Motion-Aware Decoding of Compressed-Sensed Video

Ying Liu, *Student Member, IEEE,* Ming Li, *Member, IEEE,* and Dimitris A. Pados, *Member, IEEE*

***Abstract*—Compressed sensing is the theory and practice of sub-Nyquist sampling of sparse signals of interest. Perfect reconstruction may then be possible with much fewer than the Nyquist required number of data. In this paper, in particular, we consider a video system where acquisition is carried out in the form of direct compressive sampling (CS) with no other form of sophisticated encoding. Therefore, the burden of quality video sequence reconstruction falls solely on the receiver side. We show that effective implicit motion estimation and decoding can be carried out at the receiver or decoder side via sparsity-aware recovery. The receiver performs sliding-window interframe decoding that adaptively estimates Karhunen–Loève bases from adjacent previously reconstructed frames to enhance the sparse representation of each video frame block, such that the overall reconstruction quality is improved at any given fixed CS rate. Experimental results included in this paper illustrate the presented developments.**

***Index Terms*—Compressed sensing, compressive sampling, dimensionality reduction, motion estimation, sparse representation, video codecs, video streaming.**

I. Introduction

CONVENTIONAL signal acquisition schemes follow the general Nyquist or Shannon sampling theory: to reconstruct a signal without error, the sampling rate must be at least twice as much as the highest frequency of the signal. Compressive sampling (CS), also referred to as compressed sensing, is an emerging body of work that deals with sub-Nyquist sampling of sparse signals of interest [1]–[3]. Rather than collecting an entire Nyquist ensemble of signal samples, CS can reconstruct sparse signals from a small number of (random [3] or deterministic [4]) linear measurements via convex optimization [5], linear regression [6], [7], or greedy recovery algorithms [8].

A somewhat extreme example of a CS application that has attracted much interest is the "single-pixel camera" architecture [9] where a still image can be produced from significantly fewer captured measurements than the number of desired or reconstructed image pixels. Arguably, a natural highly desirable next-step development is compressive video streaming. In this paper, we consider a video transmission system where the transmitter or encoder performs nothing more than compressed sensing acquisition without the benefits of the familiar sophisticated forms of video encoding. Such a setup may be of particular interest, e.g., in problems that involve large wireless multimedia networks of primitive low-complexity, low-cost video sensors. For video streaming across such networks, conventional predictive video encoding at individual sensors would be untenable when large deployments with power-limited devices are considered. CS can be viewed as a potentially enabling technology in this context [10], as video acquisition would require minimal or no computational power at all, yet transmission bandwidth would still be greatly reduced. In such a case, the burden of quality video reconstruction will fall solely on the receiver or decoder side.

The quality of the reconstructed video is determined by the number of collected measurements, which, based on CS principles, should be proportional to the sparsity level of the signal. Therefore, the challenge of implementing a well-compressed and well-reconstructed CS-based video streaming system rests on developing effective sparse representations and corresponding video recovery algorithms. Several important methods for CS video recovery have already been proposed, each relying on a different sparse representation. 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 under a 3-D DWT [12] or 2-D DWT applied on interframe difference data [13].

In standard video compression technology, effective encoder-based motion estimation (ME) is a defining matter in the feasibility and success of digital video. In the case of CS-only video acquisition that we study in this paper, ME can be exploited only at the receiver or decoder side. In current approaches [14], [15], a video sequence is divided into key frames and CS frames. While each key frame is reconstructed individually using a fixed basis (e.g., 2-D DWT or 2-D DCT), each CS frame is reconstructed conditionally using an adaptively generated basis from adjacent already reconstructed key frames. In our recent preliminary work [16], we proposed an iterative forward–backward decoding algorithm operating on successive pairs of frames where each pair of odd and even frames is reconstructed using adaptively generated Karhunen–Loève transform (KLT) bases.

Manuscript received December 5, 2011; revised February 29, 2012 and May 2, 2012; accepted June 3, 2012. Date of publication July 5, 2012; date of current version March 7, 2013. This work was supported by the National Science Foundation under Grant CNS-1117121. This paper was presented in part at the 17th International Conference on Digital Signal Processing, Corfu, Greece, July 2011 (invited paper). This paper was recommended by Associate Editor W. Zhang.

The authors are with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: yl72@buffalo.edu; mingli@buffalo.edu; pados@buffalo.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.2012.2207269

In this paper, we propose a new sparsity-aware video decoding algorithm for compressive video streaming systems to exploit long-term interframe similarities and pursue the most efficient and effective utilization of all available measurements. For each video frame, we operate block by block and recover each block using a KLT basis adaptively generated or estimated from previously reconstructed reference frame(s) defined in a fixed-width sliding window manner. The scheme essentially implements ME and motion compensation at the decoder by sparsity-aware reconstruction using interframe Karhunen–Loève basis estimation.

The remainder of this paper is organized as follows. In Section II, we briefly review the CS principles that motivate our compressive video streaming system. In Section III, the proposed sliding-window sparsity-aware video decoding algorithm is described in detail. Some experimental results are presented and analyzed in Section IV. Finally, a few conclusions are drawn in Section V.

II. CS Background and Formulation

In this section, we briefly review the CS principles for signal acquisition and recovery that are pertinent to our CS video streaming problem. A signal vector $\mathbf{x} \in \mathbb{R}^N$ can be expanded or represented by an orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ in the form of $\mathbf{x} = \mathbf{\Psi s}$. If the coefficients $\mathbf{s} \in \mathbb{R}^N$ have at most k nonzero components, we call $\mathbf{x}$ a k-sparse signal with respect to $\mathbf{\Psi}$. Many natural signals, images most notably, can be represented as a sparse signal in an appropriate basis.

Traditional approaches to sampling signals follow the Nyquist or Shannon theorem by which the sampling rate must be at least twice the maximum frequency present in the signal. CS emerges as an acquisition framework under which sparse signals can be recovered from far fewer samples or measurements than Nyquist. With a linear measurement matrix $\mathbf{\Phi}_{P\times N}$, $P \ll N$, CS measurements of a k-sparse signal $\mathbf{x}$ are collected in the form of

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

If the product of the measurement matrix $\mathbf{\Phi}$ and the basis matrix $\mathbf{\Psi}$, $\mathbf{A} \triangleq \mathbf{\Phi\Psi}$, satisfies the restricted isometry property (RIP) of order k [3], that is

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

holds for all k-sparse vectors $\mathbf{s}$ for a small "isometry" constant $0 < \delta_k < 1$, then the sparse coefficient vector $\mathbf{s}$ can be accurately recovered via the following linear program:

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\widetilde{\mathbf{s}}\|_{\ell_1} \qquad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}. \tag{3}$$

Afterward, the signal of interest $\mathbf{x}$ can be reconstructed by

$$\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\mathbf{s}}. \tag{4}$$

In most practical situations, $\mathbf{x}$ is not exactly sparse but approximately sparse and measurements may be corrupted by noise. Then, the CS acquisition or compression procedure can be formulated as

$$\mathbf{y} = \mathbf{\Phi\Psi s} + \mathbf{e} \tag{5}$$

where $\mathbf{e}$ is the unknown noise bounded by a known power amount $\|\mathbf{e}\|_{\ell_2} \le \epsilon$. To recover $\mathbf{x}$, we can use ℓ_1 minimization with relaxed constraint in the form of

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\widetilde{\mathbf{s}}\|_{\ell_1} \qquad \text{subject to} \quad \|\mathbf{y} - \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}\|_{\ell_2} \le \epsilon \tag{6}$$

which can be solved via convex optimization with computational complexity $\mathcal{O}(N^3)$. Specifically, if $\mathbf{A} \triangleq \mathbf{\Phi\Psi}$ satisfies RIP of order $2k$, that is

$$(1-\delta_{2k})||\mathbf{s}||_{\ell_2}^2 \le ||\mathbf{As}|| \le (1+\delta_{2k})||\mathbf{s}||_{\ell_2}^2 \tag{7}$$

holds for all $2k$-sparse vectors $\mathbf{s}$ with isometry constant $0 < \delta_{2k} < \sqrt{2}-1$ [3], then recovery by (6) guarantees

$$\|\widehat{\mathbf{s}} - \mathbf{s}\|_{\ell_2} \le c_0\|\mathbf{s} - \mathbf{s}_k\|_{\ell_1}/\sqrt{k} + c_1\epsilon \tag{8}$$

where c_0 and c_1 are positive constants and $\mathbf{s}_k$ is the k-term approximation of $\mathbf{s}$ by enforcing all but the largest k components of $\mathbf{s}$ to be zero.

Equivalently, the optimization problem in (6) can be reformulated as the following unconstrained problem:

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\mathbf{y} - \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}\|_{\ell_2}^2/2 + \lambda\|\widetilde{\mathbf{s}}\|_{\ell_1} \tag{9}$$

where λ is a regularization parameter that tunes the sparsity level. The problem in (9) can be efficiently solved via the least absolute shrinkage and selection operator (LASSO) algorithm [6], [7] with computational complexity $\mathcal{O}(P^2N)$. Again, after we obtain $\widehat{\mathbf{s}}$, $\mathbf{x}$ can be reconstructed by (4). As for selecting a proper measurement matrix $\mathbf{\Phi}$, it is known [3] that with overwhelming probability probabilistic construction of $\mathbf{\Phi}$ with entries drawn from independent and identically distributed (i.i.d.) Gaussian random variables with mean 0 and variance $1/P$ obeys RIP provided that $P \ge c\cdot k\log(N/k)$. For deterministic measurement matrix constructions, the reader is referred to [4] and references therein.

III. Proposed CS Video Decoding System

The CS-based signal acquisition technique described in Section II can be applied to video acquisition on a frame-by-frame, block-by-block basis. In the simple compressive video encoding block diagram shown in Fig. 1, each frame F_t, $t = 1, 2, \ldots$, is virtually partitioned into M nonoverlapping blocks of pixels with each block viewed as a vectorized column of length N, $\mathbf{x}_t^m \in \mathbb{R}^N$, $m = 1, \ldots, M$, $t = 1, 2, \ldots$. CS is performed by projecting $\mathbf{x}_t^m$ onto a $P \times N$ random measurement matrix $\mathbf{\Phi}$

$$\mathbf{y}_t^m = \mathbf{\Phi x}_t^m \tag{10}$$

with the entries of $\mathbf{\Phi}$ drawn from i.i.d. Gaussian random variables of zero mean and unit variance. Then, the resulting measurement vector $\mathbf{y}_t^m \in \mathbb{R}^P$ is processed by a fixed-rate uniform scalar quantizer. The quantized indices $\widetilde{\mathbf{y}}_t^m$ are encoded and transmitted to the decoder.

In the CS video decoder of [11], each frame is individually decoded via sparse signal recovery algorithms with fixed bases such as block-based 2-D DCT (or frame-based 2-D DWT). With a received (dequantized) measurement vector $\widehat{\mathbf{y}}$ and

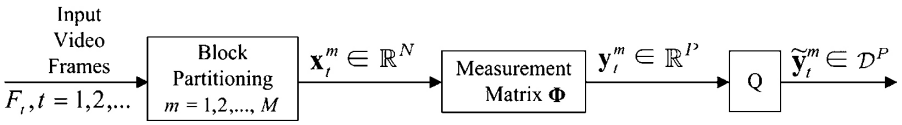


Fig. 1. Simple compressed sensing video encoder system with quantization alphabet $\mathcal{D}$.

a block-based 2-D DCT basis $\mathbf{\Psi}_{\text{DCT}}$, video reconstruction becomes an optimization problem as in (9)

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\widehat{\mathbf{y}} - \mathbf{\Phi}\mathbf{\Psi}_{\text{DCT}}\widetilde{\mathbf{s}}\|_{\ell_2}^2/2 + \lambda\|\widetilde{\mathbf{s}}\|_{\ell_1} \tag{11}$$

where the original video block $\mathbf{x}$ is recovered as

$$\widehat{\mathbf{x}} = \mathbf{\Psi}_{\text{DCT}}\widehat{\mathbf{s}}. \tag{12}$$

However, such intraframe decoding using a fixed basis does not provide sufficient sparsity level for the video block signal. Consequently, higher number of measurements is needed to ensure a required level of reconstruction quality. To enhance sparsity, in [12], the correlation among successive frames was exploited by jointly recovering several frames with a 3-D DWT basis, assuming that the video signal is more sparsely represented in a 3-D DWT domain. In [13], a sparser representation is provided by exploiting small interframe differences within a spatial 2-D DWT basis. Nevertheless, in all cases, these decoders cannot pursue or capture local motion effects that can significantly increase sparseness and are well known to be a critical attribute to the effectiveness of conventional video compression. Below, we propose and describe a new motion-capturing sparse decoding approach.

The founding concept of the proposed CS video decoder is shown in Fig. 2. The decoder consists of an initialization stage that decodes F_t, $t = 1, 2$, and a subsequent operational stage that decodes F_t, $t \geq 3$. At the initialization stage, F_1 is first reconstructed using the block-based fixed DCT basis exactly as described in (11) and (12). Then, we attempt to reconstruct each block of F_2 based on the reconstructed previous frame $\widehat{F}_1$. Our sparsity-aware ME decoding approach is based on the fact that the pixels of a block in a video frame may be satisfactorily predicted by using a linear combination of a small number of nearby blocks in adjacent (previous or next) frame(s). In particular, for our setup, the blocks in F_2 may be sparsely represented by a few neighboring blocks in $\widehat{F}_1$. We propose to use the KLT basis for this representation. The KLT is a linear transform where the basis vectors are deduced from the statistical properties of the frame block data and are, thus, data adaptive. KLT is optimal in the sense of energy compaction, i.e., it captures as much energy as possible in as few coefficients as possible [17]. For each block $\mathbf{x}_2^m$ in F_2, $m = 1, \ldots, M$, a group of neighboring blocks that lie in a window of a square $w \times w$ region centered at $\mathbf{x}_1^m$ are extracted from $\widehat{F}_1$. Then, the KLT basis for $\mathbf{x}_2^m$, $\mathbf{\Psi}_{2,\text{KLT}}^m$, is formed by the eigenvectors of the correlation matrix of the extracted blocks from $\widehat{F}_1$. Fig. 3 illustrates the block extraction procedure. Given a block $\mathbf{x}_2^m$ to estimate or reconstruct (block in bold of size $\sqrt{N}\times\sqrt{N}$ in F_{t+1}, $t = 1$, of Fig. 3), one can find its colocated block $\widehat{\mathbf{x}}_1^m$ (block in bold of size $\sqrt{N}\times\sqrt{N}$ in $\widehat{F}_t$, $t = 1$). Neighboring blocks (other overlapping blocks of size

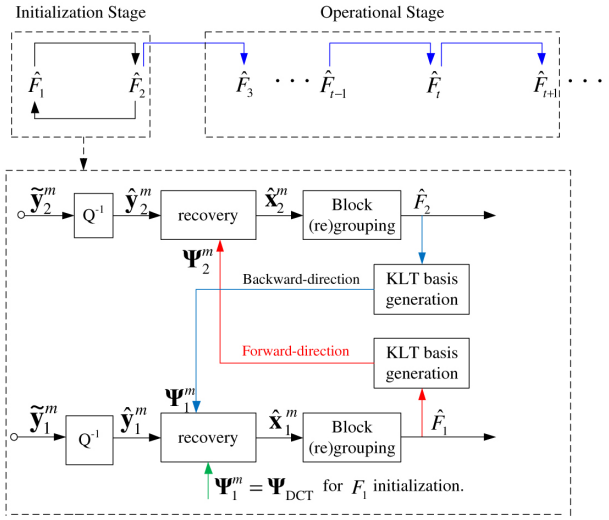


Fig. 2. Proposed CS decoder system (first-order decoding algorithm).

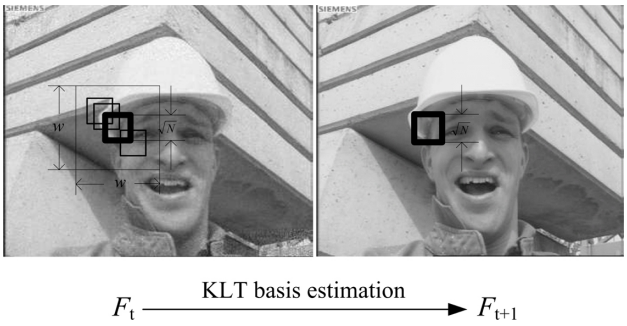


Fig. 3. KLT basis estimation illustration (first-order decoding).

$\sqrt{N}\times\sqrt{N}$ in $\widehat{F}_t$, $t = 1$) $\mathbf{d}_i$, $i = 1, \ldots, B$, can be extracted from a $w \times w$ area carrying out one-pixel shifts in all directions. When, say, e.g., w equals three times the block width $\sqrt{N}$ and block $\mathbf{x}_2^m$ is well in the interior of F_2, then the total number of available neighboring blocks is $B = (w - \sqrt{N})^2$; for blocks near the edge of F_2, B will be accordingly smaller.

Considering now all the extracted neighboring blocks as different realizations of an underlying vector stochastic process, the correlation matrix can be estimated by the sample average

$$\widehat{\mathbf{R}}_2^m = \frac{1}{B}\sum_{i=1}^{B} \mathbf{d}_i\mathbf{d}_i^T. \tag{13}$$

We form the KLT basis for Frame 2, block m, $\mathbf{\Psi}_{2,\text{KLT}}^m$, by the eigenvectors of $\widehat{\mathbf{R}}_2^m = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$

$$\mathbf{\Psi}_{2,\text{KLT}}^m = \mathbf{Q} \tag{14}$$

where $\mathbf{Q}$ is the matrix with columns, the eigenvectors of $\widehat{\mathbf{R}}_2^m$, and Λ is the diagonal matrix with the corresponding

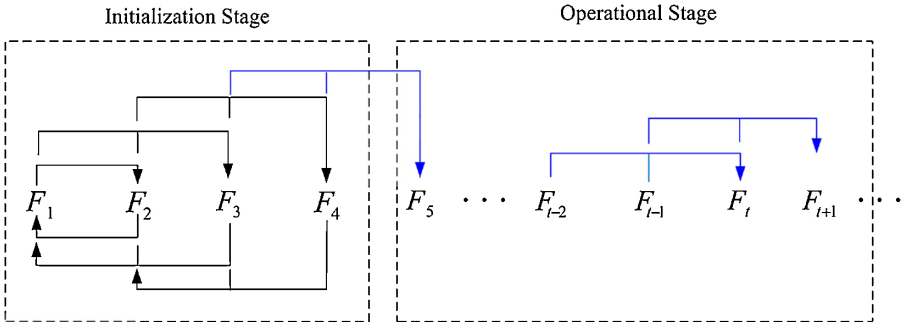


Fig. 4. CS decoder of order 2.

eigenvalues. The computational complexity, therefore, for the calculation of the KLT basis is $\mathcal{O}(N^3)$ per block. Next, using the (dequantized) measurement vector $\widehat{\mathbf{y}}_2^m$, we recover the sparse coefficients $\mathbf{s}_2^m$ by solving

$$\widehat{\mathbf{s}}_2^m = \arg\min_{\widetilde{\mathbf{s}}} \|\widehat{\mathbf{y}}_2^m - \mathbf{\Phi}\mathbf{\Psi}_{2,\mathrm{KLT}}^m \widetilde{\mathbf{s}}\|_{\ell_2}^2/2 + \lambda\|\widetilde{\mathbf{s}}\|_{\ell_1} \tag{15}$$

and we reconstruct the video block $\mathbf{x}_2^m$ by

$$\widehat{\mathbf{x}}_2^m = \mathbf{\Psi}_{2,\mathrm{KLT}}^m \widehat{\mathbf{s}}_2^m. \tag{16}$$

After all M blocks are reconstructed, they are grouped again to form the complete decoded frame $\widehat{F}_2$.

So far, during the initialization stage, we have carried out forward only frame F_2 reconstruction accounting for motion from the DCT reconstructed frame F_1. For improved initialization, we may repeat the algorithm backward and reconstruct again F_1 using KLT bases generated from $\widehat{F}_2$. This forward–backward approach iterates for the two initial frames F_1 and F_2 only, as shown in some detail in Fig. 2, until no significant further reconstruction quality improvement can be achieved.[1]

At the normal operational stage that follows, the decoder recovers the blocks of F_t, $t \geq 3$, based on the KLT bases estimated from $\widehat{F}_{t-1}$. Since only one previous reconstructed frame is used as the reference frame in KLT bases estimation, we refer to this approach as first-order sparsity-aware ME decoding.

To exploit the correlation within multiple successive frames and achieve higher ME effectiveness in decoding, we may extend the first-order sparsity-aware ME decoding algorithm to an nth-order procedure. At the initialization stage, the first $2n$ only frames are recovered via forward–backward KLT basis estimation from n reconstructed (previous or next) frames. Then, at the operational stage each frame F_t, $t \geq 2n+1$, is recovered from the previous n reconstructed frames. For illustration purposes, Fig. 4 depicts the order $n = 2$ scheme. At the initialization stage, F_1 and F_2 are first reconstructed with forward–backward estimation as in first-order decoding. Then, F_3 is decoded with KLT bases estimated from both $\widehat{F}_1$ and $\widehat{F}_2$. After $\widehat{F}_3$ is obtained, F_1 is decoded again in the backward direction with KLT bases estimated from both $\widehat{F}_2$ and $\widehat{F}_3$. The same second-order decoding is performed in the forward direction for F_4 and in the backward direction for F_2, so that each of the initial frames F_t, $1 \leq t \leq 4$, has been reconstructed with implicit ME from two adjacent frames (Fig. 4). In the subsequent operational stage, each frame F_t ($t \geq 5$) is decoded by the two previous reconstructed frames $\widehat{F}_{t-1}$ and $\widehat{F}_{t-2}$. The concept is immediately generalizable to nth-order decoding with $2n$ initial frames $F_1, F_2, \ldots, F_{2n}$. Empirically, the iterative initialization for the first-order and second-order schemes converges after four iterations with no significant further reconstruction quality improvement thereafter. For higher-order decoders (i.e., $n \geq 3$), initialization convergence is attained typically after two iterations.

[1]Therefore, the decoder proposed in [16] is used herein to provide the initialization stage.

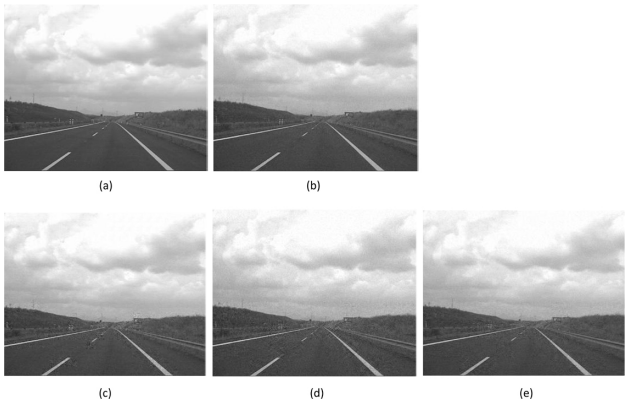


Fig. 5. Different decodings of the 54th frame of *Highway*. (a) Original frame. (b) Using the proposed order-10 sparsity-aware ME decoder. (c) Using the K-SVD basis decoder [14]. (d) Using the 2-D DWT basis interframe decoder [13]. (e) Using the 2-D DCT basis intraframe decoder [11] ($P = 0.625N$).

A defining characteristic of the proposed CS video decoder in comparison with existing CS video literature [11]–[16], [18]–[20] is that the order-n sliding-window decoding algorithm utilizes the spatial correlation within a video frame and the temporal correlation between successive video frames, which essentially results to implicit joint spatial-temporal motion-compensated video decoding. The adaptively generated block-based KLT basis provides a much sparser representation basis than fixed block-based basis approaches [11]–[13], [18] and the size-K singular value decomposition (K-SVD) adaptive basis approach [14] as demonstrated experimentally in the following section.

IV. Experimental Results

In this section, we experimentally study the performance of the proposed CS video decoders by evaluating the peak signal-to-noise ratio (PSNR) (as well as the perceptual quality) of

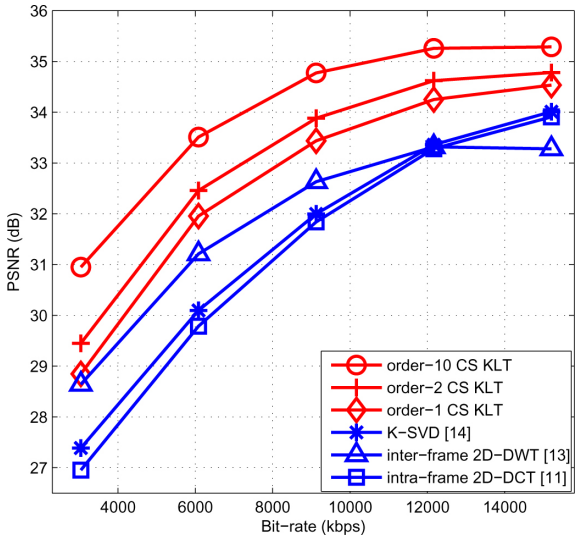


Fig. 6. Rate-distortion studies on the *Highway* sequence.

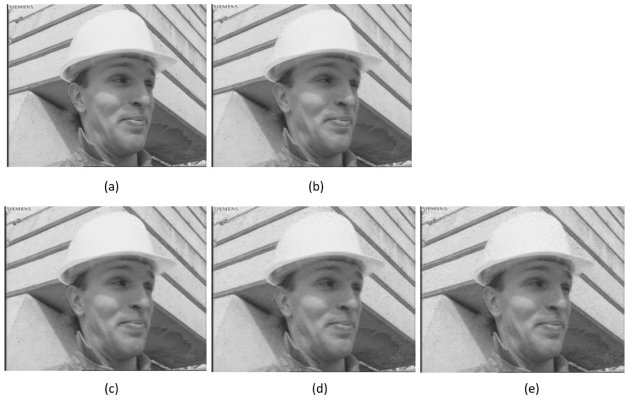


Fig. 7. Different decodings of the sixth frame of *Foreman*. (a) Original frame. (b) Using the proposed order-10 sparsity-aware ME decoder. (c) Using the K-SVD basis decoder [14]. (d) Using the 2-D DWT basis interframe decoder [13]. (e) Using the 2-D DCT basis intraframe decoder [11] ($P = 0.625N$).

reconstructed video sequences. Three test sequences, *Highway*, *Foreman*, and *Container*, with a CIF resolution of 352×288 pixels and frame rate of 30 f/s are used. Processing is carried out only on the luminance component.

At the trivial CS encoder side, each frame is partitioned into nonoverlapping blocks of 32×32 pixels. Each block is viewed as a vectorized column of length $N = 1024$ and multiplied by a $P \times N$ measurement matrix with elements drawn from i.i.d. zero-mean, unit-variance Gaussian random variables. The elements of the captured P-dimensional measurement vector are quantized individually by an 8-bit uniform scalar quantizer and then transmitted to the decoder. In our experiments, $P = 128$, 256, 384, 512, and 640 are used to provide the corresponding bit rates 3041.28, 6082.56, 9123.84, 12165.12, and 15206.4 kb/s. With an Intel i5-2410M 2.30 GHz processor, the encoding time per frame is well within 0.1 of a second, while the H.264/AVC JM reference software programmed in C++ requires about 1.55 s with low-complexity configurations [21].

At the decoder side, we choose the LASSO algorithm [6], [7] for sparse recovery motivated by its low-complexity and satisfactory recovery performance characteristics. While

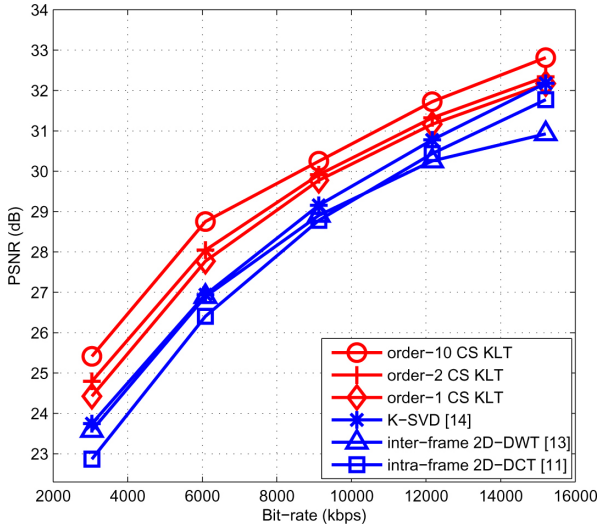


Fig. 8. Rate-distortion studies on the *Foreman* sequence.

at the operational stage each frame is reconstructed from individually compressed-sensed frame data, the reconstruction bases (KLT) come from previously reconstructed frames. Hence, some decreasing effectiveness of the estimated content-adaptive bases will be experienced. To enhance operational robustness of the proposed decoder to basis degradation, we perform reinitialization (repeat the initialization stage) after every 20 frames. In our experimental studies, three proposed CS video decoders are examined for all three sequences: order-1, order-2, and order-10 sparsity-aware ME decoding. For comparison purposes, we also include three existing typical CS video decoders:[2] fixed 2-D DCT basis intraframe decoder used as a reference benchmark [11], fixed 2-D DWT basis interframe decoder [13], and fixed 2-D DWT basis odd-frame decoding combined with global K-SVD trained basis even-frame decoding [14].[3]

Fig. 5 shows the decodings of the 54th frame of *Highway* produced by the order-10 CS decoder [Fig. 5(b)], K-SVD basis decoder [14] [Fig. 5(c)], 2-D DWT basis interframe decoder [13] [Fig. 5(d)], and the 2-D DCT basis intraframe decoder [11] [Fig. 5(e)]. It can be observed that the fixed basis interframe or intraframe decoders as well as the K-SVD basis decoder suffer noticeable performance loss over the whole image, while the proposed order-10 sparsity-aware ME decoder demonstrates considerable reconstruction quality improvement.[4]

Fig. 6 shows the rate-distortion characteristics of the six decoders for the *Highway* video sequence. The PSNR values (in dB) are averaged over 100 frames. Evidently, the proposed order-1 sparsity-aware ME decoder outperforms significantly the fixed basis interframe or intraframe decoders, as well as

[2]The video decoders in [19] and [20] both utilize H.264 instead of pure CS video acquisition and cannot be included in these comparisons.

[3]The video decoder proposed in [14] is an advanced version of the decoder in [18], therefore, [18] is not considered herein.

[4]As usual, portable document formatting of this paper tends to dampen perceptual quality differences between Figs. 5(a)–(e) that are in fact pronounced in video playback. Fig. 6 is the usual attempt to capture average differences quantitatively.

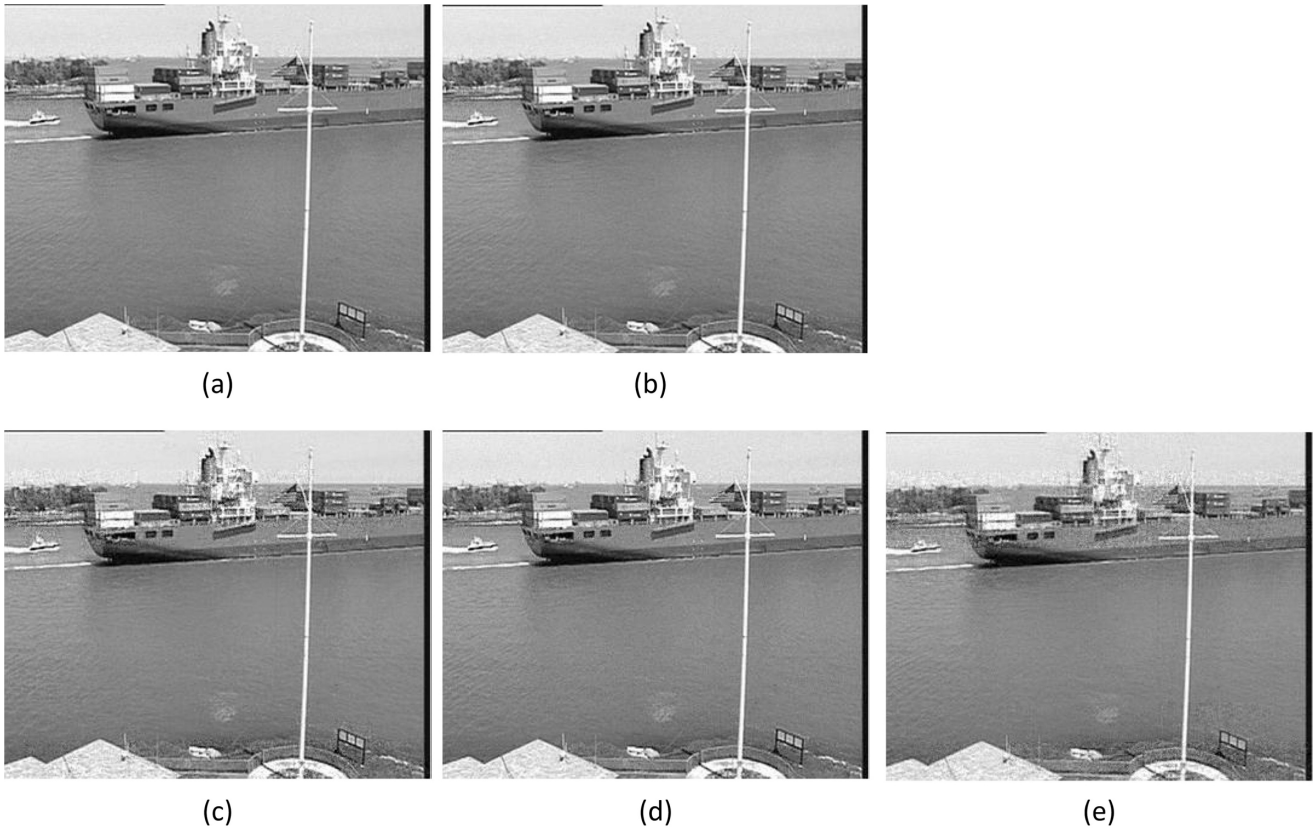


Fig. 9. Different decodings of the 28th frame of *Container*. (a) Original frame. (b) Using the proposed order-10 sparsity-aware ME decoder. (c) Using the K-SVD basis decoder [14]. (d) Using the 2-D DWT basis interframe decoder [13]. (e) Using the 2-D DCT basis intraframe decoder [11] ($P = 0.625N$).

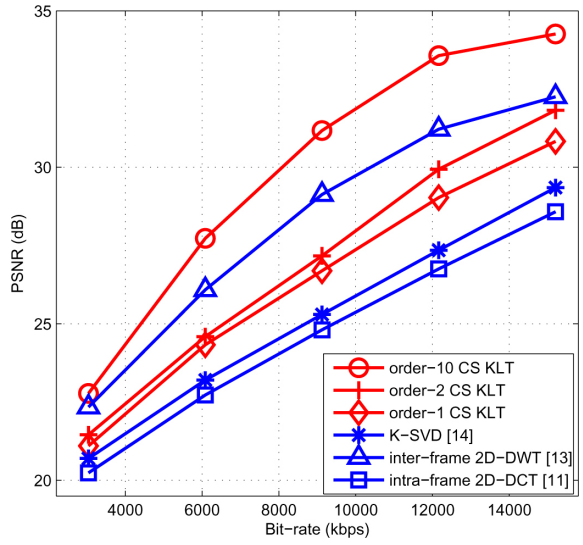


Fig. 10. Rate-distortion studies on the *Container* sequence.

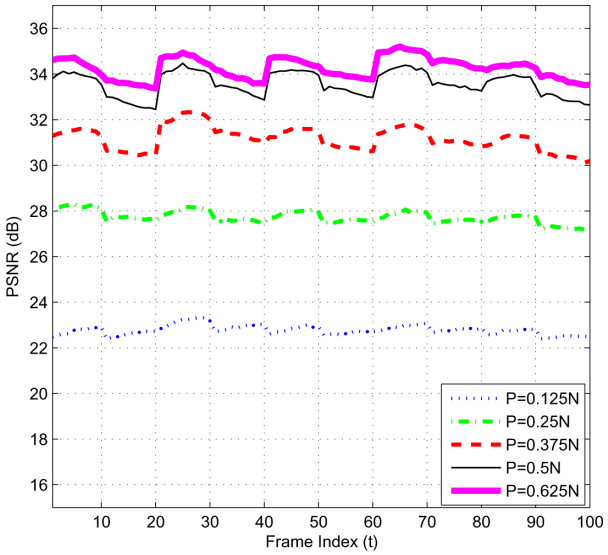


Fig. 11. Per-frame error rate of *Container* sequence (order-10 decoding, reinitialization every 20 frames).

the K-SVD basis decoder at the low-to-medium bit rate ranges of interest with gains as much as 1.5 dB. The second-order and tenth-order proposed decoders further improve performance by up to 2 dB.

The same rate-distortion performance study is repeated in Figs. 7 and 8 for the *Foreman* sequence. By Fig. 8, the proposed first-order sparsity-aware ME decoder again outperforms the fixed basis interframe or intraframe decoders and K-SVD basis decoder. The performance is enhanced by as much as 1 dB as the decoder order increases to 10. Similar conclusions can be drawn by Figs. 9 and 10 (*Container* sequence, and order-1, order-2, and order-10 proposed CS decoding).

Finally, Fig. 11 depicts the per-frame error characteristics of the *Container* sequence under the proposed order-10 decoding. It can be observed that KLT basis degradation is effectively mitigated with reinitialization every 20 frames.

V. Conclusion

We proposed a sparsity-aware motion-accounting decoder for video streaming systems with plain CS encoding. The decoder performed sliding-window interframe decoding that adaptively generated KLT bases from adjacent previously reconstructed frames to enhance the sparse representation of each video frame block, such that the overall reconstruction quality is improved at any given fixed CS rate. Experimental results demonstrated that the proposed sparsity-aware decoders significantly outperform the conventional fixed-basis intraframe and interframe, as well as the K-SVD, decoders. Performance was improved as the number of reference frames (what we call "decoder order") increases, with order values in the range of 2–10 appearing as a good compromise between computational complexity and reconstruction quality. In terms of future work, to further reduce the decoder complexity and improve video reconstruction quality, we may seek other effective and efficient basis representations and recovery algorithms, together with rate-adaptive compressive sensing at the encoder. Measurement matrices of deterministic design may also be pursued to facilitate efficient encoding or decoding.

Acknowledgment

The authors would like to thank the Associate Editor W. Zhang and the five anonymous reviewers for their comments and suggestions that helped improve this manuscript significantly, both in presentation and in content.

References

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

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

[3] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, 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, no. 10, pp. 4759–4766, Oct. 2011.

[5] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, pp. 407–451, Apr. 2004.

[8] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[9] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[10] S. Pudlewski, T. Melodia, and A. Prasanna, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comp.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2011.

[11] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Proc. Eur. Signal Proc. Conf. (EUSIPCO)*, Aug. 2008.

[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. PCS*, Apr. 2006, pp. 711–716.

[13] R. F. Marcia and R. M. Willet, "Compressive coded aperture video reconstruction," in *Proc. Eur. Signal Proc. Conf. (EUSIPCO)*, Aug. 2008.

[14] H. W. Chen, L. W. Kang, and C. S. Lu, "Dynamic measurement rate allocation for distributed compressive video sensing," in *Proc. VCIP*, Jul. 2010, pp. 1–10.

[15] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. PCS*, May 2009, pp. 197–200.

[16] Y. Liu, M. Li, K. Gao, and D. A. Pados, "Motion compensation as sparsity-aware decoding in compressive video streaming," in *Proc. 17th Int. Conf. DSP*, Jul. 2011, pp. 1–5.

[17] Z.-N. Li and M. S. Drew, *Fundamentals of Multimedia*. Upper Saddle River, NJ: Pearson, Prentice-Hall, 2004.

[18] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 1393–1396.

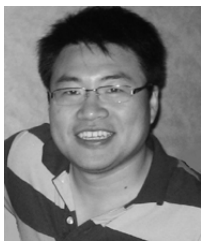
[19] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. PCS*, May 2009, pp. 165–168.

[20] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. IEEE ICIP*, Nov. 2009, pp. 1169–1172.

[21] I. E. Richardson, *The H.264 Advanced Video Compression Standard*, 2nd ed. New York: Wiley, 2010.

Ying Liu (S'11) received the B.S. degree in telecommunications engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2006, and the M.S. degree in electrical engineering from the State University of New York at Buffalo, Buffalo, in 2008, where she is currently pursuing the Ph.D. degree in electrical engineering.

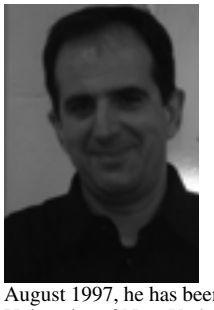
She is currently a Research Assistant with the Signals, Communications, and Networking Research Group, State University of New York at Buffalo. Her research interests include video streaming, compressed sensing, and adaptive signal processing.



Ming Li (M'11) received the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo, Buffalo, in 2005 and 2010, respectively.

He is currently a Post-Doctoral Research Associate with the Signals, Communications, and Networking Research Group, Department of Electrical Engineering, State University of New York at Buffalo. His current research interests include wireless multiple access communications, multimedia and covert communications, statistical signal processing, and cognitive radios.

Dr. Li is a member of the IEEE Communications and Signal Processing Societies.



Dimitris A. Pados (M'95) was born in Athens, Greece, on October 22, 1966. He received the Diploma degree in computer science and engineering (five-year program) from the University of Patras, Patras, Greece, in 1989, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1994.

From 1994 to 1997, he was an Assistant Professor with the Department of Electrical and Computer Engineering and the Center for Telecommunications Studies, University of Louisiana, Lafayette. Since August 1997, he has been with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, where he is currently a Professor. He served the Department as an Associate Chair from 2009 to 2010. He was elected the University Faculty Senator thrice, from 2004 to 2006, 2008–2010, and 2010–2012, and was a Faculty Senate Executive Committee Member from 2009 to 2010. His current research interests include communication theory and adaptive signal processing with applications to interference channels, signal waveform design, compressive sampling, and multimedia communications.

Dr. Pados was a recipient of the IEEE International Conference on Telecommunications Best Paper Award in 2001, the IEEE Transactions on Neural Networks Outstanding Paper Award in 2003, and the IEEE International Communications Conference Best Paper Award in Signal Processing for Communications in 2010, as a co-author, the SUNY-System-wide Chancellor's Award for Excellence in Teaching in 2009, and the University at Buffalo Exceptional Scholar Sustained Achievement Award in 2011. He is a member of the IEEE Signal Processing, Communications, Information Theory, and Computational Intelligence Societies. He was an Associate Editor of the IEEE Signal Processing Letters from 2001 to 2004 and the IEEE Transactions on Neural Networks from 2001 to 2005.