DISTRIBUTED VIDEO CODING BASED ON COMPRESSED SENSING

Yousuf Baig, Edmund M-K. Lai† and Amal Punchihewa‡*

School of Engineering and Advanced Technology
Massey University, New Zealand
Email:*m.y.baig,†e.lai,‡g.a.punchihewa@massey.ac.nz

ABSTRACT

Compressed Sensing (CS) is a new approach to signal acquisition that can potentially allow us to design very simple video encoders that can be implemented on mobile devices with limited resources. However, previously proposed CS based video codec either require a conventional video codec or a feedback channel for effective operation, thus increasing the complexity. In this paper, a distributed Compressed Video Sensing codec is proposed that only makes use of CS at the encoder. A novel Side Information generation scheme is incorporated in the decoder which exploits the correlation between CS measurements of nearby frames. It is much simpler than other schemes found in the literature and yet effective. Simulation results demonstrate that effective video coding can be achieved using this codec.

***Keywords*-Compressed sensing, distributed compressed video sensing, distributed video coding, low-complexity video coding, side information**

I. INTRODUCTION

Video coding standards such as MPEG [1] and H.26x [2] are well developed and widely deployed. The exploitation of spatial and temporal redundancies for data compression at the encoder causes the encoding process to be typically 5 to 10 times computationally more complex than the decoder [3]. This often means that the camera needs to have a dedicated chip to perform real time encoding of the captured video. However, many mobile devices such as mobile phones do not have such dedicated hardware and so the computing of the compressed video consumes a lot of power and ties up most of the computing power available. An alternative, efficient and low-complexity encoding scheme is needed.

In the past few years, a new theory called Compressed Sensing (CS) [4]–[6] has been developed which provides us with a completely new approach to data acquisition. In essence, CS tells us that for signals which possess some "sparsity" properties, the sampling rate required to reconstruct these signals with good fidelity can be much lower than the lower bound specified by Shannon's sampling theorem. Since video signals contain substantial amount of redundancy, they are sparse signals and CS can potentially be applied. The simplicity of the encoding process is traded off by a more complex, iterative decoding process. The reconstruction process of CS is usually formulated as an optimization problem which potentially allows one to tailor the objective function and constraints to the specific application. The use of CS can potentially provide the simplicity and efficiency in data acquisition and encoding while shifting the computational burden to the decoder. However, CS alone will not give us a low enough compression rate. This can be remedied by combining CS with distributed Video Coding (DVC) techniques [7]. DVC allows us to encode several pieces of data independently while the decoding is performed jointly. Thus it removes the need for motion estimation and prediction which is computationally the most complex part of the conventional video encoder.

In this paper, we propose a new distributed video coding system based only on CS. Side information (SI) is generated based on the correlation of CS measurements in video frames. This SI generation scheme is simple yet effective, without putting extra complexity at the decoder. Simulation results show that significant performance gains are possible using the proposed video codec.

The rest of this paper is organized as follows. Section II is a brief overview of CS, followed by a brief review of DVC and in particular previously proposed DVC systems which makes use of CS in Section III. Our proposed distributed video coding solution is presented in Section IV. It is tested using several typical video sequences and the results are presented in Section V. Finally, Section VI concludes the paper.

II. COMPRESSED SENSING

Compressed Sensing [4], [5] is applicable to signals that are sparse or compressible in some domain. This applies to most natural signals including video. Let $x \in R^N$ be a discrete time signal. If x can be represented in a transform domain Ψ by s, then

$$x = \Psi s = \sum_{i=1}^{N} s_i \psi_i \tag{1}$$

where $s_i =< x, \psi >$. When all but $K \ll N$ coefficients s_i are zero, then x is said to be K-sparse. In practice, most compressible signals have only a few significant coefficients

while the rest have relatively small magnitudes which can be assumed to be zero.

Let y be the length-M ($M < N$) measurement vector obtained by applying a certain measurement (sensing) matrix Φ to x such that

$$y = \Phi x \tag{2}$$

It has been shown that x can be recovered from $M \sim K$ or more measurements [4], [5]. In order to achieve that, it is necessary for $A = \Phi\Psi$ to have the restricted isometry property [5]. A class of sensing matrices known as Structurally Random Matrix (SRM) [8] has recently been proposed that has performance similar to completely random matrices but is much more computationally efficient.

The reconstruction problem can be expressed as a linear program:

$$\min \|x\|_{l_1} \text{ subject to } Ax = y \tag{3}$$

Many algorithms that are based on basis pursuit and matching pursuit are available to solve it. One of them is the Gradient Projection for Sparse Reconstruction (GPSR) algorithm [9].

III. DISTRIBUTED COMPRESSED VIDEO SENSING

Distributed Video Coding (DVC) is an application of distributed source coding (DSC). It involves the encoding of two or more dependent random sources where each source is coded by an independent encoder. At the receiver (decoder), the independently encoded data are jointly decoded by exploiting correlation between them. The basic theory for lossless compression was established by Slepian and Wolf [10]. Later, Wyner and Ziv [11] extended it to lossy compression with side information (SI) at the decoder. The idea of DVC is to remove the need for joint encoding and motion estimation/prediction at the encoder. Several video codecs based on this principle have been developed [7], [12], [13]. In this section, we will review the work in DVC that only involves using CS.

In [14], a framework called Distributed Compressed Video Sensing (DISCOS) is introduced. At the encoder, video frames are grouped into group of pictures (GOP) consisting of a key frame and a number of non-key frames. Key frames are encoded using traditional MPEG/H.264 encoding. For non-key frames, both local block-based and global frame-based CS measurements are taken. Side information is generated by using a block-based prediction frame which is created by sparsity-constraint block prediction. In this approach, the block-based measurements of a CS frame are compared with two neighbouring decoded key-frames. The measurement vector of the prediction frame is subtracted from that of the input frame to form a new measurement prediction error vector. The reconstructed CS frame is simply the sum of the prediction error and the prediction frame. The disadvantage of this framework is that the complex MPEG/H.264 encoding is still required.

DVC and CS are combined in [15] to simultaneously capture and compress video data. Their approach is different from [14] in that CS measurement is applied to both key and non-key frames. Key-frames are reconstructed using GPSR [9] at the decoder. For every non-key frame, a stopping criteria based on side information generated from the key-frames is used during the reconstruction process. Side information is generated by an efficient frame rate up-conversion tool. This work is extended in [16], [17] with the concept of dictionary learning. The dictionary is learned from adjacent video frames.

Another distributed approach to CVS is reported in [18]. For each image block in non-key frame, two different coding modes, SKIP and SINGLE, are used. In the SKIP mode, a block is skipped for decoding if it does not change much from the co-located decoded key frame. This is achieved by increasing the complexity at the encoder. In the SINGLE mode, CS measurements for a block are compared with the CS measurements in a dictionary using the MSE criterion. If it is below some threshold, then the block is marked as a decoded block. A feedback channel is used to communicate with the encoder that this block has been decoded and no more measurements are required. For blocks that are not encoded by either the SKIP or the SINGLE mode, normal CS reconstruction is performed. A somewhat similar approach is taken by [19] without using a feedback channel.

For these methods, either a feedback channel is required or a complex side information generation technique is employed. There are many application scenarios where a feedback channel either does not exist or is rather unreliable. Thus there is a need for a CS based video codec which does not require a feedback channel with simple side information generation methods to keep the complexity of the encoder low.

IV. PROPOSED DCVS CODEC

A block diagram of our proposed distributed compressed video sensing (DCVS) codec is shown in Figure 1. This codec is entirely based on CS and does not involve traditional video encoders. Both the key frames and non-key frames are encoded as CS measurements. It also does not require a feedback channel.

IV-A. Encoder

At the encoder, a video sequence is broken up into a sequence of GOPs. Video frames are divided into key frames and non-key frames, which are also called Wyner-Ziv (WZ) frames, within a group of pictures (GOP). Each GOP consists of a key frame followed by some non-key WZ frames. Both key and non-key frames are encoded in a similar way using CS. The data of each frame is converted to a column vector $x \in R^{N\times 1}$ where N is the total number of pixels in the

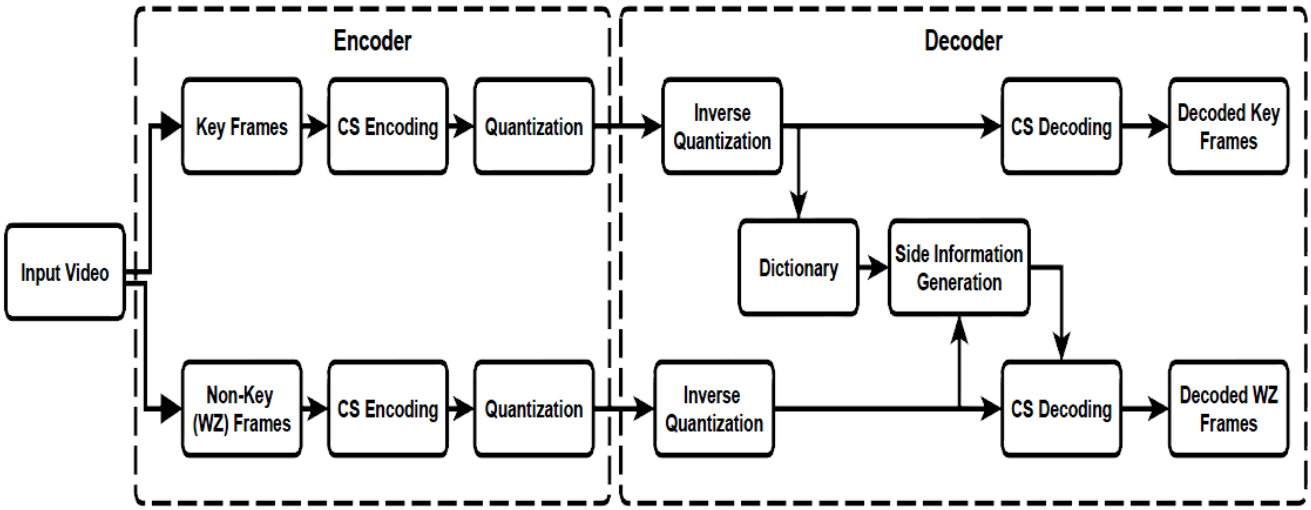


Fig. 1. Proposed Video Codec

frame. CS measurements y are obtained by taking random projections through a measurement matrix Φ, i.e. $y = \Phi x$. The number of measurements for key and non-key frames are denoted by M_k and M_w respectively. The corresponding measurement rates can be defined by M_k/N and M_w/N. Key frames are encoded with a higher measurement rate that WZ frames with $M_w < M_k < N$.

The measurements y are then quantized by a Gaussian quantization scheme proposed in [20]. While previous works ignored the quantization step, it is crucial for practical codecs. Conventionally, different quantization matrices are used for intra-frame and inter-frame coding. For MPEG, the DC and the lower frequency Discrete Cosine Transform (DCT) coefficients are finely quantized while the higher frequency coefficients are coarsely quantized [1]. This design is based on the fact that the human visual system is less sensitive to errors in higher frequencies than it is for lower frequencies. Also, the values of the DCT coefficients tend to be larger at the lower end of the spectrum. For the H.264 baseline, main and extended profiles, the quantization matrix gives equal weight to all coefficients and uses a uniform quantization scheme [2]. The CS measurement process is very different from orthogonal transforms such as the DCT. The distribution of CS coefficients is directly related to the measurement matrix used. The authors in [20] investigated the quantization effects on CS measurements and recovery for video signals. They have shown that both uniform quantization and the standard quantization matrix in MPEG do not perform well for compress-sensed videos. They proposed that Gaussian quantization performs better than uniform and MPEG quantization in CS based videos. In this work, we have adopted this quantization technique.

IV-B. Decoder

At the decoder, each key frame is reconstructed by the GPSR algorithm [9]. It reformulates the l_1-minimization problem given by (3) as

$$\min_{\theta_x} \frac{1}{2}||y_k - A\alpha_k||_2^2 + \tau||\alpha_k||_1 \tag{4}$$

where y_k is the CS measurements of key frame received at the decoder, $A = \Phi\Psi$ as decribed in Section II, $\alpha_k \in R^N \times 1$ is the sparse coefficient vector which is solved by GPSR algorithm. The key frame $\hat{x}_k$ is obtained by $\hat{x_k} = \Psi\hat{\alpha}_k$ where $\hat{\alpha}_k$ is the optimal solution for α_k in (4).

The decoding WZ frames are aided by side information which is generated through a dictionary. Side information is generated from the inverse quantized CS measurements of the key frame. This side information is only useful if the CS measurements of the key and WZ frames exhibit sufficient correlation. Therefore, before discussing dictionary and side information generation, we shall first present a correlation analysis of the CS measurements between video frames.

IV-C. Correlation Analysis

In a video sequence, adjacent frames in same scenes are highly correlated with each other. Therefore we postulate that the CS measurements of such adjacent frames are also highly correlated even though the CS measurement process is very different from linear transforms such as the DCT. DCT coefficients follows the Laplacian distribution [21]. On the other hand, the CS measurements follows a more or less normal (Gaussian) distribution. So the CS measurements can be modelled as random Gaussian sources. The dependence between two random quantities is indicated by Pearson's correlation coefficent [22].

To analyze the correlation among video frames, we used the first 100 frames of four standard video sequences (Foreman, News, Coastguard and Harbour). CS measurements

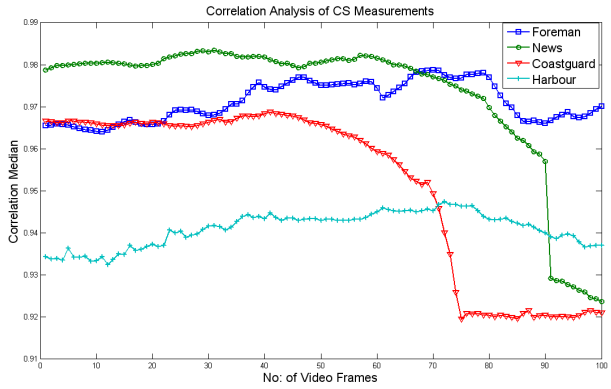


Fig. 2. Correlation Analysis for CS Measurements

of the luminance data are obtained for each frame with a measurement rate of 50%. The correlation coefficient of the measurements of each frame with all other 99 frames are computed. Figure 2 shows the median of correlation of each frame. All video frames of each video sequence show high correlation between CS measurements with median correlation coefficient above 0.9. Figure 3 shows the correlation coefficient between all analyzed frame in the "foreman" video.

Total 50% CS measurement per frame are taken. Figure 2 shows the median values of correlation between all frames. All video frames of each video sequence show high correlation between CS measurements. This is more clear from Figure 3, which shows the visual representation in a mesh plot for correlation in foreman video.

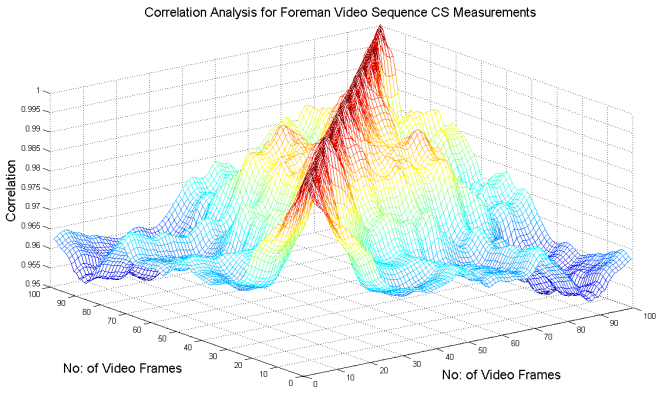


Fig. 3. Correlation of CS Measurements for 100 frames of Foreman video

IV-D. Side Information Generation

Side information (SI) plays an important role in DVC decoding. If SI is not accurate, then RD performance will suffer. We propose a novel Side Information technique based on Correlation of CS measurements. We showed in Section IV-C that CS measurements of adjacent video frames are highly correlated. We create a dictionary D from CS measurements of Key-frames y_k. The columns of the D consists of the CS measurements of key-frames available at the decoder. This is different from [16], where a dictionary is learned from the neighbouring frames of a video frames. For a given Wyner-Ziv frame (non-key frame) x_w, the maximum correlation of its CS measurements y_w with the dictionary D is given by

$$\max_{r(i)}, r(i) = (y_w, D_i)\, i = 1, 2, 3....N \tag{5}$$

In above equation, $r(i)$ is the CS measurement correlation between current WZ frame CS measurements y_w and each atom (column) of dictionary D_i. The column of dictionary D_i which has the maximum correlation $max_{r(i)}$ with the CS measurements y_w will be selected as the side information. Following algorithm is used for reconstructing WZ frames with Side Information.

Algorithm 1 Reconstruction with Side Information

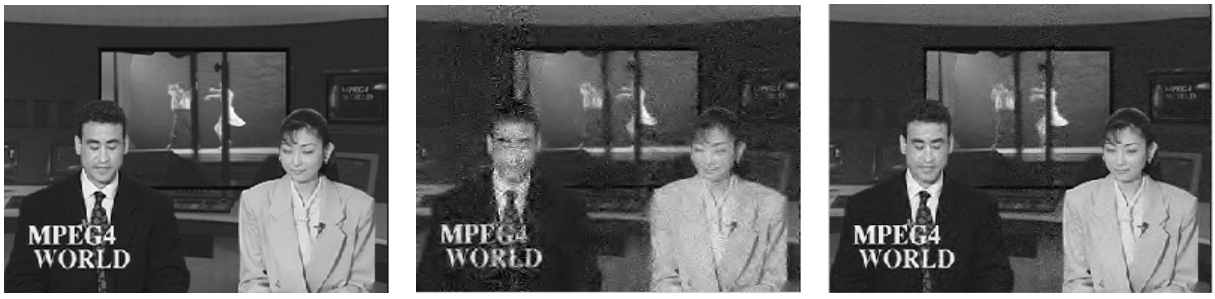
Input:y_w, D
Output: Reconstructed WZ Frame, $\hat{x}_w$
for each column i inD **do**
  Calculate $r[i] = Correlation(y_w, D)$
end for
Calculate $\beta_s = max[r]$
$\beta_w = [y_w, \beta_s]$
$\hat{x}_w = Reconstruction(\beta_w)$

In algorithm 1, β_s is the maximum correlated column in Dictionary D which is then used as the side information. $\beta_w = [y_w, \beta_s]$ is an important step in the algorithm which incorporate the side information β_s with current WZ frame CS coefficients y_w. As β_s and y_w are highly correlated, β_w is the updated measurement rate M_{w+k} for current WZ frame equal to measurement rate of M_k of key frame. β_w is then used with GPSR algorithm for WZ frame reconstruction.

This is very simple, yet efficient SI technique and does not involve complexities like other SI generation techniques. In [14] both block based and frame based CS measurements are combined with decoded key frames to generate the SI. In our proposed technique, it is not necessary to first decode key frames. SI is generated with direct CS measurements of Key frames y_k. In other SI techniques, feedback channel is required for the successful recovery and SI generation [17], [18]. The proposed codec does not require feedback channel. Even for non-feedback architectures [19], motion estimation is required for the generation of SI. The proposed SI does not require any motion estimation.

V. SIMULATION RESULTS

To demonstrate the effectiveness of our proposed distributed compressed sensing video codec, several QCIF



(a) (b) (c)

Fig. 4. Reconstructed 32nd News frame. (a) Original Frame; (b) Reconstruction without SI, PSNR=23.55dB; (c) Reconstruction with SI, PSNR=28.22dB

(frame size: 176x144) video sequences, (full 300 frames for each) are used. Only the luminance component is used in the simulations. CS measurements are obtained using the Hadamard structurally random matrices [8]. The GPSR [9] algorithm incorporating SI is used for CS reconstruction. The GOP size used in simulation is 3, i.e. for every key frame, two non-key (WZ) frames are inserted between them. Different measurement rates (MRs) were used to evaluate the proposed DCVS method. For example, the average MR = 37% means that the MRs for each key and non-key (WZ) frames are 50% and 30%, respectively.

In this paper, a compressive video sensing schemes without side information is used for comparison with our proposed DCVS scheme (denoted by Proposed). In without SI scheme, key frames and non-key frames are reconstructed with respect to the frame-based DWT basis without using any side information. This is similar to the approach used in [8].

Table I shows the average PSNR performance for the test video sequences for an average measurement rate of 37%. The values shown are averaged over all frames and also for WZ frames only, both with and without side information. The overall PSNR performance increases by over 1dB for all videos when SI is used at the decoder. Considering only the WZ frames, the average improvement is over 1.5dB. The highest improvement of 4dB is with the "news" video where the scenes are relatively static. This clearly shows the effectiveness of our proposed Side information generation technique.

Figure 4 shows the visual reconstruction quality for 32nd frame (which is a WZ frame) with and without SI. The proposed DCVS scheme improves reconstruction quality of WZ frame significantly. Figure V, 6 and 7 shows the R-D curve for average PSNR performance for Foreman, News and Coastguard videos. For lower MRs, the proposed DCVS codec gives substantial improvements in PSNR quality. For "Foreman" sequence, at higher MRs, the PSNR improvement is not much significant. For "Coastguard" and "News"sequences, the improvements for both low and high MRs are significant. Overall, the proposed DCVS scheme with SI out performs Frame based DWT without SI.

Table I. Average Reconstructed PSNR (dB)

Video	All frames without SI	All frames with SI	WZ frames without SI	WZ frames with SI
Foreman	26.91	27.89	25.49	26.97
News	25.26	28.10	23.56	27.82
Coastguard	25.12	26.81	23.79	26.33

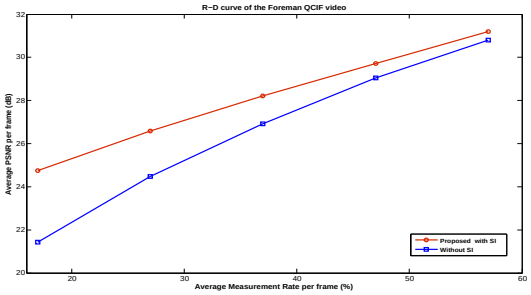


Fig. 5. MR-PSNR Performance for Foreman Video

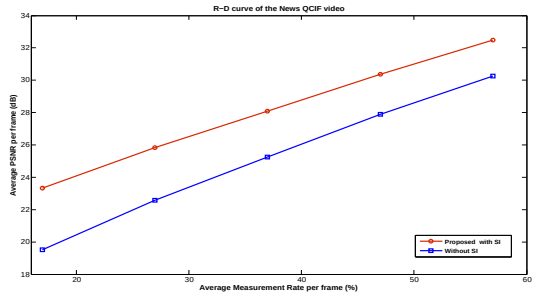


Fig. 6. MR-PSNR Performance for News Video

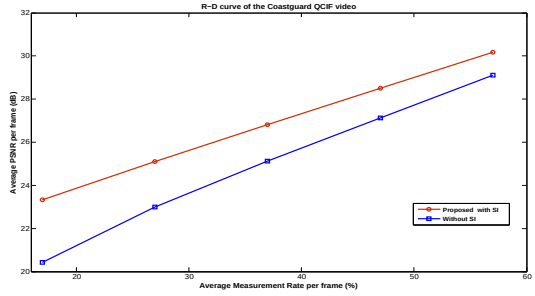


Fig. 7. MR-PSNR Performance for Coastguard Video

VI. CONCLUSIONS

In this paper, we proposed a simple but effective distributed video compressed sensing codec. The encoding is entirely performed using compressed sensing which can be implemented with much reduced hardware complexity compared with conventional video coders. At the decoder, we proposed a simple Side Information generation technique based on our correlation analysis of CS measurements between video frames. This technique does not require a feedback channel nor motion estimation as in some previously proposed methods. It does not even need the decoded key frames. Simulation results show the effectiveness of the proposed video codec.

VII. REFERENCES

[1] P. Symes, *Digital Video Compression*. McGraw-Hill, 2004.

[2] ITU, "Advanced video coding for generic audiovisual services," *ITU-T Recommendations for H.264*, 2005.

[3] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

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

[5] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[7] F. Pereira, "Distributed video coding: Basics, main solutions and trends," in *Proceedings of IEEE International Conference on Multimedia and Expo*, Cancun, Mexico, 28 June-3 July 2009, pp. 1592–1595.

[8] T. Do, L. Gan, and T. Tran, "Fast compressive sampling with structurally random matrices," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, 30 March - 4 April 2008, pp. 3369–3372.

[9] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

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

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

[12] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Proceedings of Picture Coding Symposium*, Lisbon, Portugal, 7-9 November 2007.

[13] J. Ascenso, C. Brites, F. Dufaux, A. Fernando, T. Ebrahimi, F. Pereira, and S. Tubaro, "The VISNET II DVC codec: Architecture, tools and performance," in *Proc. 18th European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, 23-27 Aug. 2010.

[14] T. Do, Y. Chen, D. Nguyen, N. Nguyen, L. Gan, and T. Tran, "Distributed compressed video sensing," in *Proceedings of 43rd Annual Conference on Information Sciences and Systems*, Baltimore, USA, Mar. 2009, pp. 1–2.

[15] L.-W. Kang and C.-S. Lu, "Distributed compressive video sensing," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, 19-24 Apr. 2009, pp. 1169–1172.

[16] H.-W. Chen, L.-W. Kang, and C.-S. Lu, "Dictionay learning-based distributed compressive video sesning," in *Proceedings of Picture Coding Symposium*, Dec. 2010, pp. 210–213.

[17] H. W. Chen, L. W. Kang, and C. S. Lu, "Dynamic measurement rate allocation for distributed compressive video sensing," in *Proceedings of SPIE Visual Communications and Image Processing*, July 2010, pp. 774 401–774 410.

[18] J. Prades-Nebot, M. Yi, and T. Huang, "Distributed video coding using compressive sampling," in *Proceedings of Picture Coding Symposium*, Chicago, IL, USA, 6-8 May 2009.

[19] Z. Gan, L. Qi, and X. Zhu, "Wyner-Ziv coding of video using compressive sensing without feedback channel," in *Proc. IEEE 10th International Conference on Signal Processing*, Oct. 2010, pp. 1129 –1132.

[20] Y. Baig, E. Lai, and J. Lewis, "Quantization effects on compressed sensing video," in *Proceedings of 17th International Telecommunications Conference*, Doha, Qatar, 4-7 April 2010.

[21] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. 31, no. 6, pp. 835 – 839, jun 1983.

[22] S. Hoggar, *Mathematics of Digital Images*. Cambridge, 2006.