely low encoder cost of our proposed distributed CS encoder, we compare the average encoding time per image versus the reconstruction PSNR of our proposed CS encoder with DMC-ATV2D decoding against the HEVC codec [69], [70] for the *Balloons* and *Ballet* data sets (Fig. 17). For HEVC, three group-of-pictures structures are considered: single-view intra-frame encoder (HEVC-SV-I) that encodes each view independently in intra-frame mode, single-view inter-frame encoder with IBBBP structure (HEVC-SV-IBBBP) that encodes each view independently in inter-frame mode, and multiview inter-frame encoder with IBBBP structure (HEVC-MV-IBBBP) that jointly encodes different views in inter-frame mode. With an Intel i5-2410M 2.30-GHz processor and MATLAB programming, the average encoding time of our distributed CS encoder is 0.1195, 0.1245, 0.1320, 0.1592, and 0.1775 seconds per image for the five adopted CS ratios, respectively. In contrast, the HEVC codec in C++ implementation can take more than 13 seconds on average to encode one image of *Balloons* and can take more than 17 seconds on average to encode one image of *Ballet*.

E. Decoding Complexity

In this section, we analyze the decoding complexity of the proposed DMC-ATV2D reconstruction algorithm and compare with existing compressed-sensed multiview video reconstruction algorithms. For a 4D multiview video data set of size $q \times p \times m \times n$, our findings are summarized in Table IV.

TABLE IV
DECODING COMPLEXITY OF THE $q \times p \times m \times n$ 4D VIDEO DATA SET

Algorithm	Operations	Additions per Operation	Multiplications per Operation	Total Additions	Total Multiplications	Execution Time (sec/image)
Proposed	$\widetilde{\mathbf{\Phi}}(\cdot)$, $\widetilde{\mathbf{\Phi}}^T(\cdot)$	$O(qpPmn)$		$N_{\text{out}}N_{\text{in}}O(qpPmn)$	$N_{\text{out}}[N_{\text{in}}O(2qpmn)$ $+O(2qp(mn)^2)]$	536.99 $N_{\text{out}}=3$
	$\widetilde{\mathbf{G}}(\cdot)$, $\widetilde{\mathbf{G}}^T(\cdot)$	$O(2qpmn)$				
	$\widetilde{\mathbf{D}}(\cdot)$, $\widetilde{\mathbf{D}}^T(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$\widetilde{\mathbf{M}}(\cdot)$, $\widetilde{\mathbf{M}}^T(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k-1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k-1})$ $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k+1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k+1})$		$O(2(q-1)p(2mn)^2)$			
	$(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$		$O(2q(p-1)(2mn)^2)$			
DMC-residue	$\widetilde{\mathbf{\Phi}}(\cdot)$, $\widetilde{\mathbf{\Phi}}^T(\cdot)$	$O(qpPmn)$		$N_{\text{out}}N_{\text{in}}O(qpPmn)$	$N_{\text{out}}O(2qp(mn)^2)$	174.68 $N_{\text{out}}=3$
	$\widetilde{\mathbf{G}}(\cdot)$, $\widetilde{\mathbf{G}}^T(\cdot)$	$O(2qpmn)$				
	$\widetilde{\mathbf{D}}(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$\widetilde{\mathbf{M}}(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k-1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k-1})$ $(\mathbf{\Delta}_{\mathbf{h}_{t,k}}^{k+1}, \mathbf{\Delta}_{\mathbf{v}_{t,k}}^{k+1})$		$O(2(q-1)p(2mn)^2)$			
	$(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$		$O(2q(p-1)(2mn)^2)$			
MC-ATV2D	$\widetilde{\mathbf{\Phi}}(\cdot)$, $\widetilde{\mathbf{\Phi}}^T(\cdot)$	$O(qpPmn)$		$N_{\text{out}}N_{\text{in}}O(qpPmn)$	$N_{\text{out}}[N_{\text{in}}O(qpmn)$ $+O(qp(mn)^2)]$	287.15 $N_{\text{out}}=3$
	$\widetilde{\mathbf{G}}(\cdot)$, $\widetilde{\mathbf{G}}^T(\cdot)$	$O(2qpmn)$				
	$\widetilde{\mathbf{M}}(\cdot)$, $\widetilde{\mathbf{M}}^T(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$		$O(2q(p-1)(2mn)^2)$			
MC-DWT	$\widetilde{\mathbf{\Phi}}(\cdot)$, $\widetilde{\mathbf{\Phi}}^T(\cdot)$	$O(qpPmn)$		$N_{\text{out}}N_{\text{in}}O(qpPmn)$	$N_{\text{out}}[N_{\text{in}}O(qp\max(mn^2,$ $m^2n))+O(qp(mn)^2)]$	447.95 $N_{\text{out}}=3$
	$\widetilde{\mathbf{\Psi}}(\cdot)$, $\widetilde{\mathbf{\Psi}}^T(\cdot)$		$O(qp\max(mn^2,m^2n))$			
	$\widetilde{\mathbf{M}}(\cdot)$, $\widetilde{\mathbf{M}}^T(\cdot)$	$O(3qpmn)$	$O(4^3qpmn)$			
	$(\sigma_{\mathbf{h}_t}^{k,t-1}, \sigma_{\mathbf{v}_t}^{k,t-1})$ $(\sigma_{\mathbf{h}_t}^{k,t+1}, \sigma_{\mathbf{v}_t}^{k,t+1})$		$O(2q(p-1)(2mn)^2)$			
Intra-ATV2D	$\widetilde{\mathbf{\Phi}}(\cdot)$, $\widetilde{\mathbf{\Phi}}^T(\cdot)$	$O(qpPmn)$		$N_{\text{in}}O(qpPmn)$		77.18
	$\widetilde{\mathbf{G}}(\cdot)$, $\widetilde{\mathbf{G}}^T(\cdot)$	$O(2qpmn)$				
Intra-GSR	SVD		$O(qp\text{IterNum}N_GB_sc^2)$		$O(qp\text{IterNum}N_GB_sc^2)$	6552.99 $\text{IterNum}=100$

The operators needed for decoding are listed for each algorithm where transpose (T) indicates the adjoint operation. Since we adopt random partial WH sensing (± 1 entries), the CS operation $\widetilde{\mathbf{\Phi}}$ requires only additions. The gradient operation $\widetilde{\mathbf{G}}(\cdot)$ calculates the intensity difference of two adjacent pixels, and therefore, it also requires additions only. $\widetilde{\mathbf{D}}(\cdot)$ and $\widetilde{\mathbf{M}}(\cdot)$ are the DC and MC predictors, respectively, which are each performed via pixel-by-pixel bilinear interpolation from four surrounding pixels. This computation is dominated by the solution of a 4×4 linear equation with complexity 4^3 multiplications and 3 additions at most. The disparity fields, $(\mathbf{\Delta}_{\boldsymbol{h}_{t,k}}^{k-1}, \mathbf{\Delta}_{\boldsymbol{v}_{t,k}}^{k-1})$, $(\mathbf{\Delta}_{\boldsymbol{h}_{t,k}}^{k+1}, \mathbf{\Delta}_{\boldsymbol{v}_{t,k}}^{k+1})$, and motion fields, $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t-1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t-1})$, $(\boldsymbol{\sigma}_{\boldsymbol{h}_t}^{k,t+1}, \boldsymbol{\sigma}_{\boldsymbol{v}_t}^{k,t+1})$, are estimated by the OF algorithm, which solves a set of linear equations of size $2mn \times 2mn$ by the conjugate gradient method with complexity $O((2mn)^2)$. For MC-DWT, $\widetilde{\mathbf{\Psi}}(\cdot)$ denotes the 2D wavelet transform that requires $O(\max(m^2n, mn^2))$ multiplications for an $m \times n$ frame.

The proposed DMC-ATV2D and DMC-residue, MC-ATV2D, and MC-DWT algorithms all iterate between: 1) ME (or motion- and disparity-estimation) and 2) MC (or DMC) ATV (or DWT) minimization. We call such iterations "outer" iterations. Step 2) is the core optimization program (DMC ATV minimization in (21) for the proposed DMC-ATV2D scheme, MC ATV minimization for MC-ATV2D, MC DWT minimization for MC-DWT, residue ATV minimization for DMC-residue, and ATV minimization for Intra-ATV2D) that is solved by minimizing an augmented Lagrangian function, which iterates between updating the variables and the Lagrangian multipliers. We call such iterations "inner" iterations. For Intra-ATV2D, only "inner" iteration is required since DE (or ME) is not needed. If N_{out} is the number of outer iterations and N_{in} is the number of inner iterations, the total complexity in terms of additions and multiplications, respectively, is summarized in Table IV for all algorithms. Our studies conclude that all algorithms have the same complexity in terms of *additions*, which is dominated by the CS operation $\widetilde{\mathbf{\Phi}}$, except for Intra-ATV2D that only requires inner iterations. In addition, the proposed DMC-ATV2D that simultaneously considers motion and disparity has approximately twice the number of *multiplications* compared with MC-ATV2D, which only considers motion. MC-DWT also considers motion only, and hence the number of multiplications it requires is less than the proposed DMC-ATV2D. However, compared with MC-ATV2D, MC-DWT performs 2D-DWT ($\widetilde{\mathbf{\Psi}}$ and its adjoint $\widetilde{\mathbf{\Psi}}^T$) that requires multiplications, while MC-ATV2D performs a gradient operation ($\widetilde{\mathbf{G}}$ and its adjoint $\widetilde{\mathbf{G}}^T$) that requires additions only. As a result, MC-DWT has higher complexity than MC-ATV2D. Furthermore, multiplication operations

required by DMC-residue are less than DMC-ATV2D since MC prediction $\widetilde{\mathbf{M}}(\cdot)$ and DC prediction $\widetilde{\mathbf{D}}(\cdot)$ are performed in the outer iteration, while the proposed DMC-ATV2D scheme executes these operations in the inner iteration.

The Intra-GSR decoder [27] reconstructs each image independently. It partitions an $m \times n$ image into N_G overlapping patches of size $\sqrt{B_s} \times \sqrt{B_s}$. For each patch, it searches for c most correlated patches surrounding it and forms a group/matrix of vectorized patches of size $B_s \times c$. For each group of patches, a dictionary is learned via singular value decomposition (SVD) with complexity $O(B_s c^2)$ followed by ℓ_0-norm minimization that recovers the group of patches from the CS measurements with the learned dictionary. If $\mathrm{IterNum}$ is the number of iterations between SVD dictionary learning and group-of-patches recovery, the total complexity of Intra-GSR in terms of multiplications is $O(qp\mathrm{IterNum}N_G B_s c^2)$ for all q views and p frames in the 4D video data. The experiments show that our proposed DMC-ATV2D scheme only needs $N_{\mathrm{out}} = 3$ outer iterations until convergence, while the Intra-GSR decoder usually needs $\mathrm{IterNum} > 100$. The last column of Table IV shows the average decoding time per image in our experiments. $N_{\mathrm{out}} = 3$ is adopted for the proposed DMC-ATV2D and the existing iterative algorithms DMC-residue, MC-ATV2D, and MC-DWT, while $\mathrm{IterNum} = 100$ is adopted for Intra-GSR. Our proposed DMC-ATV2D decoder is significantly faster than the Intra-GSR decoder.

VII. Conclusion

We proposed a novel joint reconstruction algorithm for multiview video reconstruction from independently compressed-sensed frames and views. The decoder simultaneously utilizes intra-frame, inter-view, and temporal correlations to enhance signal sparsity. The decoder consists of an initial independent frame reconstruction stage (which performs simple 2D ATV minimization frame recovery) and a follow-up joint decoding stage. The joint decoding stage iterates between DE among the multiple views, ME along the temporal direction, and a DMC ATV minimization step. Since both inter-view and temporal sparsity is considered, as well as intra-frame sparsity, the reconstruction quality is significantly improved compared with existing CS multiview video reconstruction algorithms, such as independent frame 2D ATV minimization, MC 2D-ATV (or 2D-DWT ℓ_1-norm) minimization, DMC residue recovery, and intra-frame group sparse representation approach. At the same time, the proposed DMC-ATV2D decoder is robust to a certain degree of occlusion/corruption in the multiview video data as demonstrated by experimental studies.

References

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 10–21, Nov. 2007.

[2] A. Smolic *et al.*, "3D video and free viewpoint video—Technologies, applications and MPEG standards," in *Proc. IEEE Int. Conf. Multimedia Expo*, Toronto, ON, Canada, Jul. 2006, pp. 2161–2164.

[3] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 67–76, Jan. 2011.

[4] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66–76, Nov. 2007.

[5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[6] D. Ren, S.-H. G. Chan, G. Cheung, and P. Frossard, "Coding structure and replication optimization for interactive multiview video streaming," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1874–1887, Nov. 2014.

[7] 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.

[8] C. Li, H. Jiang, P. Wilford, Y. Zhang, and M. Scheutzow, "A new compressive video sensing framework for mobile broadcast," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 197–205, Mar. 2013.

[9] J. Xiao, M. M. Hannuksela, T. Tillo, M. Gabbouj, C. Zhu, and Y. Zhao, "Scalable bit allocation between texture and depth views for 3-D video streaming over heterogeneous networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 139–152, Jan. 2015.

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

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[12] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

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

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

[15] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

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

[17] M. Trocan, E. W. Tramel, J. E. Fowler, and B. Pesquet, "Compressed-sensing recovery of multiview image and video sequences using signal prediction," *Multimedia Tools Appl.*, vol. 72, no. 1, pp. 95–121, Sep. 2014.

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

[19] C. Li, W. Yin, and Y. Zhang. (2009). *TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithm*. [Online]. Available: http://www.caam.rice.edu/∼optimization/L1/TVAL3/

[20] C. Li, W. Win, H. Jing, and Y. Zhang, "An efficient augmented Lagrangian method with applications to total variation minimization," *Comput. Optim. Appl.*, vol. 56, no. 3, pp. 507–530, Dec. 2013.

[21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–124, Jan. 2011.

[22] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008, pp. 1–5.

[23] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[24] J. Yang, Y. Zhang, and W. Yin, "An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise," *SIAM J. Sci. Comput.*, vol. 31, no. 4, pp. 2842–2865, 2009.

[25] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM J. Imag. Sci.*, vol. 6, no. 2, pp. 1035–1058, Jun. 2013.

[26] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Image compressive sensing recovery via collaborative sparsity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 380–391, Sep. 2012.

[27] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.

[28] M. B. Wakin *et al.*, "Compressive imaging for video representation and coding," in *Proc. Picture Coding Symp. (PCS)*, Beijing, China, Apr. 2006, pp. 1–6.

[29] R. F. Marcia and R. M. Willet, "Compressive coded aperture video reconstruction," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008, pp. 1–5.

[30] C. Li, H. Jiang, P. Wilford, and Y. Zhang, "Video coding using compressive sensing for wireless communications," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2011, pp. 2077–2082.

[31] Y. Liu and D. A. Pados, "Rate-adaptive compressive video acquisition with sliding-window total-variation-minimization reconstruction," *Proc. SPIE*, vol. 8717, p. 871703, May 2013.

[32] Y. Liu and D. A. Pados, "Decoding of framewise compressed-sensed video via interframe total variation minimization," *J. Electron. Imag.*, vol. 22, no. 2, p. 021012, Apr. 2013.

[33] 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.

[34] M. S. Asif, F. Fernandes, and J. Romberg, "Low-complexity video compression and compressive sensing," in *Proc. ACSSC*, Nov. 2013, pp. 579–583.

[35] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization for accelerated dynamic MRI," *Magn. Reson. Med.*, vol. 70, no. 3, pp. 800–821, Sep. 2013.

[36] M. B. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," in *Proc. Picture Coding Symp. (PCS)*, Chicago, IL, USA, May 2009, pp. 1–4.

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

[38] 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 (ICME)*, Suntec City, Singapore, Jul. 2010, pp. 1225–1229.

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

[40] V. Thirumalai and P. Frossard, "Joint reconstruction of multiview compressed images," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1969–1981, May 2013.

[41] X. Wang and J. Liang, "View interpolation confidence-aided compressed sensing of multiview images," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 1651–1655.

[42] C. Fu, X. Ji, and Q. Dai, "Compressed multi-view imaging with joint reconstruction," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2011, p. 448.

[43] Q. Dai, C. Fu, X. Ji, and Y. Zhang, "Robust joint reconstruction in compressed multi-view imaging," in *Proc. Picture Coding Symp. (PCS)*, May 2012, pp. 13–16.

[44] K. Chang, T. Qin, W. Xu, and A. Men, "A joint reconstruction algorithm for multi-view compressed imaging," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 221–224.

[45] 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. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 1458–1462.

[46] H. Ganapathy, D. A. Pados, and G. N. Karystinos, "New bounds and optimal binary signature sets—Part I: Periodic total squared correlation," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1123–1132, Apr. 2011.

[47] H. Ganapathy, D. A. Pados, and G. N. Karystinos, "New bounds and optimal binary signature sets—Part II: Aperiodic total squared correlation," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1411–1420, May 2011.

[48] E. Candès and J. Romberg. *ℓ_1-Magic: Recovery of Sparse Signals Via Convex Programming*. [Online]. Available: http://www.acm.caltech.edu/l1magic/downloads/l1magic.pdf.

[49] M. R. Dadkhah, S. Shirani, and M. J. Deen, "Compressive sensing with modified total variation minimization algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 1310–1313.

[50] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, May 2004, pp. 25–36.

[51] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Doctoral thesis, Massachusetts Inst. Technol., Boston, MA, USA, May 2009.

[52] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

[53] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1385–1392.

[54] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "DeepMatching: Hierarchical deformable dense matching," *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 300–323, Dec. 2016.

[55] C. Li, "An efficient algorithm for total variation regularization with applications to the sigle pixel camera and compressive sensing," M.S. thesis, Dept. Comput. Appl. Math., Rice Univ., Houston, TX, USA, 2009.

[56] H. Jiang, W. Deng, and Z. Shen, "Surveillance video processing using compressive sensing," *Inverse Problem Imag.*, vol. 6, no. 2, pp. 201–214, 2012.

[57] H. Jiang, S. Zhao, Z. Shen, W. Deng, P. A. Wilford, and R. Haimi-Cohen, "Surveillance video analysis using compressive sensing with low latency," *Bell Labs Tech. J.*, vol. 18, no. 4, pp. 63–74, Mar. 2014.

[58] G. Van Wallendael, S. Van Leuven, J. De Cock, F. Bruls, and R. Van de Walle, "3D video compression based on High Efficiency Video Coding," *IEEE Trans. Consum. Electron.*, vol. 58, no. 1, pp. 137–145, Feb. 2012.

[59] F. Zou, D. Tian, A. Vetro, H. Sun, O. C. Au, and S. Shimizu, "View synthesis prediction in the 3-D video coding extensions of AVC and HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1696–1708, Oct. 2014.

[60] Y. Chen, L. Zhang, V. Seregin, and Y.-K. Wang, "Motion hooks for the multiview extension of HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2090–2098, Dec. 2014.

[61] H. Yuan, S. Kwong, X. Wang, W. Gao, and Y. Zhang, "Rate distortion optimized inter-view frame level bit allocation method for MV-HEVC," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2134–2146, Dec. 2015.

[62] A. I. Purica, E. G. Mora, B. Pesquet-Popescu, M. Cagnazzo, and B. Ionescu, "Multiview plus depth video coding with temporal prediction view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 2, pp. 360–374, Feb. 2016.

[63] Y. Chen, X. Zhao, L. Zhang, and J. W. Kang, "Multiview and 3D video compression using neighboring block based disparity vectors," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 576–589, Apr. 2016.

[64] Y. Zhang, S. Kwong, L. Xu, S. Hu, G. Jiang, and C.-C. J. Kuo, "Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3497–3512, Sep. 2013.

[65] J. Chakareski, V. Velisavljević, and V. Stanković, "User-action-driven view and rate scalable multiview video coding," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3473–3484, Sep. 2013.

[66] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.

[67] L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.

[68] I. Feldmann *et al.*, *HHI Test Material for 3-D Video*, document ISO/IEC JTC1/SC29/WG11, MPEG08/M15413, Archamps, France, May 2008.

[69] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[70] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of High Efficiency Video Coding (HEVC)," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.

[71] C. Li and A. C. Bovik, "Content-weighted video quality assessment using a three-component image model," *J. Electron. Imag.*, vol. 19, no. 1, p. 011003, Jan./Mar. 2010.

[72] 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.

Ying Liu (S'11–M'13) received the B.S. degree in telecommunications engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2006 and the Ph.D. degree in electrical engineering from The State University of New York at Buffalo, Buffalo, NY, USA, in 2012.

She is currently a Research Scientist with The State University of New York at Buffalo. Her research lies in the intersection of signal processing, machine learning, and optimization, with applications to multimedia big data.

Dimitris A. Pados (M'95–SM'15) received the Diploma degree in computer science and engineering (five-year program) from the University of Patras, Greece, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA.

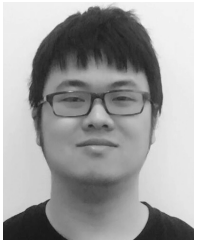
Since August 1997, he has been with the Department of Electrical Engineering, The State University of New York at Buffalo, holding in sequence the titles of Assistant Professor, Associate Professor, Professor, and currently Clifford C. Furnas Chair Professor of Electrical Engineering. He served the Department as Associate Chair in 2009-2010. Dr. Pados was elected four times University Faculty Senator (terms 2004-06, 2008-10, 2010-12, 2016-18) and served on the Faculty Senate Executive Committee (2009-10, 2016-17).

His research interests are in the general areas of communication theory and adaptive signal processing with applications to signal waveform design; multiplexing, multiple-access and interference channels; secure wireless communications; cognitive radios and networks; L1-norm data and signal processing; compressed sensing; image and video processing; error correcting coding; data embedding and hiding; antenna arrays and array radar.

Dr. Pados is a member of the IEEE Communications, Signal Processing, Information Theory, and Computational Intelligence Societies. He served as an Associate Editor for the IEEE Signal Processing Letters and the IEEE Transactions on Neural Networks. He received a 2001 IEEE International Conference on Telecommunications best paper award, the 2003 IEEE Transactions on Neural Networks Outstanding Paper Award, the 2010 IEEE International Communications Conf. (ICC) Best Paper Award in Signal Processing for Communications, the 2013 International Symposium on Wireless Communication Systems Best Paper Award in Physical Layer Communications and Signal Processing, a Best of IEEE GLOBECOM 2014 - Top 50 Papers distinction, and Best Paper Selection in 2016 IEEE Intern. Conf. on Multimedia Big Data (IEEE BigMM) for articles that he coauthored with students and colleagues. Professor Pados is a recipient of the 2009 SUNY-system-wide Chancellor's Award for Excellence in Teaching and the 2011 University at Buffalo Exceptional Scholar - Sustained Achievement Award.

Joohee Kim (S'97–M'03) received the B.S. and M.S. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea in 1991 and 1993, respectively. She received the Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, in 2003.

From 1993 to 1997, she was with Korea Telecom Research Laboratory, Seoul, Korea as a Research Engineer. She joined Samsung Advanced Institute of Technology, Suwon-si, Korea in 2003 as a Senior Research Engineer and developed various video coding algorithms. From 2005 to 2008, she was an Assistant Professor in the Department of Information and Communication Engineering at Inha University in South Korea. She joined the faculty of the Illinois Institute of Technology (IIT), Chicago, IL, in 2009 and is currently an Associate Professor of the Department of Electrical and Computer Engineering. She is the Director of Multimedia Communications Laboratory at IIT and has been actively involved in research projects funded by US Federal Agencies and Korean Government. Her current research interests include image and video signal processing, multimedia communication, multimedia systems, 3D video representation and transmission, real-time 3D reconstruction for augmented teleoperation, computer vision and machine learning.



Chen Zhang received the B.E. degree in electronic information engineering from North China Electric Power University, Beijing, China, in 2012 and the M.S. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2014, where he is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering.

His research interests include stereo image processing, computer vision, and object detection.