JOINT-VIEW KALMAN-FILTER RECOVERY OF COMPRESSED-SENSED MULTIVIEW VIDEOS[†]

Ying Liu, Shubham Chamadia, and Dimitris A. Pados[*]

Department of Electrical Engineering, State University of New York at Buffalo
Buffalo, NY, 14260 USA

ABSTRACT

We develop a novel joint-view Kalman filter for causal reconstruction of compressed-sensed multiview videos. Compressed-sensed multiview video frames are initially reconstructed individually via ℓ_1-norm minimization. Then, a joint-view state transition model is established for each pair of neighboring views using motion or motion-disparity field estimates. Experimental results demonstrate significantly improved reconstruction quality compared to conventional CS reconstruction and independent-view (single-view) motion-compensated Kalman filtering.

Index Terms— Compressed sensing, joint reconstruction, Kalman filter, motion/motion-disparity compensation, multiview video.

1. INTRODUCTION

Compressive sampling (CS), also referred to as compressed sensing, deals with sub-Nyquist sampling of sparse signals of interest [1]-[3]. Rather than collecting an entire Nyquist ensemble of signal samples, sparse signals can be reconstructed from a small number of (random [3] or deterministic [4]) linear measurements. As a fast data acquisition modality, CS has been successfully applied to time-sequence problems such as dynamic magnetic resonance imaging (dMRI) [5], hyperspectral imaging [6], distributed video coding [7], [8], and multiview video coding [9], [10].

To reconstruct compressed-sensed video sequences, conventional methods establish an appropriate sparse data domain representation and solve a convex optimization problem that minimizes the ℓ_1-norm of the sparse domain coefficients. An intuitive (JPEG-motivated) approach is to independently recover each frame using the 2-D discrete cosine transform (2-D DCT) [11] or a 2-D discrete wavelet transform (2-D DWT). To enhance sparsity by exploiting correlations among successive frames, several frames can be jointly recovered with a 3-D DWT [12] basis. Joint multi-frame recovery of this form, however, can be considered only as an off-line reconstruction method, since it requires access to the CS measurements of the entire group of video frames and has prohibitively high computational complexity. Instead, effective frame-by-frame reconstruction can be attempted with an adaptive sparse basis [13], [14], which still has significant complexity due to the adaptive calculation of the sparse basis and the iterative nature of the convex optimization method.

Alternatively, Kalman filtered compressed sensing (KF-CS) was suggested [15] to causally reconstruct time sequences of sparse signals with unknown and slow time-varying sparsity patterns from CS measurements. In [16], KF-CS was used to reconstruct real-time cardiac and brain image sequences from dMRI data with significantly improved reconstruction results compared to conventional CS methods. More recently, a generalized KF-CS method was considered [17] to incorporate motion estimation and enhance KF-CS for compressed-sensed monoview video reconstruction.

In this work, for the first time in the literature, we present a Kalman filtering algorithmic approach to reconstruct compressed-sensed multiview video sequences. Our contributions include a new joint-view state transition model that utilizes motion or motion-disparity compensation and a joint-view Kalman filter that updates each pair of adjacent views simultaneously. The multiple views are initially recovered independently via minimizing the ℓ_1-norm of the 2-D DCT coefficients. Then, a joint-view state transition model is established by estimating the motion or motion-disparity field between current and previous time instants. Afterwards, a sliding-window joint-view motion-adaptive or motion-disparity-adaptive Kalman filter is run for each pair of adjacent views and data fusion is utilized to obtain the final reconstruction. Experimental studies presented in this paper illustrate the theoretical development and demonstrate that the proposed joint-view motion-adaptive or motion-disparity-adaptive Kalman procedure outperforms independent-view motion-compensated Kalman and conventional CS methods.

2. KALMAN FILTERING NOTATION AND BACKGROUND

The KF model assumes that the state of a system at time t evolves from the prior state at time $t-1$ according to the

[†]This work was supporte d in part by the National Science Foundation under Grants CNS-1422874 and ECCS-1462341.

[*]Corresponding author.

equation

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t \tag{1}$$

where $\mathbf{x}_t$ is the state vector at time t, $\mathbf{F}_t$ is the state transition matrix which applies the effect of $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$, $\mathbf{u}_t$ is the control input, $\mathbf{B}_t$ is the control input matrix which applies the effect of $\mathbf{u}_t$ to $\mathbf{x}_t$, and $\mathbf{w}_t$ is the process noise. The process noise is assumed to be drawn from a zero mean multivariate Gaussian distribution with covariance matrix $\mathbf{Q}_t$. In this work, we assume that no control input is involved and the state transition model is simplified to

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{w}_t. \tag{2}$$

Measurements of the system can be performed according to the model

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \tag{3}$$

where $\mathbf{y}_t$ is the measurement vector, $\mathbf{H}_t$ is the transformation matrix that maps $\mathbf{x}_t$ into the measurement domain, and $\mathbf{v}_t$ is measurement noise. As with the process noise, measurement noise is assumed to be zero-mean white Gaussian with covariance matrix $\mathbf{R}_t$.

The KF algorithm that causally recovers a time sequence $\{\mathbf{x}_t\}_{t=1}^{p}$ from the measurement sequence $\{\mathbf{y}_t\}_{t=1}^{p}$ involves iterations between two stages, prediction and measurement update. With initialized variables $\widehat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$ and known $\mathbf{F}_t$, $\mathbf{H}_t$, $\mathbf{Q}_t$, and $\mathbf{R}_t$ values, the standard KF equations for the prediction stage are

$$\widehat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1|t-1} \tag{4}$$

and

$$\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^{\mathrm{T}} + \mathbf{Q}_t \tag{5}$$

where $\mathbf{P}_{t|t-1} \triangleq \mathbf{E}\{(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1})^{\mathrm{T}}\}$ is the prediction error autocorrelation matrix ((5) can be derived as in [18], for example). The measurement update equations are given by

$$\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\widehat{\mathbf{x}}_{t|t-1}) \tag{6}$$

and

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_{t|t-1} \tag{7}$$

where $\mathbf{P}_{t|t} \triangleq \mathbf{E}\{(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})^{\mathrm{T}}\}$ is the updated error autocorrelation matrix and the Kalman gain $\mathbf{K}_t$ is given by $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}$, which can be obtained by minimizing the mean squared error $\mathbf{E}\{\|\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t}\|_2^2\}$.

3. THE DEVELOPED JOINT-VIEW KALMAN FILTER

Consider a multi-camera system of q cameras that capture the same scene from q different positions. The kth camera generates a sequence of video frames $\mathbf{X}_t^k \in \mathbb{R}^{m\times n}$ vectorized as $\mathbf{x}_t^k \in \mathbb{R}^D$, $D = mn$, $t = 1, ..., p$. Each camera performs independent CS frame acquisition to obtain the CS measurement $\mathbf{z}_t^k = \mathbf{H}\mathbf{x}_t^k$ where $\mathbf{H} \in \mathbb{R}^{P\times D}$ is the sensing matrix with entries generated from a zero-mean, unit-variance Gaussian distribution. Once $\mathbf{H}$ is generated beforehand off-line, it is fixed to process all frames in all views. We assume independent identically distributed (i.i.d.) Gaussian noise in the measurement/transmission channel such that the decoder receives $\mathbf{y}_t^k = \mathbf{z}_t^k + \mathbf{v}_t^k = \mathbf{H}\mathbf{x}_t^k + \mathbf{v}_t^k$, for $k = 1, ..., q$ and $t = 1, ..., p$, where the measurement noise $\mathbf{v}_t^k$ has autocorrelation matrix $\mathbf{R}_t^k \triangleq \mathbf{E}\{\mathbf{v}_t^k\mathbf{v}_t^{k^{\mathrm{T}}}\} = \sigma^2\mathbf{I}_P$. In the following, the proposed joint-view Kalman filter is developed in detail.

3.1. Single-view state-transition model

The state transition of a monoview video sequence [17] can be used to develop the state transition model for the individual kth view,

$$\mathbf{x}_t^k = \mathbf{F}_t^k\mathbf{x}_{t-1}^k + \mathbf{w}_t^k, \qquad t = 1, ..., p, \tag{8}$$

where $\mathbf{F}_t^k$ performs linear motion-compensated prediction from $t-1$ to t for the kth view, $\mathbf{w}_t^k$ is the prediction error, and the autocorrelation matrix of the prediction error is assumed to be $\mathbf{Q}_t^k \triangleq \mathbf{E}\{\mathbf{w}_t^k\mathbf{w}_t^{k^{\mathrm{T}}}\} = \mathbf{I}_D$ for $k = 1, ..., q$ and $t = 1, ..., p$.

3.2. Joint-view state-transition model

In our joint-view state transition model, two adjacent views $\mathbf{x}_t^k$ and $\mathbf{x}_t^{k+1}$ captured at the same time t by two neighboring cameras are stacked to form a single state vector $\mathbf{x}_t = [\mathbf{x}_t^{k^T}\mathbf{x}_t^{k+1^T}]^T \in \mathbb{R}^{2D}$. For fast-motion video sequences, a block-diagonal state transition matrix $\mathbf{F}_t$ can be formed using $\mathbf{F}_{t,k}^k$ and $\mathbf{F}_{t,k+1}^{k+1}$ as the diagonal elements

$$\underbrace{\begin{bmatrix}\mathbf{x}_t^k\\ \mathbf{x}_t^{k+1}\end{bmatrix}}_{\mathbf{x}_t} = \underbrace{\begin{bmatrix}\mathbf{F}_{t,k}^k & \mathbf{0}\\ \mathbf{0} & \mathbf{F}_{t,k+1}^{k+1}\end{bmatrix}}_{\triangleq\,\mathbf{F}_t}\underbrace{\begin{bmatrix}\mathbf{x}_{t-1}^k\\ \mathbf{x}_{t-1}^{k+1}\end{bmatrix}}_{\mathbf{x}_{t-1}} + \underbrace{\begin{bmatrix}\mathbf{w}_t^k\\ \mathbf{w}_t^{k+1}\end{bmatrix}}_{\mathbf{w}_t} \tag{9}$$

where $\mathbf{F}_{t,i}^i$, $i = k, k+1$, is the motion compensation operator that predicts the ith view at time t from the ith view at time $t-1$. For slow-motion multiview video sequences, since the cross-view scene change from time $t-1$ to time t is small, both $\widehat{\mathbf{x}}_{t-1|t-1}^k$ and $\widehat{\mathbf{x}}_{t-1|t-1}^{k+1}$ can be used to predict $\mathbf{x}_{t|t}^j$, $j = k, k+1$. Therefore, the joint-view state transition model can be formulated as

$$\underbrace{\begin{bmatrix}\mathbf{x}_t^k\\ \mathbf{x}_t^{k+1}\end{bmatrix}}_{\mathbf{x}_t} = \underbrace{\begin{bmatrix}\frac{1}{2}\mathbf{F}_{t,k}^k & \frac{1}{2}\mathbf{F}_{t,k}^{k+1}\\ \frac{1}{2}\mathbf{F}_{t,k+1}^k & \frac{1}{2}\mathbf{F}_{t,k+1}^{k+1}\end{bmatrix}}_{\triangleq\,\mathbf{F}_t}\underbrace{\begin{bmatrix}\mathbf{x}_{t-1}^k\\ \mathbf{x}_{t-1}^{k+1}\end{bmatrix}}_{\mathbf{x}_{t-1}} + \underbrace{\begin{bmatrix}\mathbf{w}_t^k\\ \mathbf{w}_t^{k+1}\end{bmatrix}}_{\mathbf{w}_t} \tag{10}$$

where $\mathbf{F}_{t,i}^{j}$, $i, j = k, k+1$, $i \neq j$ is the cross-view motion-disparity-compensation operator that predicts the ith view at time t from the jth view at time $t-1$.

The compact expression for joint-view state transition is

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t \tag{11}$$

where $\mathbf{w}_t$ is the prediction error with auto-correlation matrix $\mathbf{Q}_t \triangleq \mathbf{E}\{\mathbf{w}_t \mathbf{w}_t^T\} = \mathbf{I}_{2D}$.

3.3. Joint-view Kalman filter update

Fig. 1 depicts the system architecture for the proposed work. We aim at developing a procedure for causal reconstruction of multiview video sequences that provides better reconstruction quality than ℓ_1-norm minimization based CS recovery. At time instant t, the decoder first performs CS recovery of each pair of neighboring views $\widehat{\mathbf{x}}_t^{j,\mathrm{CS}}$, $j = k, k+1$. Then, motion ($\mathbf{F}_t$ by (9)) or motion-disparity ($\mathbf{F}_t$ by (10)) estimation is performed using the currently recovered $\widehat{\mathbf{x}}_t^{j,\mathrm{CS}}$, $j = k, k+1$, and the stored previously recovered $\widehat{\mathbf{x}}_{t-1}^{j,\mathrm{CS}}$, $j = k, k+1$. The resulting motion or motion-disparity field $\mathbf{F}_{t,i}^{j}$, $i = k, k+1$, $j = k, k+1$, is then used in the joint-view motion-compensated prediction in (9) or the joint-view motion-disparity-compensated prediction in (10), respectively.

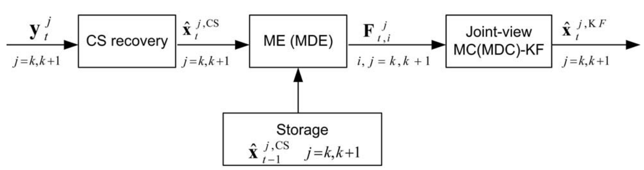

Fig. 1. Causal recovery of CS multiview video sequences with motion estimation (ME) or motion-disparity estimation (MDE).

The standard KF algorithm can be applied to multiview scenarios with the proposed joint-view state transition model and the resulting prediction equations are as in (4) and (5) with $\mathbf{Q}_t = \mathbf{I}_{2D}$. For the KF update, we collect the CS measurement vectors for the two neighboring views at time t and form a single measurement vector $\mathbf{y}_t = [\mathbf{y}_t^{k^{\mathrm{T}}} \mathbf{y}_t^{k+1^{\mathrm{T}}}]^{\mathrm{T}} \in \mathbb{R}^{2P}$. The same CS matrix $\mathbf{H} \in \mathbb{R}^{P \times D}$ is being used for each view, therefore $\mathbf{y}_t$ can be expressed as

$$\mathbf{y}_t = \widetilde{\mathbf{H}} \mathbf{x}_t + \mathbf{v}_t \tag{12}$$

where $\widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix}$ and $\mathbf{v}_t \in \mathbb{R}^{2P}$ is $2P$-dimensional independent zero-mean observation noise with covariance matrix $\mathbf{R}_t = \sigma^2 \mathbf{I}_{2P}$. The measurement update equations are then given by (6) and (7) with $\mathbf{H}_t$ replaced by $\widetilde{\mathbf{H}}$ and the Kalman gain $\mathbf{K}_t$ is given by

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \widetilde{\mathbf{H}}^{\mathrm{T}} (\widetilde{\mathbf{H}} \mathbf{P}_{t|t-1} \widetilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}_t)^{-1}. \tag{13}$$

After the KF update $\widehat{\mathbf{x}}_{t|t} = [\widehat{\mathbf{x}}_{t|t}^{k} \quad \widehat{\mathbf{x}}_{t|t}^{k+1}]$ is obtained, the decoder moves one view forward and performs joint-view motion-compensated or motion-disparity-compensated (MC/MDC) KF again for $\mathbf{x}_t = [\mathbf{x}_t^{k+1} \quad \mathbf{x}_t^{k+2}]^T$. Such sliding-window based joint-view MC/MDC-KF is conducted until all q views at time t are decoded. Finally, the "middle" views ($k = 2, ...q-1$) that have two available decodings are reconstructed by the average of the two decoding results.

4. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

In this section, we experimentally study the performance of the proposed joint-view KF decoder by evaluating the perceptual quality of the reconstructed multiview video sequences and their peak signal-to-noise ratio (PSNR). Three data sets, *Balloons*, *Bookarrival*, and *Ballet*, each with a resolution of 48×64 pixels are used. Each video sequence consists of $q = 5$ views captured by different cameras and each view contains $p = 50$ successive frames. For *Balloons* and *Bookarrival*, motion along the temporal direction is relatively slow and disparity difference between adjacent views is also relatively small. *Ballet* has faster motion and larger disparity differences. In all case studies, processing is carried out only on the luminance component.

At our independent and distributed CS encoder side, each frame of each view is handled as a vectorized column of length $N = 3072$ multiplied by the $P \times N$ Gaussian matrix $\mathbf{H}$. In our experiments, the CS ratio is set at $\frac{P}{N} = 0.5$. The elements of the captured P-dimensional measurement vector are corrupted by independent Gaussian noise of zero-mean and variance $\sigma^2 = 0.25$.

Initially, each frame is reconstructed independently by seeking a sparse representation in the 2-D DCT domain using the interior-point method [19]. With the initially reconstructed frames, the motion or motion-disparity field between adjacent frames (views) is estimated by the optical flow method [20].

In our experiments, we compare the performance of three decoders: Direct ℓ_1-norm minimization as in [11]; independent-view (one-view) motion-compensated Kalman filter decoder as in [17]; and the proposed joint-view motion-compensated or motion-disparity-compensated decoder (MC/MDC-KF) with two joint views. Fig. 2 shows the reconstructed 3rd view of the 25th frame ((a2)-(a4)) and 28th frame ((b2)-(b4)) of the *Balloons* and *Bookarrival* sequence, respectively. Since *Balloons* and *Bookarrival* have slow motion and small view differences, motion-disparity-compensated prediction is performed as in (10). It can be observed that the proposed joint-view (two views, herein) MDC-KF decoder provides much better visual quality. The same experiment is next carried out on the *Ballet* sequence. *Ballet* has fast motion and large view differences, therefore we perform motion-compensated prediction as in (9). A similar conclusion can

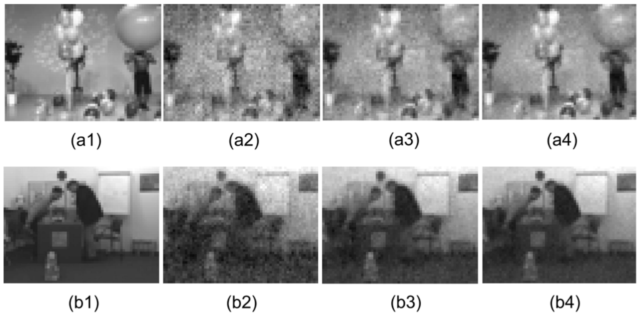

Fig. 2. Reconstructed *Balloons* ((a1)-(a4)) and *Bookarrival* ((b1)-(b4)) data sets: (a1, b1) Original frame; (a2, b2) ℓ_1-norm CS recovery; (a3, b3) single-view motion-compensated Kalman filtering; and (a4, b4) joint-view (two views) motion-disparity-compensated Kalman filtering ($P = 0.5N$).

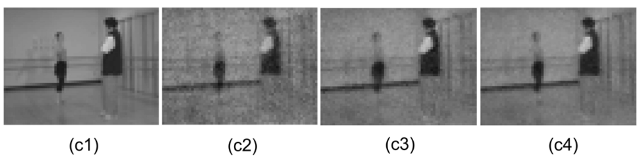

Fig. 3. Reconstructed *Ballet* data set: (c1) Original frame; (c2) ℓ_1-norm CS recovery; (c3) single-view motion-compensated Kalman filtering; and (c4) joint-view (two views) motion-compensated Kalman filtering ($P = 0.5N$).

be drawn from Fig. 3 that shows the reconstructed 3rd view of the 30th frame of the *Ballet* sequence. The joint-view MC-KF decoder offers significantly clearer reconstruction than the single-view MC-KF decoder and ℓ_1-norm CS decoding. Fig. 4 shows the decoder PSNR values versus frame/time index. KF-CS reconstructed PSNR values increase significantly as time elapses compared to conventional minimum ℓ_1-norm reconstruction. Compared to single-view MC-KF decoding, the proposed joint-view MDC-KF and joint-view MC-KF decoders offer as much as 4 dB PSNR improvement for the *Balloons* sequence, 3 dB improvement for the *Bookarrival* sequence, and 1.4 dB improvement for the *Ballet* sequence, respectively.

5. CONCLUSIONS

In this work, we developed and presented for the first time in the literature a joint-view Kalman filtering approach for causal reconstruction of compressed-sensed multiview video sequences. Conventional CS recovery is performed to obtain initial reconstructions followed by motion or motion-disparity estimation. Then, each pair of adjacent views is jointly updated by the motion-compensated or motion-disparity-compensated Kalman filter. Experimental results demonstrated that the proposed joint-view KF decoder outperforms significantly conventional minimum ℓ_1-norm reconstruction and single-view KF, both in visual perception and PSNR value.

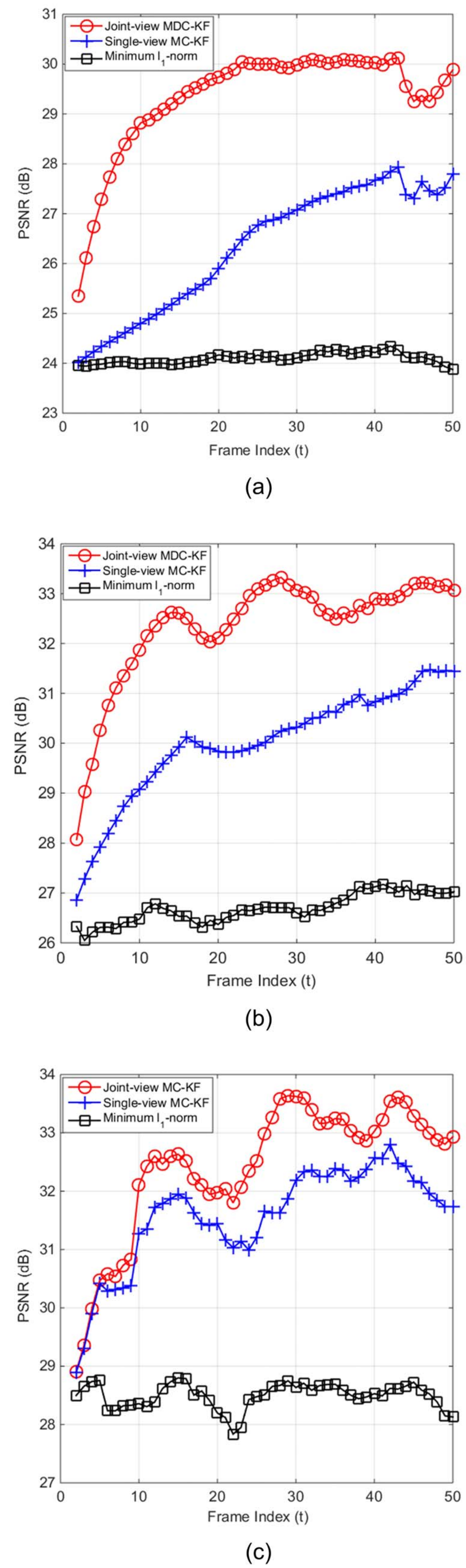

Fig. 4. Rate-distortion performance of (a) *Balloons*, (b) *Bookarrival*, and (c) *Ballet* data sets.

6. REFERENCES

[1] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406-5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, Apr. 2006.

[3] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, pp. 21-30, Mar. 2008.

[4] K. Gao, S. N. Batalama, D. A. Pados, and B. W. Suter, "Compressive sampling with generalized polygons," *IEEE Trans. Signal Process.*, vol. 59, pp. 4759-4766, Oct. 2011.

[5] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," vol. 59, pp. 365-373, Feb. 2008.

[6] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Optics Express*, vol. 17, pp. 6368-6388, Apr. 2009.

[7] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1393-1396.

[8] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 1169-1172.

[9] M. Trocan, E. W. Tramel, J. E. Fowler, and B. Pesquet, "Compressed-sensing recovery of multiview image and video sequences using signal prediction," *Multimed. Tools Appl.*, vol. 72, pp. 95-121, Sept. 2014.

[10] 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, Australia, Oct. 2015, pp. 1458-1462.

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

[12] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Proc. Picture Coding Symposium (PCS)*, Beijing, China, Apr. 2006, pp. 1-6.

[13] H. W. Chen, L. W. Kang, and C. S. Lu, "Dynamic measurement rate allocation for distributed compressive video sensing," in *Proc. Visual Comm. and Image Proc. (VCIP)*, Huang Shan, China, July 2010, pp. 1-10.

[14] Y. Liu, M. Li, and D. A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Trans. Circuits and Syst. for Video Tech.*, vol. 23, pp. 438-444, Mar. 2013.

[15] N. Vaswani, "Kalman fitered compressed sensing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, San Diego, CA, Oct. 2008, pp. 893-896.

[16] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MR image reconstruction using Kalman filtered compressed sensing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 393-396.

[17] X. Ding, W. Chen, and I. Wassell, "Generalized-KFCS: motion estimation enhanced kalman filtered compressive sensing for video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 1297-1301.

[18] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation [lecture notes]," *IEEE Signal Process. Magazine*, vol. 29, pp. 128-132, Sept. 2012.

[19] E. Candès and J. Romberg, "ℓ1-magic: Recovery of sparse signals via convex programming," URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf.

[20] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Doctoral Thesis. Massachusetts Institute of Technology, Boston, MA, May 2009.