LOW COMPLEXITY SIDE INFORMATION FOR DISTRIBUTED COMPRESSED VIDEO CODING

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

Distributed Video coding based on compressed sensing is considered in this paper. Side information plays an important role in the quality of decoded non-key video frames. Existing systems generate side information based on the decoded key frames and the processes are quite complicated, increasing the computation burden at the decoder. We propose a side information generation method that is founded on the high statistical correlation between compressed sensing measurements of key frames and non-key frames. The proposed technique is simple and simulation results show that better rate distortion performance can be obtained in comparison with motion compensated interpolation.

Index Terms— Distributed Video Coding, Side Information Generation, Compressed Sensing.

1. INTRODUCTION

The encoding of video data in conventional video compression standards [1] is a computationally demanding process mainly because it involves motion estimation to give us higher compression rates. Decoding videos compressed with conventional standards, on the other hand, is much simpler. For modern applications where video acquisition is performed by resource limited devices such as mobile phones, and decoding is performed by relatively resource rich computers, a new approach to video encoding and decoding is needed. It basically requires a low-power, low-complexity encoder while the computational burden is shifted from the encoder to the decoder.

Research in this direction has been developed along the lines of Distributed Video Coding (DVC) [2]. DVC is an application of distributed source coding, pioneered by Slepian and Wolf [3] and also Wyner and Ziv [4], which involves the encoding of two or more dependent sources. Each source is coded by an independent encoder. At the receiver (decoder), the independently encoded data are jointly decoded by exploiting any correlation between them. When applied to video coding, this implies that no motion estimation is performed at the encoder. More recently, DVC has been combined with the concept of Compressed Sensing (CS) [5, 6]. The theory of CS tells us that for signals which are sparse in a certain domain, the sampling rate required to reconstruct these signals can be much lower than what is required by Shannon's sampling theorem. Since video signals contain substantial amounts of redundancy and are therefore sparse, CS is applicable. A number of DVC schemes that makes use of CS has recently been proposed [7, 8, 9, 10]. Typically, video frames are classified into key and non-key frames. Key frames are encoded at substantially higher rates than non-key frames. Since these frames are encoded independently in DVC systems, the quality of the decoded non-key frames will be substantially lower than that for key frames. In order to overcome this problem, non-key frames are decoded with the help of side information which is generated at the decoder. The way side information is generated varies in complexity for the various systems mentioned above. A common technique is motion compensated interpolation which involves motion estimation, increasing the computational burden at the decoder. More importantly, side information is derived from decoded key frames and therefore depends on the quality of the reconstructed data.

In this paper, we propose a simple and effective side information generation technique for DVC based solely on CS. This technique is based on the fact that CS measurements between video frames are highly correlated. Side information is generated directly from CS measurements of the key frames. Thus it does not depend on the decoded key frames, unlike other DVC techniques. The performance of this technique is compared with motion compensated interpolation proposed in [8] using four different video sequences. Experimental results show that the proposed technique produces better rate distortion performance and is computationally less demanding.

The rest of this paper is organized as follows. In Section 2 we give a brief overview of compressed sensing and distributed video coding based on CS. In Section 3, we provide some details on role of side information in DVC. Our proposed side information generation scheme is presented in Section 4. It is tested using several video sequences and re-

978-1-4673-2193-8/12/$31.00 ©2012 IEEE

sults are presented in Section 5. Finally, Section 6 concludes the paper.

2. BACKGROUND

2.1. Compressed Sensing

Compressed Sensing [5, 6] suggests that signals which are sparse in some domain can be efficiently acquired much lower than the sampling rate required by Shannon's sampling theorem. In practice, most compressible signals have only a few significant coefficients while the rest have relatively small magnitudes. A signal is more compressible if it has higher sparsity in some representation domain Ψ that is less coherent to the sensing (or sampling) domain Φ. Let $x = \{x[1], \dots x[N]\}$ be a discrete time real-valued random process. If x is represented in a transform domain Ψ by s, then

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

where $s = [s_1 \dots s_N]$, $s_i =< x, \psi >$ and $\Psi = [\psi_1, \psi_2 \dots \psi_N]$ is the basis matrix. Let y be the length-M ($M < N$) measurement vector,obtained by applying a certain measurement matrix Φ to x such that

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

It has been proven that x can be recovered from $M \sim K$ or more measurements [5, 6]. In order to achieve that, it is necessary for $A = \Phi\Psi$ to have a restricted isometry property [11]. The reconstruction problem can be expressed as a linear program:

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

This under-determined linear program can be efficiently solved by algorithms based on basis pursuit [12, 13], matching pursuit [14, 15], and gradient projection [16].

2.2. Distributed Compressed Video Coding

Distributed video coding applies the theory of distributed source coding to video signals where each video frame is encoded independently. At the decoder, the independently encoded data are jointly decoded by exploiting statistical dependencies between them. We are interested in DVC that makes use of CS.

A framework called Distributed Compressed Video Sensing (DISCOS) is proposed in [7]. It is a hybrid video codec that uses traditional MPEG/H.264 encoding for key frames and CS for non-key or Wyner-Ziv (WZ) frames. Side information for a block in a WZ frame is essentially motion vectors that are estimated in the same way as MPEG. 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. A similar distributed CS video codec is reported in [9]. For each block in a non-key frame, two different coding modes, known as 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. In the SINGLE mode, CS measurements for a block are compared with those in a dictionary using the MSE criterion. 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 SKIP or SINGLE mode, normal CS reconstruction is performed.

While the above DVC's make use of CS only for non-key frames, a CS only DVC codec is proposed in [8]. It will be referred to as the distributed compressed video sensing (DCVS) system. Both key frames and non-key frames are encoded using CS. A higher measurement rate is used for key frames compared with non-key frames. Side information for the decoding of non-key frames is generated using a frame rate up conversion tool that make use of motion compensated interpolation which is described in more detail in Section 3.1. This work has been extended to use dictionary learning techniques for selecting the best side information [10]. For these codecs, side information generation involves motion estimation which is computationally demanding. Since the CS reconstruction process is itself computationally complex, this way of generating SI will increase the computational burden at the decoder even further.

3. SIDE INFORMATION GENERATION

In DVC, WZ frames are encoded at much lower rates than key frames. To compensate for this, side information is generated using the key frames at the decoder for the reconstruction of WZ frames. Side information plays an important role in DVC decoding. If SI is not accurate, then the rate-distortion (RD) performance will suffer.

The Laplacian distribution is commonly used to model the correlation noise [2, 17, 18]. It provides a good trade-off between model accuracy and complexity and, therefore, is often chosen [19]. In [8], the statistical dependency between a WZ frame W and its side information SI is modelled as a virtual correlation channel, where SI can be viewed as a noisy version of W. The correlation between W and SI can then be modelled using a Laplacian distribution as follows:

$$p([W(x,y) - SI(x,y)) = \frac{\alpha}{2} e^{-\alpha|W(x,y)-SI(x,y)|} \tag{4}$$

Here, $p(\cdot)$ is the probability density function, $W(x,y)$ and $SI(x,y)$ are the (x,y)-th pixel in W and SI and α is the

Laplacian distribution model parameter defined by

$$\alpha = \sqrt{\frac{2}{\sigma^2}} \tag{5}$$

where σ^2 is the variance of the residue between the W and SI. Therefore, the more similar W and SI are, the higher will be the value of α.

3.1. Motion Compensated Interpolation

Motion-compensated interpolation (MCI) is a general side information method used in [8, 10]. It is similar to motion estimation used for B-frames in MPEG. Let W_n denote a WZ frame at time n, and let K_{n-1} and K_{n+1} be the key frames adjacent to W_n. We need to estimate the motion compensated prediction for W_n. If the motion contained in three successive frames can be assumed to be linear, then the motion vectors for W_n can be derived from the motion vectors from the adjacent two key frames. For forward prediction, if the motion vector of a block b_i in W_n is MV_f, then MV_f can be derived from the motion vector of co-located block in K_{n+1} by $MV_f = MV_{n+1}/2$. Using the same method, we can obtain the backward prediction motion vector by $MV_b = MV_{n-1}/2$. After that, we can compute the two motion predicted blocks of b_i from K_{n-1} and K_{n+1}. Let P_b represent the prediction value of b_i, then $P_b = (P(MV_f)+P(MV_b))/2$ where $P(MV_f)$ and $P(MV_b)$ are the predicted values based on the forward and backward motion vectors respectively. In this way, most blocks of W_n can be predicted and the side information SI can be achieved.

4. PROPOSED SIDE INFORMATION GENERATION

The DVC system that we are considering involves only CS measurements. In other words, both key and WZ frames are encoded by CS measurements only. Therefore the side Information that are generated will be used directly by the CS reconstruction algorithm. This is different from methods used by other DVC codecs discussed above.

First, we need to establish the extent of correlation between CS measurements of adjacent frames. Our side information generation method is based on this knowledge. It should be emphasized here that our method does not depend on any decoded data at the decoder. Hence errors in the reconstructed frames will not affect the quality of the side information generated.

4.1. Correlation Analysis of CS Measurements

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. The dependence between two random

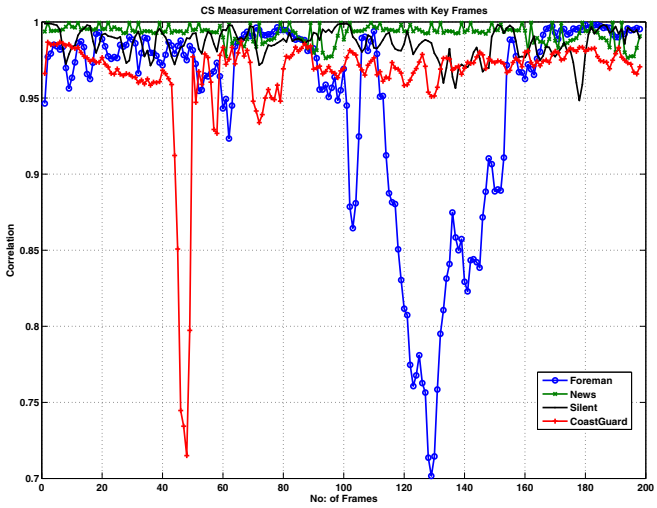

Fig. 1. Correlation Analysis for CS Measurements

quantities can be indicated by the Pearson's correlation coefficient [20]. It can be obtained by dividing the covariance of two variables by the product of their standard deviation.

To analyze the correlation among video frames, we have tested all 300 frames of four different test video sequences – "Foreman","News","Coastguard" and "Silent" available from [21]. For each frame, 50% of random CS measurements are used. The first frame in a sequence of 3 consecutive frames is considered a key frame, followed by two non-key (WZ) frames. Hence there are 200 WZ frames per video sequence. Correlation coefficients of the CS measurements between each WZ frame and key frames are computed and shown Figure 1. All WZ frames show high correlation with their key frames. In particular, the videos "News" and "Silent" which have slow motion show very high correlation throughout. On the other hand, the correlation coefficients for the part of the "Foreman" sequence where there is relatively fast motion are lower. However, they are still above 0.7.

4.2. Correlation Based Side Information

We showed in Section 4.1 that CS measurements of WZ frames are highly correlated with adjacent key frames. Therefore we can directly make use of the CS measurements of key frames as side information. Starting with an empty dictionary D, we populate the first column of D, denoted D_1 with the CS measurements of the first key frame received. Subsequent columns of D are populated with the CS measurements of the corresponding key frames.

Assume that each key frame is followed by two WZ frames. When the first two WZ frames are decoded, D has only one column D_1. So D_1 is used as the side information to reconstruct these two WZ frames. For the third WZ frame W_3, the dictionary will have two columns from the two key frames received so far. One of them will be used as side

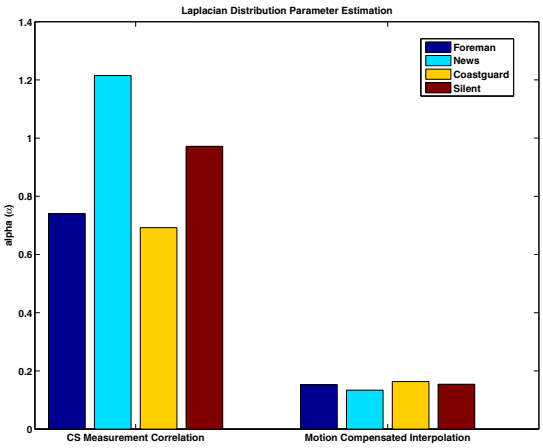

Fig. 2. Laplacian Distribution Parameter Estimation

information for W_3. The best choice will be the one that has higher correlation with W_3. Thus we need to compute the Pearson correlation coefficients

$$r(i) = corr\,(W_3, D_i) \qquad i = 1, 2 \tag{6}$$

and choose the column D_i that gives the largest $r(i)$ for all i in the dictionary. This continues until all the WZ frames are reconstructed.

This method does not require the key frames to be decoded. Furthermore, the side information can be directly used by the CS reconstruction algorithm. In order to limit the size of the dictionary, only the measurements of the most recent N key frames need to be stored as the most recent WZ frame will most likely be more correlated with the most recent key frames.

To further evaluate the effectiveness of our proposed side information technique, we estimated the laplacian distribution parameter α as discussed in section 3 for the proposed correlation based SI and motion compensated SI. We used the same video sequences as in Section 4.1 and generated the side information using MCI and our proposed method. Figure 2 shows the median of Laplacian distribution parameter α for these videos. It can be observed that α is substantially larger for side information generated by the correlation-based method compared with MCI. These results suggest that the proposed correlation-based side information performs much better than MCI based ones.

5. RECONSTRUCTION RESULTS

To evaluate the proposed side information generation scheme, a DVC codec that uses only CS for encoding is implemented in MATLAB. The video sequences used for testing are "Foreman","News","Coastguard" and "Silent" available from [21] which are in QCIF format (174×144 pixels). Both key frames and WZ frames are encoded by CS measurements only. Key frames are encoded with higher measurement rate (MR) compared with WZ frames. A group of picture (GOP) consists of three frames – a key frame followed by 2 WZ frames is used. Only the luminance component is encoded.

The measurements are quantized by the quantization matrix proposed in [22]. Structurally Random Matrices [23] are used to acquire CS measurements at the encoder. The Gradient Projection for Sparse Reconstruction (GPSR) [16] algorithm is used for reconstruction at the decoder. The proposed side information generation scheme is compared with motion compensated interpolation (MCI) [8] and with frame based measurement (Frame DWT) without side information [23].

First, the computational complexities are compared. The complexity of different side information generation schemes are evaluated by calculating the average reconstruction time (in seconds) for key frames and WZ frames using an average measurement rate of 27%. The programs are run on the same computer. The results for the four video sequences are shown in Figure 3. It can be observed that using side information improves the reconstruction time regardless of the side information used. The proposed side information performs better than MCI based side information because it does not require motion estimation. Its performance is also comparable to DCVS [8] although they have used a relative stopping criteria to reduce the number of iterations during reconstruction.

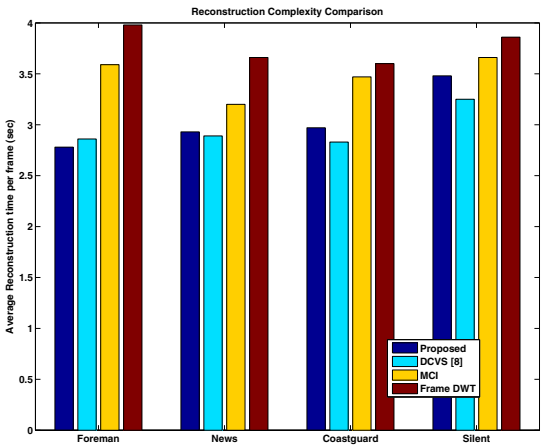

Fig. 3. Reconstruction complexity comparison for average measurement rate of 27%

Next, the rate-distortion performance of the reconstructed videos are compared. Figures 4 to 7 show the average reconstructed PSNR at various average measurement rates for the four video sequences. For video sequences involving slower motion ("News" and "Silent"), our proposed method outperforms MCI based side information and DCVS [8]. For the "Foreman" and "Coastguard" sequences, the proposed

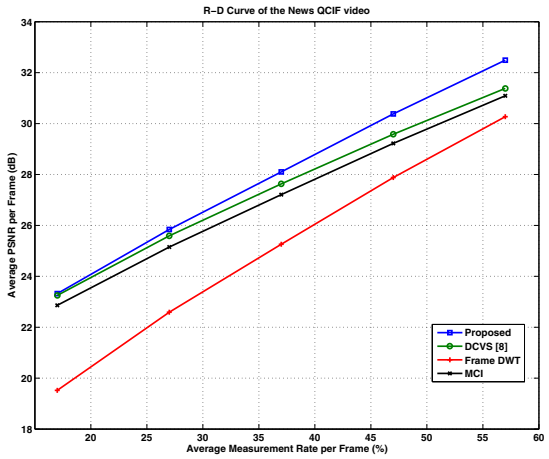

Fig. 4. MR-PSNR Performance for News video

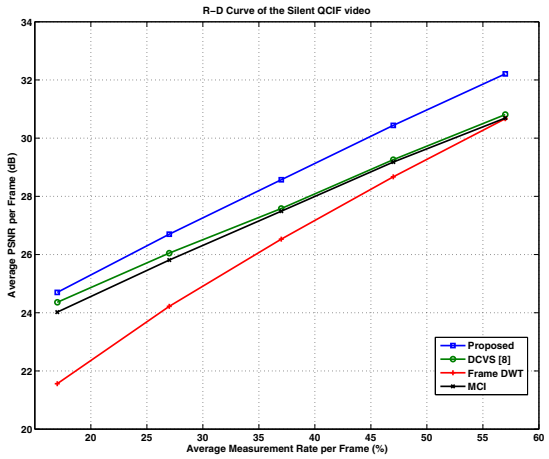

Fig. 5. MR-PSNR Performance for Silent video

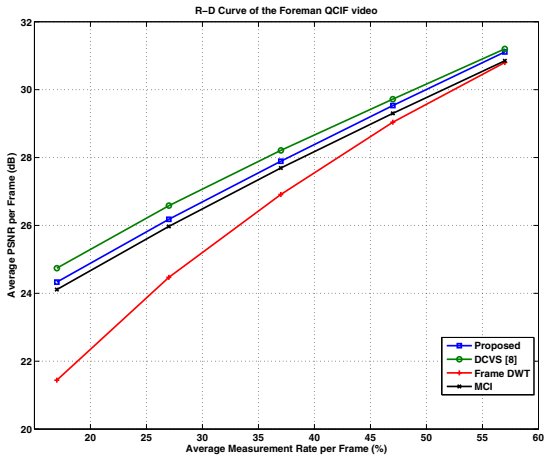

Fig. 6. MR-PSNR Performance for Foreman video

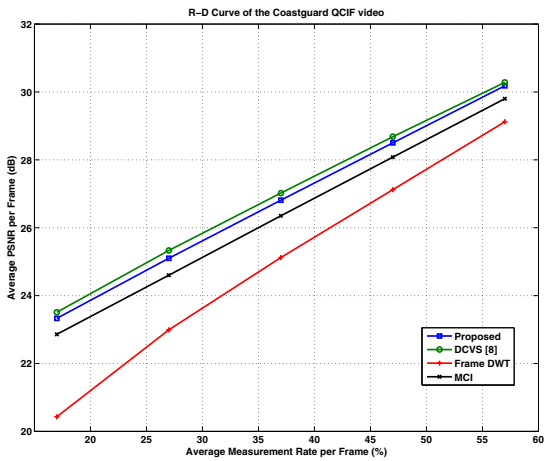

Fig. 7. MR-PSNR Performance for Coastguard video

method performs better than MCI based side information. It is marginally worse than DCVS which requires more computational resources.

6. CONCLUSIONS

In this paper, we proposed a simple and low complexity side information generation scheme for distributed compressed video coding. We showed that there is strong correlation between the CS measurements of the key and non-key frames. Therefore, the CS measurements of the key frames can be used directly as side information. This eliminates the need for side information generated from the decoded key frames. Using the proposed side information, the complexity at decoder can be significantly low as no motion estimation or other sophisticated techniques are needed. Experimental results show that the quality of reconstructed videos using the proposed side information is better than motion compensated interpolation and comparable to a more sophisticated method proposed in [8].

7. REFERENCES

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

[2] 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.

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

[4] 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.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] S. Chen and D. Donoho, "Basis pursuit," in *Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 1994, pp. 41–44.

[13] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[14] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[15] T. Do, L. G. N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proceedings of 42nd IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 27-29 October 2008.

[16] 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.

[17] A. Aaron, R. Zhang, and B. Girod, "Wyner Ziv coding of motion video," in *Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2002.

[18] 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.

[19] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the dct coefficient distributions for images," *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1661 – 1666, oct 2000.

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

[21] http://media.xiph.org/video/derf/.

[22] 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.

[23] 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.