Rate-distortion optimization for compressive video sampling

Ying Liu, Krishna Rao Vijayanagar, and Joohee Kim

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616

ABSTRACT

The recently introduced compressed sensing (CS) framework enables low complexity video acquisition via sub-Nyquist rate sampling. In practice, the resulting CS samples are quantized and indexed by finitely many bits (bit-depth) for transmission. In applications where the bit-budget for video transmission is constrained, rate-distortion optimization (RDO) is essential for quality video reconstruction. In this work, we develop a double-level RDO scheme for compressive video sampling, where frame-level RDO is performed by adaptively allocating the fixed bit-budget per frame to each video block based on block-sparsity, and block-level RDO is performed by modelling the block reconstruction peak-signal-to-noise ratio (PSNR) as a quadratic function of quantization bit-depth. The optimal bit-depth and the number of CS samples are then obtained by setting the first derivative of the function to zero. In the experimental studies the model parameters are initialized with a small set of training data, which are then updated with local information in the model testing stage. Simulation results presented herein show that the proposed double-level RDO significantly enhances the reconstruction quality for a bit-budget constrained CS video transmission system.

Keywords: Rate-distortion optimization, bit-budget, bit-depth, compressive sampling, compressed sensing, sub-Nyquist rate, video acquisition, video reconstruction

1. INTRODUCTION

Compressive sampling (CS), also referred to as compressed sensing, is an emerging bulk 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]. An example of a CS application that has attracted much interest is compressive video sampling (CVS) systems, where the encoder performs nothing more than compressed sensing acquisition without the benefits of the familiar sophisticated forms of transform-based video encoding, and the decoder reconstructs video signals via sparsity-aware decoding. Such a set-up may be of particular interest, for example, in problems that involve large wireless multimedia networks of primitive low-complexity, low-cost video sensors, where conventional predictive video encoding at individual sensors would be untenable when large deployments with power limited devices are considered [9].

In practice, several issues may arise in CVS systems. First, the real-valued CS samples will be mapped to discrete bits via a quantizer. Second, the transmission bandwidth is often limited, imposing a constraint on the number of bits used for CS sample quantization. Third, for successful reconstruction, the number of samples should be proportional to signal sparsity captured by the utilized sparsifying basis. Thus, a tradeoff exists between the number of CS samples and the number of quantization bits per sample (bit-depth) in bit-budget constrained CVS systems. To obtain high quality video playback, rate-distortion optimization (RDO) is necessary at the encoder to assign the optimal number of CS samples and quantization bit-depth to minimize the reconstruction distortion. In existing CVS systems, quantization bit-depth is usually fixed, while RDO is performed by adaptively allocating the available number of CS samples per frame for each block based on

Further author information: (Send correspondence to Y.L.)
Y.L.: E-mail: yliu81@iit.edu, Telephone: 1 312 567 3421
K.R.V.: E-mail: kvijayan@hawk.iit.edu, Telephone: 1 312 567 3421
J.K.: E-mail: joohee@ece.iit.edu, Telephone: 1 312 567 3421

Compressive Sensing III, edited by Fauzia Ahmad, Proc. of SPIE Vol. 9109,
91090R · © 2014 SPIE · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2053407

block-sparsity [10],[11], or rate control is achieved by adapting the number of CS samples according to network conditions [9]. Nevertheless, the influence of quantization bit-depth was not investigated in these studies.

In this paper, we consider a CVS system where both frame-level and block-level RDO are considered. Since the number of CS measurements required for quality reconstruction is proportional to the signal sparsity captured by the utilized basis, and different areas in a video sequence or video frame may have different sparsity, frame-level RDO can be achieved by adaptively allocating the total number of available CS measurements per frame to each frame block based on block sparsity level*.

Afterwards, block-level RDO is performed within each block via sparsity-adaptive bit-depth quantization. On the one hand, we can increase the bit-depth as we decrease the number of samples, thereby increasing the precision of each sample. On the other hand, we can decrease the bit-depth so that the requirement for accurate ℓ_1-based reconstruction can be satisfied with more number of samples. Intuitively, it is straightforward to use a large number of samples for low-sparsity blocks at the expense of sample accuracy, and a small number of samples for highly sparse blocks, where each sample is sufficiently accurate by using a large bit-depth for quantization. The key issue is to find an accurate relationship between block-sparsity and an optimal combination of bit-depth and the number of CS samples that leads to maximal reconstruction quality. In the present work, we propose a doubly-model to describe the relationship among block-sparsity, bit-depth for CS sample quantization, and the reconstruction quality under certain block bit-budget constraint. First, the reconstruction peak signal-to-noise ratio (PSNR) is modeled as a quadratic function of bit-depth B (PSNR-B model). Second, each parameter of the PSNR-B model is approximated as an individual function of block-sparsity measured in block spatial total-variation (TV) [11]. Hence, the optimal bit-depth for a particular video block can be obtained by fitting its spatial TV into the doubly-model and setting the first derivative of the PSNR-B model to zero.

The remainder of this paper is organized as follows. In Section 2, we briefly review the CVS background. In Section 3, the proposed double-level RDO is described in detail. Experimental results and comparison studies are presented in Section 4 and finally, a few conclusions are drawn in Section 5.

2. COMPRESSIVE VIDEO SAMPLING BACKGROUND

Traditional approaches to sampling signals follow the Nyquist/ 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 $\mathbf{\Phi}_{P\times L}, P \ll L$, CS samples of a signal $\mathbf{x} \in \mathbb{R}^L$ with at most k non-zeros coefficients in basis $\mathbf{\Psi}$ 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], i.e.

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

holds for all k-sparse vectors $\mathbf{s}$ for a small "isometry" constant δ_k, then the sparse coefficient vector $\mathbf{s}$ can be accurately recovered via the following convex optimization program

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

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

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

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 independently and identically distributed (i.i.d.) Gaussian random

*The quantization bit-depth is assumed to be constant for each CS measurement of each frame block in frame-level RDO.

variables with mean 0 and variance $1/P$ obeys RIP with any basis $\mathbf{\Psi}$ provided that $P \geq c \cdot k \log(L/k)$, where c is some constant depending on each instance.

When compressed sensing is applied to video compression, video frames are typically divided into non-overlapping blocks [10]-[14] and each block in vectorized form $\mathbf{x}$ is encoded via compressed sensing. For practical transmission systems, the CS samples are quantized into finite number of bits. Then, the CS acquisition/compression procedure can be formulated as

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

where $\mathbf{\Psi}$ can be any sparsifying basis such as the 2-D DCT basis [12] or adaptively generated bases [10],[13], and [14], and $\mathbf{e}$ is the quantization noise bounded by a known power amount $\|\mathbf{e}\|_{\ell_2} \leq \epsilon$. For uniform scalar quantization with quantization step size Δ, $\epsilon = \Delta\sqrt{\frac{12}{P}}$. To recover $\mathbf{x}$, we can use ℓ_1 minimization with quadratic constraint in the form of

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

which can be recast as a second-order cone program and solved using a log-barrier algorithm [15]. Again, after we obtain $\widehat{\mathbf{s}}$, $\mathbf{x}$ can be reconstructed by (4).

3. PROPOSED DOUBLE-LEVEL RATE-DISTORTION OPTIMIZATION

In the proposed CVS system, each video frame is assigned a fixed bit-budget of $\mathfrak{B}^{\mathrm{f}}$ bits. In frame-level RDO, the bit-depth per CS sample is assumed a constant B_{c}, leading to a fixed total number of CS samples per frame $P_{\mathrm{total}} = \frac{\mathfrak{B}^{\mathrm{f}}}{B_{\mathrm{c}}}$. Assuming there are a total number of M non-overlapping blocks of equal size per frame, these CS samples can be adaptively allocated to each block based on block sparsity measured in two-dimensional total-variation† (2D-TV) in the following form

$$P_m^{\mathrm{f}} = P_{\mathrm{total}} \times \frac{(\mathrm{TV}_{\mathrm{2D}}(\mathbf{X}_m))^{\alpha}}{\sum_{m=1}^{M} (\mathrm{TV}_{\mathrm{2D}}(\mathbf{X}_m))^{\alpha}} \tag{7}$$

where m is the block index, and α is a decision constant between 0 and 1. Therefore, the bit-budget for the m^{th} block is $\mathfrak{B}_m = B_{\mathrm{c}} \times P_m^{\mathrm{f}}$.

To perform the block-level RDO for the m^{th} block, we can vary the actual number of CS samples P_m and quantization bit-depth B_m as long as their product $B_m \times P_m$ equals the block bit-budget $\mathfrak{B}_m$ assigned in the frame-level RDO. As shown in Fig. 1 (a), under the fixed block bit-budget $\mathfrak{B}$, the PSNR of the block reconstructed via (6) does not increase monotonically with bit-depth, rather, it decreases when the bit-depth is larger than a certain number. In addition, the empirical study in Fig. 1 (b) reveals that the true optimal bit-depth that leads to maximal reconstruction PSNR tends to be large if the block has small spatial TV, and tends to be small if the block has large spatial TV, which indicates that adaptive bit-depth quantization based on block spatial TV would potentially perform block-level RDO. In order to associate the optimal bit-depth with block spatial TV, we propose to model the reconstruction block PSNR under fixed block bit-budget $\mathfrak{B}$ as a quadratic function in a small neighborhood around the optimal bit-depth in the following form

$$\mathrm{PSNR}(B) = p_1(A)B^2 + p_2(A)B + p_3(A), \tag{8}$$

with model parameters $p_i(A)$, $i = 1, 2, 3$ as functions of block spatial TV A. By setting the first derivative of (8) to zero, the optimal bit-depth B^* can then be obtained as

$$B^* = \mathrm{round}\Big(-\frac{p_2(A)}{2p_1(A)}\Big). \tag{9}$$

†The 2D-TV $\mathrm{TV}_{\mathrm{2D}}(\mathbf{X}_m)$ is computed in the same way as specified in [11].

Since the shape of the quadratic model in (8) varies for blocks with different spatial TV, the model parameters $\mathbf{p}(A) = [p_1(A) \;\; p_2(A) \;\; p_3(A)]^{\mathrm{T}}$ corresponding to a certain block with spatial TV A can be obtained via least-squares regression in the following form

$$\mathbf{p}(A) = \arg\min_{\mathbf{p}} ||\mathbf{B}\mathbf{p} - \mathbf{q}(A)||_2^2, \tag{10}$$

where $\mathbf{B}$ is the matrix

$$\mathbf{B} = \begin{bmatrix} B_1^2 & B_1 & 1 \\ B_2^2 & B_2 & 1 \\ \vdots & \vdots & \vdots \\ B_N^2 & B_N & 1 \end{bmatrix}$$

with B_k, $k = 1, 2, ..., N$ be the N candidate bit-depths, and $\mathbf{q}(A) \in \mathbb{R}^{N\times 1}$ are the reconstruction PSNR values corresponding to the same training block with spatial TV A encoded with N different bit-depths and the fixed block bit-budget $\mathfrak{B}$.

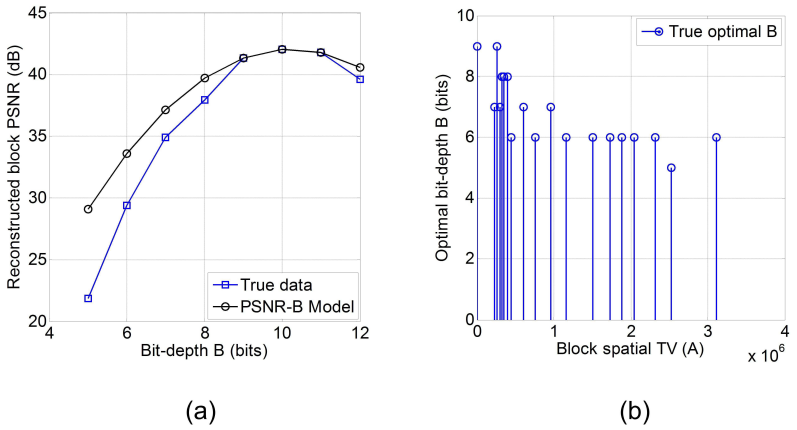

Figure 1. (a) An illustration of PSNR-B model training for one block in the *Container* sequence with block bit-budget $\mathfrak{B} = 4096$ bits. (b) The true optimal bit-depth for *Stefan* sequence with block bit-budget $\mathfrak{B} = 2048$ bits and varying block spatial TV.

A natural next step is to develop sparsity-aware models for parameters $p_1(A)$ and $p_2(A)$ needed for computing the optimal bit-depth from (9). For simplicity, $p_i(A)$, $i = 1, 2$ can each be modeled again as an individual quadratic function of A in the following form

$$p_i(A) = c_{i,1}A^2 + c_{i,2}A + c_{i,3}, \qquad i = 1, 2. \tag{11}$$

where the model parameters $\mathbf{c}_i = [c_{i,1} \quad c_{i,2} \quad c_{i,3}]^{\mathrm{T}}$, $i = 1, 2$ can be obtained again using least-squares regression in the following form

$$\mathbf{c}_i = \arg\min_{\mathbf{c}} ||\mathbf{A}\mathbf{c} - \mathbf{p}_i||_2^2, \quad i = 1, 2, \tag{12}$$

where $\mathbf{A}$ is the matrix

$$\mathbf{A} = \begin{bmatrix} A_1^2 & A_1 & 1 \\ A_2^2 & A_2 & 1 \\ \vdots & \vdots & \vdots \\ A_M^2 & A_M & 1 \end{bmatrix},$$

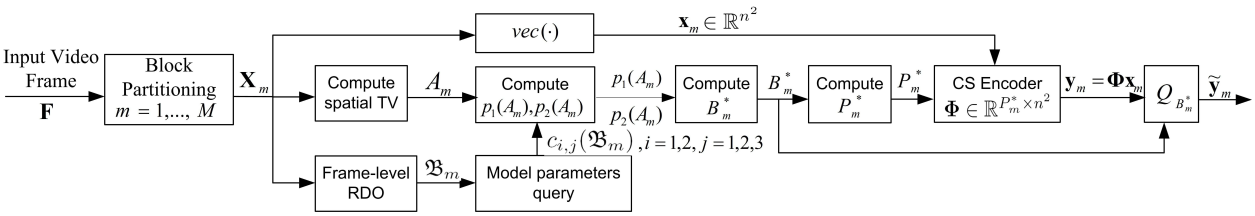

Figure 2. CS video encoder with double-level RDO.

with A_k, $k = 1, 2, ..., M$ be the M spatial TV values of the M training blocks, and $\mathbf{p}_i = [p_i(1), \quad \ldots, \quad p_i(M)]^{\mathrm{T}}$, $i = 1, 2$ are the model parameters obtained from (10) with M training blocks. After model training, the resultant six parameters $c_{i,1}$, $c_{i,2}$, and $c_{i,3}$, $i = 1, 2$ associated with block bit-budget $\mathfrak{B}$ are stored in the encoder memory as $c_{i,j}(\mathfrak{B})$, $i = 1, 2$, $j = 1, 2, 3$ for adaptive bit-depth quantization in the model testing stage.

The block diagram of the CVS system with the proposed double-level RDO is shown in Fig. 2. At the encoder, the video frame is first partitioned into non-overlapping blocks of size $n \times n$, then frame-level RDO is performed to obtain the optimal block bit-budget $\mathfrak{B}_m$ for the m^{th} block $\mathbf{X}_m$, $m = 1, ..., M$. To perform block-level RDO for $\mathbf{X}_m$, its spatial TV A_m is computed and parameters $p_1(A_m)$ and $p_2(A_m)$ are obtained using (11) with $c_{i,j}(\mathfrak{B}_m)$, $i = 1, 2$, $j = 1, 2, 3$. The optimal bit-depth B_m^* is then determined by (9), and corresponding optimal number of CS samples P_m^* is computed as $P_m^* = \frac{\mathfrak{B}_m}{B_m^*}$. For CS acquisition, $\mathbf{X}_m$ is first vectorized into a length n^2 signal via $\mathbf{x}_m = \mathrm{vec}(\mathbf{X}_m)$, which is then projected onto the sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{P_m^* \times n^2}$ generated with mean zero, variance $\frac{1}{P_m^*}$ Gaussian distributed entries. Afterwards, each element of the measurement vector $\mathbf{y}_m$ is quantized with B_m^*-bit uniform scalar quantization. Finally, the quantized indices $\widetilde{\mathbf{y}}_m$ are encoded and transmitted.

4. EXPERIMENTAL RESULTS

In this section, we study experimentally the performance of the proposed double-level RDO for CVS systems by evaluating the PSNR of the reconstructed video sequences. Two test sequences, *Container* and *Stefan* with CIF resolution 352×288 pixels are used. Processing is carried out only on the luminance component.

At the trivial CS encoder side, each frame is partitioned into non-overlapping blocks of 32×32 pixels. The bit-rate per frame is fixed at R equals to 1 to 5 bits per pixel (bpp). For each bit-rate, all blocks in one frame are first assigned the optimal block bit-budget through frame-level RDO, afterwards block-level RDO is performed on each block. For the model training of block-level RDO, 99 blocks in the first frame of each video sequence are encoded with candidate bit-depth values varying from 3 to 13. At the model testing stage, the sparsity level of each block is measured in spatial TV A, after which the optimal bit-depth B^* and number of CS samples P^* are determined through the proposed doubly-model with the block bit-budget $\mathfrak{B}$ assigned in frame-level RDO. Then, each block is viewed as a vectorized column of length $L = 1024$ and multiplied by a $P^* \times L$ measurement matrix with elements drawn from i.i.d. zero-mean, $\frac{1}{P^*}$-variance Gaussian random variables. The elements of the captured P^*-dimensional measurement vector are quantized individually by a B^*-bit uniform scalar quantizer and then transmitted to the decoder. At the decoder side, we choose the log-barrier algorithm [15] to solve the reconstruction problem in (6). For simplicity, the 2-D DCT basis is used for sparse representation.

The proposed CVS system with double-level RDO is compared with existing CVS systems with fixed eight-bit uniform scalar quantization [14]. Fig. 3 (a) shows the rate-distortion characteristics for the *Container* video sequence. The PSNR values (in dB) are averaged over the first 50 frames. Evidently, using the proposed adaptive bit-depth quantization, the performance is better than the fixed bit-depth quantization even when the bit-budget $\mathfrak{B}$ for all blocks are the same, with the maximal gain close to 1dB at low bit-rate. The performance is further improved by 0.7dB at the low bit-rate to 1.8dB at the high bit-rate when frame-level RDO is performed as well to optimize the block bit-budget. The same rate-distortion performance study is repeated in Fig. 3 (b) for

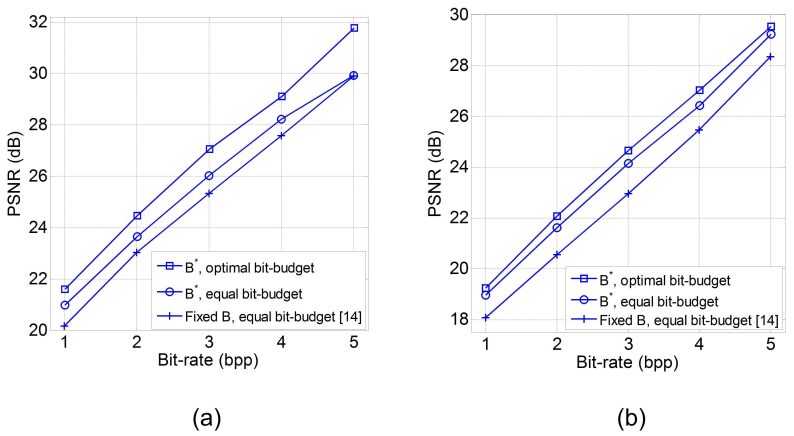

Figure 3. Rate-distortion performance of (a) the *Container* sequence and (b) the *Stefan* sequence.

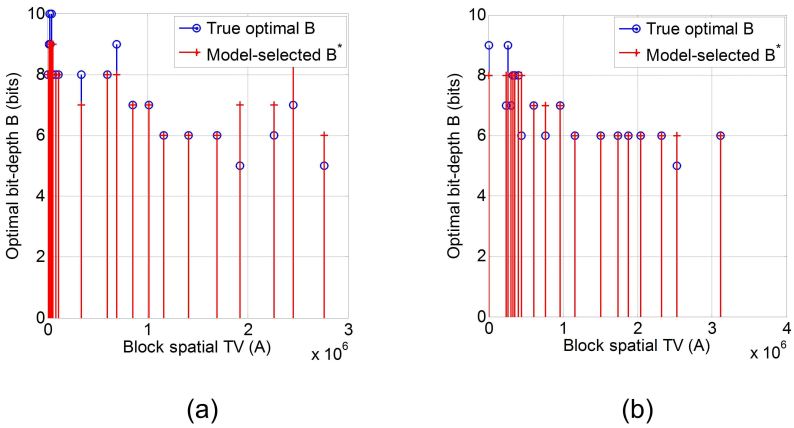

Figure 4. Model-selected optimal bit-depth B^* versus the true optimal bit-depth for (a) *Container* sequence and (b) *Stefan* sequence with block bit-budget $\mathfrak{B} = 2048$ bits and varying block spatial TV.

the *Stefan* sequence, where the proposed adaptive bit-depth quantization algorithm with-or-without frame-level RDO again outperforms the fixed bit-depth quantization scheme with the maximal gain close to 2dB at median bit-rate. Finally, Fig. 4 compares the optimal bit-depth selected by the proposed model with the true optimal bit-depth. It is observed that the model selected bit-depth B^* is the same as or close to the true optimal bit-depth most of the time for both sequences.

5. CONCLUSIONS

We proposed a double-level RDO algorithm for a frame bit-budget constrained CVS system. In the frame-level RDO, each frame block is assigned an optimal bit-budget according to its block sparsity, then the optimal number of CS samples and quantization bit-depth are determined via block-level RDO. In block-level RDO, the reconstruction block PSNR is modeled as a quadratic function of the bit-depth, where the model parameters are approximated again as quadratic functions of the block-sparsity measured in block spatial TV. Experimental results demonstrate that the proposed double-level RDO outperforms the conventional fixed rate compressed sensing with fixed bit-depth quantization for the bit-budget constrained CVS system, as well as adaptive bit-depth quantization without frame-level RDO.

REFERENCES

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

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

[3] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Magazine*, 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 Proc.*, vol. 59, pp. 4759-4766, Oct. 2011.

[5] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure and Applied Math.*, vol. 59, pp. 1207-1223, Aug. 2006.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, 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, pp. 4655-4666, Dec. 2007.

[9] S. Pudlewski, T. Melodia, and A. Prasanna, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comp.*, vol. 11, pp. 1060-1072, June 2011.

[10] 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.

[11] Y. Liu and D. A. Pados, "Rate-adaptive compressive video acquisition with sliding-window total-variation-minimization reconstruction," in *Proc. SPIE, Compressive Sensing Conf., SPIE Defense, Security, and Sensing*, Baltimore, MD, vol. 8717, May, 2013.

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

[13] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. Picture Coding Symposium (PCS)*, Chicago, IL, May 2009.

[14] Y. Liu, M. Li, and D. A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, pp. 438-444, May 2012.

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