ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/224256211

Motion compensation as sparsity-aware decoding in compressive video streaming

Conference Paper · August 2011

DOI: 10.1109/ICDSP.2011.6005006 · Source: IEEE Xplore

CITATIONS
3

READS
25

4 authors:

Ying Liu
University at Buffalo, The State University o…
14 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Ming Li
Dalian University of Technology
37 PUBLICATIONS **110** CITATIONS
SEE PROFILE

Kanke Gao
University at Buffalo, The State University o…
14 PUBLICATIONS **86** CITATIONS
SEE PROFILE

Dimitris A. Pados
University at Buffalo, The State University o…
197 PUBLICATIONS **2,002** CITATIONS
SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Dimitris A. Pados
Retrieved on: 17 June 2016

MOTION COMPENSATION AS SPARSITY-AWARE DECODING IN COMPRESSIVE VIDEO STREAMING

Ying Liu, Ming Li, Kanke Gao, and Dimitris A. Pados[†]
Department of Electrical Engineering
State University of New York at Buffalo
Buffalo, NY 14260 USA
E-mail: {yl72,mingli,kgao,pados}@buffalo.edu

ABSTRACT

We consider a video transmission system where the transmitter performs merely direct compressive sensing with no other forms of encoding/processing and the burden of quality video sequence reconstruction falls solely on the receiver side. We show that effective implicit motion compensation can be carried out at the receiver/decoder via iterative sparsity-aware recovery on adaptively forward-backward estimated Karhunen-Loève bases. Experiments illustrate these developments.

Index Terms— Compressive sampling, compressed sensing, motion compensation, sparse representation, video codecs, video streaming.

1. INTRODUCTION

Traditional sampling schemes follow the general Nyquist/ Shannon's sampling theory: To reconstruct a signal without error, the sampling rate must be at least twice the highest frequency of the signal. Compressive sampling (CS), also known 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 attention is the single-pixel camera architecture [9] where a still image can be produced from significantly fewer captured measurements than the number of desired/reconstructed image pixels. A natural highly desirable next-step development is, arguably, compressive video streaming. In this present work, we consider a video transmission system where the transmitter/encoder performs nothing more than compressed sensing acquisition. The burden for quality video reconstruction falls solely on the receiver/decoder. Such a set-up may be of particular interest in problems that involve large wireless multimedia sensor networks.

[†]Corresponding author.

The quality of video reconstruction 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, well reconstructed CS-based video streaming system comes from developing effective sparse representations and corresponding video recovery algorithms. Several methods for CS video recovery have been proposed, each relying on a different sparse representation. An intuitive (JPEG-motivated) approach is to independently recover each frame using the 2-dimensional discrete cosine transform (2D-DCT) [10] or a 2-dimensional discrete wavelet transform (2D-DWT). To enhance sparsity by exploiting correlations among successive frames, several frames can be jointly recovered under a 3D-DWT [11] or a 2D-DWT applied on inter-frame difference data [12].

In standard video compression technology around us, effective encoder-based motion compensation (MC) is a defining matter in the success and feasibility of digital video. For the problem of CS video capture and recovery, MC matters can be exploited at the receiver/decoder only. In current approaches [13],[14], a video sequence is divided into key frames and CS frames. While each key frame is reconstructed individually using a fixed basis (e.g., 2D-DWT or 2D-DCT), each CS frame is reconstructed conditionally using an adaptively generated basis from adjacent reconstructed key frames.

To exploit inter-frame similarities and pursue most efficient utilization of all available measurements, in this work we propose a new sparsity-aware video decoding algorithm for compressive video streaming systems. For each pair of successive frames, we alternate recovering one frame using Karhunen-Loève transform (KLT) bases adaptively generated/estimated from the other. This scheme essentially implements motion compensation at the decoder by sparsity-aware reconstruction using iterative forward-backward inter-frame (KL) basis estimation.

The rest of the paper is organized as follows. In Section 2, we briefly review the CS principles that motivate our compressive video streaming system. In Section 3, the proposed iterative forward-backward sparsity-aware video decoding al-

978-1-4577-0274-7/11/$26.00 ©2011 IEEE

gorithm is described in detail. Some experimental results are presented and analyzed in Section 4 and, finally, a few conclusions are drawn in Section 5.

2. COMPRESSIVE SAMPLING

In this section we briefly review the CS principles for signal acquisition and recovery. A signal vector $\mathbf{x} \in \mathbb{R}^N$ can be expanded in 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 non-zero components, we call $\mathbf{x}$ a k-sparse signal with respect to $\mathbf{\Psi}$. Many natural signals can be represented as a sparse signal in an appropriate basis.

Conventional approaches to sampling signals follow Nyquist/Shannon's theorem: 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) [3], then the sparse coefficient vector $\mathbf{s}$ can be accurately (with very high probability) recovered via the following linear program

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\widetilde{\mathbf{s}}\|_{\ell_1} \qquad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}. \tag{2}$$

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

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

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

$$\mathbf{y} = \mathbf{\Phi\Psi s} + \mathbf{e} \tag{4}$$

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 constraints in the form of

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\widetilde{\mathbf{s}}\|_{\ell_1} \qquad \text{subject to} \quad \|\mathbf{y} - \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}\|_{\ell_2} \le \epsilon. \tag{5}$$

Specifically, if the isometry constant δ_{2k} associated with RIP satisfies $\delta_{2k} < \sqrt{2} - 1$ [3], then recovery by (5) guarantees

$$\|\widehat{\mathbf{s}} - \mathbf{s}\|_{\ell_2} \le c_0\|\mathbf{s} - \mathbf{s}_k\|_{\ell_1}/\sqrt{k} + c_1\epsilon \tag{6}$$

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.

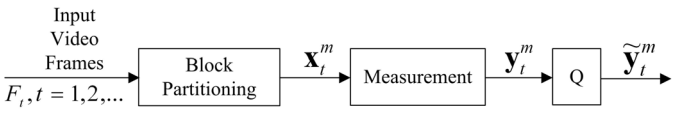

Fig. 1. A simple compressive video encoder.

Equivalently, the optimization problem in (5) can be reformulated as the following unconstrained problem

$$\widehat{\mathbf{s}} = \arg\min_{\widetilde{\mathbf{s}}} \|\mathbf{y} - \mathbf{\Phi\Psi}\widetilde{\mathbf{s}}\|_{\ell_2}^2/2 + \lambda\|\widetilde{\mathbf{s}}\|_{\ell_1}, \tag{7}$$

where λ is a regularization parameter that tunes the sparsity level. The problem in (7) is a convex quadratic minimization program that can be efficiently solved. Again, after we obtain $\widehat{\mathbf{s}}$, $\mathbf{x}$ can be reconstructed by (3). 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 identical 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.

3. PROPOSED CS VIDEO CODING SYSTEM

The CS-based signal acquisition technique reviewed in Section 2 can be applied to video acquisition on a frame-by-frame, block-by-block basis. In the simple compressive video encoding system shown in Figure 1, each frame F_t, $t = 1, 2, ...$, is virtually partitioned into M non-overlapping blocks of pixels with each block viewed as a vectorized column of length N, $\mathbf{x}_t^m$, $m = 1, ..., M$, $t = 1, 2, ...$ Compressive sampling of $\mathbf{x}_t^m$ is performed by random projection in the form of

$$\mathbf{y}_t^m = \mathbf{\Phi x}_t^m \tag{8}$$

with a Gaussian generated measurement matrix $\mathbf{\Phi}_{P\times N}$. 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 [10], each frame is individually decoded via sparse signal recovery algorithms with fixed bases such as block-based 2D-DCT (or frame-based 2D-DWT). With a received (dequantized) measurement vector $\widehat{\mathbf{y}}$ and a block-based 2D-DCT basis $\mathbf{\Psi}_{\rm DCT}$, video reconstruction becomes an optimization problem as in (7),

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

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

$$\widehat{\mathbf{x}} = \mathbf{\Psi}_{\rm DCT}\widehat{\mathbf{s}}. \tag{10}$$

However, such intra-frame 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 [11] the correlation among successive frames was exploited by jointly recovering several frames with a 3D-DWT basis, assuming that the video signal is more sparsely represented in a 3D-DWT domain. In [12], a sparser representation is provided by exploiting small inter-frame differences within a spatial 2D-DWT basis. Nevertheless, in all cases, these decoders cannot pursue/capture local motion effects which can significantly increase sparseness and are well-known to be a critical attribute to the effectiveness of conventional video compression. In this paper, we propose a motion-capturing sparse decoding approach.

The proposed CS video decoder shown in Figure 2 consists of a lower branch that decodes odd frames F_t, $t = 1, 3, 5, ...$, and the upper branch that decodes even frames F_{t+1}, $t = 1, 3, 5, ...$ At initial stage, every odd frame F_t is reconstructed using the block-based fixed DCT basis as shown in (9) and (10). Then, we attempt to reconstruct each block of the even frames F_{t+1} based on the reconstructed previous odd frames $\widehat{F}_t$. Our sparsity-aware MC decoding approach is based on the fact that the pixels of a block in a video frame can be accurately predicted by using a linear combination of a small number of nearby blocks in adjacent (previous and next) frames. As a result, the blocks in F_{t+1} can be sparsely represented by a few neighboring blocks in $\widehat{F}_t$. We propose to use the KLT basis for this representation. For each block $\mathbf{x}_{t+1}^m$, $m = 1, ...M$, in the even frame F_{t+1}, a group of neighboring blocks that lie in a window of a square $w \times w$ region centered at $\mathbf{x}_{t+1}^m$ are extracted from the odd frame $\widehat{F}_t$. Then, the KLT basis for $\mathbf{x}_{t+1}^m$, $\mathbf{\Psi}_{t+1,\mathrm{KLT}}^m$, is formed *by the eigenvectors of the correlation matrix of the extracted blocks from* $\widehat{F}_t$. The sparse coefficients $\mathbf{s}_{t+1}^m$ are recovered by solving

$$\widehat{\mathbf{s}}_{t+1}^m = \arg\min_{\widetilde{\mathbf{s}}} \|\widehat{\mathbf{y}}_{t+1}^m - \mathbf{\Phi}\mathbf{\Psi}_{t+1,\mathrm{KLT}}^m \widetilde{\mathbf{s}}\|_{\ell_2}^2/2 + \lambda\|\widetilde{\mathbf{s}}\|_{\ell_1} \quad (11)$$

and the video block $\mathbf{x}_{t+1}^m$ is reconstructed by

$$\widehat{\mathbf{x}}_{t+1}^m = \mathbf{\Psi}_{t+1,\mathrm{KLT}}^m \widehat{\mathbf{s}}_{t+1}^m. \quad (12)$$

After all M blocks are reconstructed, they are regrouped to form the complete decoded even frame $\widehat{F}_{t+1}$. So far, we call this decoder *forward-only sparsity-aware MC decoding* since only *forward* (*even* in our language) frames are reconstructed accounting for motion.

To further exploit the inter-frame similarities and achieve more efficient utilization of the available measurements, we extend the forward-only MC sparse decoding idea and propose a sequential forward-backward sparse decoding algorithm which performs implicit *iterative* motion estimation and compensation between each pair of (odd, even) frames, in both directions. In particular, in a straight forward manner, after the forward step in which frame F_{t+1} is reconstructed as described above based on $\widehat{F}_t$, we repeat the algorithm

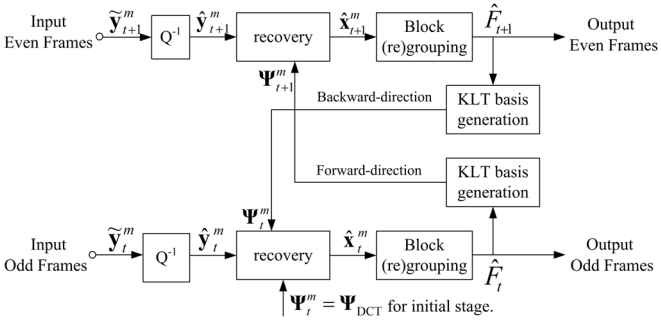

Fig. 2. The block diagram of the decoder: Motion compensation in the form of iterative forward-backward KLT sparse recovery.

backwards, and reconstruct again $\widehat{F}_t$ using KLT bases generated from $\widehat{F}_{t+1}$. For each pair of successive odd and even video frames, this forward-backward approach is repeated until no significant reconstruction quality improvement can be achieved.

A defining characteristic of the proposed CS video decoder in comparison with existing CS video literature [10]-[17] is that the repeated iterative forward-backward decoding algorithm reuses the spatial correlation within a video frame and the temporal correlation between successive video frames, which essentially results to implicit joint spatial motion-compensated video decoding. The adaptively generated block-based KLT basis is seen to provide a much sparser representation basis than fixed block-based bases approaches [10]-[12],[15].

4. EXPERIMENTAL RESULTS

In this section, we study experimentally the performance of the proposed compressive video decoder by evaluating the peak-signal-to-noise ratio (PSNR) and the perceptual quality of reconstructed video sequences. Two test sequences, Foreman and Highway, with CIF resolution 352×288 pixels and frame rate of 30 frames/second are used. Processing is carried out only on the luminance component.

At the encoder side, each frame is partitioned into non-overlapping blocks of 32×32 pixels. Each block is compressively sampled using a $P \times N$ measurement matrix with elements drawn from i.i.d. Gaussian random variables. The captured measurements are quantized by an 8-bit uniform scalar quantizer and then sent to the decoder.

At the decoder side, we choose the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm [6],[7] for sparse recovery (low complexity and satisfactory recovery performance). In our experimental studies, three CS video decoders are examined: (*i*) fixed DCT basis infra-frame decoder, as a reference; (*ii*) forward-only sparsity-aware MC decoder as described by (11)-(12); and (*iii*) iterative forward-

(a) (b) (c)

Fig. 3. Different decodings of the 23rd frame of Foreman: (a) original, (b) using the DCT basis intra-frame decoder ($P = 0.625N$), (c) using the forward-backward sparsity-aware MC decoder ($P = 0.625N$).

(a) (b) (c)

Fig. 4. Different decodings of the 51st frame of Highway: (a) original, (b) using the DCT basis intra-frame decoder ($P = 0.625N$), (c) using the forward-backward sparsity-aware MC decoder ($P = 0.625N$).

backward sparsity-aware MC decoding. In both the forward-only and iterative forward-backward sparsity-aware MC decoders, a square search window with a width of $w = 65$ pixels is used to adaptively generate KLT bases.

Figure 3 shows two different decodings of the 23rd frame of Foreman reconstructed by the DCT basis intra-frame decoder (Fig. 3(b)) and the iterative forward-backward sparsity-aware MC decoder (Fig. 3(c)). It can be observed that the DCT basis intra-frame decoder suffers much noticeable performance loss over the whole image, while our proposed iterative forward-backward sparsity-aware MC decoder demonstrates considerable reconstruction quality improvement. Figure 4 replicates this experiment on the Highway sequence and shows Frame 51 with similar conclusions[1].

Figure 5 and Figure 6 illustrate the rate-distortion characteristics of the three (fixed DCT intra-frame, forward-only, and iterative forward-backward MC) decoders for the Foreman and Highway video sequences, respectively. The PSNR values (in dB) are averages over 100 frames. Apparently,

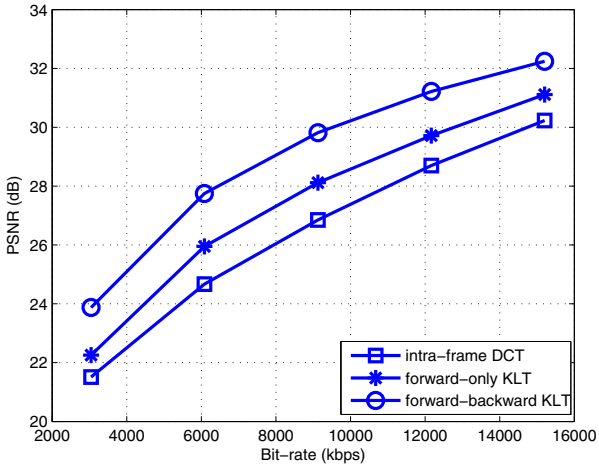

Fig. 5. Rate-distortion studies on the Foreman sequence.

the proposed iterative forward-backward sparsity-aware MC decoder outperforms significantly the forward-only sparsity-aware MC decoder (and, of course, the fixed basis intra-frame decoder), especially at medium bit rate ranges with gains approaching 2dB.

[1] As usual, pdf formatting of the present article tends to dampen perceptual quality differences between Figs. 3 (a), (b) and (c) or Figs. 4 (a), (b) and (c) that are in fact much pronounced in video playback. Figs. 5 and 6 are the usual attempt to capture average differences quantitatively.

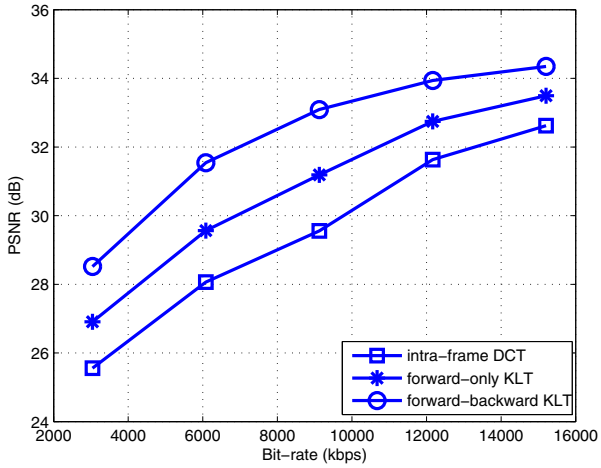

Fig. 6. Rate-distortion studies on the Highway sequence.

5. CONCLUSIONS

We proposed a sparsity-aware motion-accounting decoder for video streaming systems with compressive sampling encoding. The decoder performs an iterative inter-frame forward-backward decoding procedure that adaptively generates KLT bases to enhance the sparse representation of each video frame block, such that the overall reconstruction quality is improved at any given fixed compressive sampling rate. Experimental results demonstrate that the proposed iterative forward-backward sparsity-aware decoder outperforms significantly the conventional fixed basis intra-frame CS decoder as well as non-iterative one-direction sparsity-aware decoding.

6. 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.*, submitted.

[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] 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 Proc. Magazine*, vol. 25, pp. 83-91, Mar. 2008.

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

[11] 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.

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

[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.

[14] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. Picture Coding Symposium (PCS)*, Chicago, IL, May 2009.

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

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

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