# Quantization Effects on Compressed Sensing Video

Yousuf Baig, Edmund M-K. Lai and J.P. Lewis
School of Engineering and Advanced Technology, Massey University
Wellington, New Zealand
Email: {m.y.baig,e.lai,j.lewis}@massey.ac.nz

***Abstract*—Compressed Video Sensing (CVS) is the application of the theory and principles of Compressed Sensing to video coding. Previous research has largely ignored the effects of quantization on the random measurements. In this paper, we showed that Gaussian quantization of the CVS coefficients produce higher quality reconstructed videos compared to using MPEG and uniform quantization. Furthermore, the quantization matrix is robust against variations in the mean and standard deviations of the CS measurements among frames. Our work shows how quantization can be implemented for a practical CVS codec.**

## I. Introduction

The conventional approach to video coding removes redundancies that exist within individual images of each video frame and also between successive frames [1]. For instance, in MPEG and H.264 coding standards, discrete cosine transform (DCT) or wavelet transforms are used to exploit spatial redundancies within a video frame and motion estimation and compensation techniques are used to remove temporal redundancies between frames. These coding procedures are computationally expensive, resulting in a complex encoder. On the other hand, the decoding process is relatively simple. This approach makes sense when encoding is performed once and the video is played back many times using devices that may not have much computing power.

However, there are many other applications, such as video surveillance, where we want to deploy a large number of sensors (cameras) and the encoded video streams are not required to be played back often. In these cases, the system costs will be lower if the sensors, and hence the encoders, are relatively cheap and simple. The computing burden can be shifted to the decoding process which is performed by backend computers. Developments in the theory of Compressive Sampling or Compressed Sensing (CS) [2] has provided a theoretical foundation for us to take this new approach to video coding which we shall refer to as Compressed Video Sensing (CVS) in this paper.

CS builds upon the assumption that almost all signals contain some kind of structure that enables a compact representation. For these signals, a relatively small number of linear projections of the signals onto a random basis captures most of its essential information. The signals can be recovered from these projected measurements, or CS coefficients, through a suitable optimization process [3].

Some approaches to CVS has recently been reported [4], [5]. These works have generally assumed that the infinite precision CS measurements are available at the decoder. In practice, these measurements have to be quantized. The effects of quantization on the performance of CS based system have recently been studied theoretically for certain signal models [6], [7]. But the effects of quantization on CS video have yet to be explored.

In this paper, we focus on investigating the quantization effects on CS measurements and recovery for video signals. We found that both uniform quantization and the standard quantization matrix in MPEG do not perform well for compress-sensed videos. On the other hand, our simulation results show that Gaussian quantization performs better than uniform and MPEG quantization. Furthermore, performance is robust against mismatch between the mean and standard deviation of the quantizer and that of the actual video signals.

The rest of this paper is organized as follows. In Section II we give a brief overview of compressed sensing. Existing work on compressed video sensing is reviewed in Section III. Our proposed quantization scheme for CVS is presented in Section IV. It is tested using several standard video sequences and the results are compared to those obtained using uniform quantization as well as the standard MPEG quantization matrix. Finally, Section V concludes the paper.

## II. Compressed Sensing

Let $x = \{x[1], \dots x[N]\}$ be a discrete time real-valued random process. If $x$ is represented in a transform domain $\Psi$ 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. The sparsity of signal $x$ is measured by the number of non-zero elements in $s$. If there are $K$ non-zero coefficients (out of the $N$), then $x$ is called $K$-sparse.

The conventional approach to compression is to acquire sampled values of the signal in the time or spatial domain first. Then a suitable transform (e.g. DCT) is subsequently applied to obtain the coefficients $s$. Coefficients that are zero or insignificant are then discarded to achieve compression.

978-1-4244-5247-7/09/$26.00 ©2009 IEEE 

In compressed sensing, the idea is to acquire the significant transform coefficients directly.

Let $y$ be the length-$M$ ($M < N$) measurement vector, obtained by applying a certain measurement matrix $\Phi$ 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 [2], [3]. In order to achieve that, it is necessary for $A = \Phi\Psi$ to have a restricted isometry property [3]. In this case, the reconstruction problem can be expressed as a linear program:

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

Algorithms such as basis pursuit [8], [9] and matching pursuit [10] and their variants have been proposed to solve it.

Further results from [2], [3], [11] show that an independent identically distributed (i.i.d.) Gaussian matrix $\Phi$ satisfies the restricted isometry property for any orthonormal $\Psi$ with high probability if $M \geq cK \log(N/K)$ for some small constant $c$. The recovery of the $N$ measurements of $x$ is highly probable from only $M \approx cK \log(N/K) < N$ random Gaussian measurements $y$ under the assumption that $x$ is $K$-sparse in some domain. It is important to note that it is not known in advance which coefficients $s_i$ are zeros, or which samples $x[i]$ are not needed.

Recently, a more computationally efficient way to perform CS using structurally random matrices has been proposed [12]. The process involves pre-randomizing the signal and then a fast transform (e.g. DCT, DFT) is applied to the randomized signal. Finally, the transform coefficients are subsampled randomly to obtain the compressed measurements. Structurally random matrices are orthonormal matrices with columns permuted randomly or the sign of its entries in each column reversed simultaneously with the same probability. These matrices has the same advantage as Gaussian matrices in that they are universally incoherent with signals that are sparse in any domain except in time.

The reconstruction problem can also be formulated as a bound-constrained quadratic program. The Gradient Projection for Sparse Reconstruction (GPSR) [13] solves the quadratic program:

$$\min_x \frac{1}{2}||y - Ax||_2^2 + \tau||x||_1 \tag{4}$$

GPSR is a gradient projection algorithm has proved to be very efficient in terms of CPU utilization compared to basis pursuit and orthogonal matching pursuit. It has been used in [4], [5] to recover Compressed Sensing Video. This is also the reconstruction algorithm used in our simulations.

## III. Compressed Video Sensing

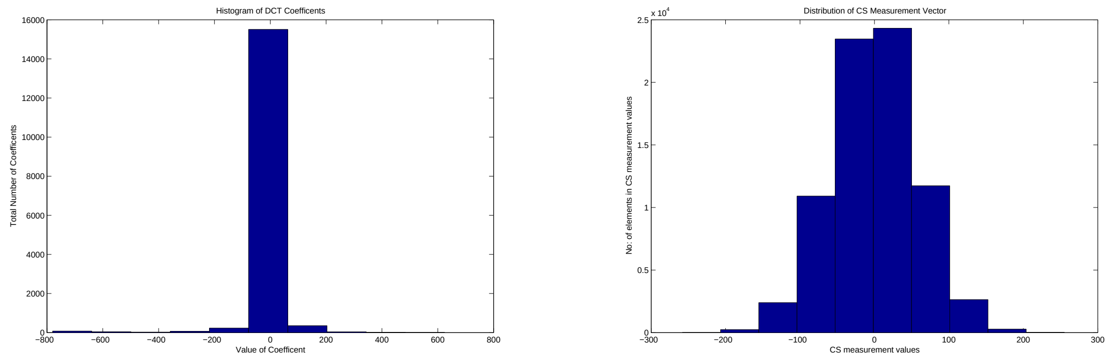
Compressed sensing has been applied in many application areas. However, its application to video coding is still at an early stage.

The first combination of CS and video is proposed in [14]. Their approach is based on single pixel camera [15]. The camera architecture employs a digital micromirror array to perform optical calculations of linear projections of an image onto pseudorandom binary patterns. It directly acquires random projections without first collecting the N pixels/voxels. They have assumed that image changes slowly across a group of snapshots which completes one frame. The have measured video sequence using a total of M measurements, which are either 2D random measurements or 3D random measurements. For 2D frame-by-frame reconstruction they have used 2D wavelets as a sparsity-inducing basis and for 3D joint reconstruction, 3D wavelets as a sparsity-inducing basis. Matching Pursuit reconstructipon algorithm is used to reconstuct compressed measurements.

In [4], CS is combined with distributed video coding which is a technique based on theorems of Slepian-Wolf and Wyner-Ziv that allows the source statistics, partially or totally, to be exploited at the decoder only [16]. The frames are divided into key frames and non-key frames. Key frames are encoded using MPEG/H.264 intra-frame coding. For non-key frames, block-based as well as frame-based CS measurement are acquired. At the decoder, the block-based measurements of a CS frame along with the two neighbouring key frames are used for generating sparsity-constraint block prediction. The block-based prediction frame is then used as side information (SI) to recover the input frame from its measurements. Unfortunately, the use of MPEG/H.264 coding for the key frames does not reduce the complexity of the encoder while it still requires substantial decoder complexity to deal with the CS-encoded non-key frames.

Another way of incorporating CS into distributed video coding is proposed in [5]. In this case, both key frames and non-key frames are acquired using CS measurements. At the decoder, each key frame is reconstructed using the GPSR algorithm. The side information generated by previously reconstructed key frames are used for reconstructing the non-key frames.

Compared with the works described above, the approach described in [17] involves the least computationally intensive encoder. Random measurements are taken independently for each frame. Motion estimation is performed at the decoder. A multiscale framework for reconstruction is proposed which iterates between motion estimation and sparsity-based reconstruction of the frames. It is built around the LIMAT method [18] for standard video compression. This approach is computationally inefficient as the decoder has to iterate




(a) DCT Coefficients (b) CS Measurements

Fig. 1. Historgram of Coefficients

through different scales to produce an efficient approximation. Another implementation of CS video coding is proposed in [19]. A video frame is split into non-overlapping blocks of equal size. The sparsity of blocks is determined by preceding reference frames which were sampled conventionally. The non-sparse blocks are sampled in a conventional manner while the sparse blocks are sampled by compressive sampling. Again, this approach does not fully utilize CS.

In all these works, it has been assumed that the data available at the decoder are of infinite precision. In practice, some form of quantization is always required.

## IV. Quantization in Compressed Video Sensing

The quantizer is very important part of the encoding process. An optimal quantizer should be tailored to the signal concerned and minimize the amount of distortion in the reconstructed signal [20]. However, for practical reasons, fixed quantizers that are sub-optimal are always used.

In video coding standards, 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 [21].

### *A. Quantizer Design for CVS*

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. Due to the need for satisfying the restricted isometry

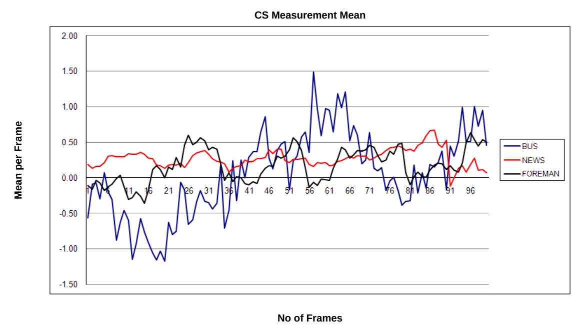



Fig. 2. Mean Graph

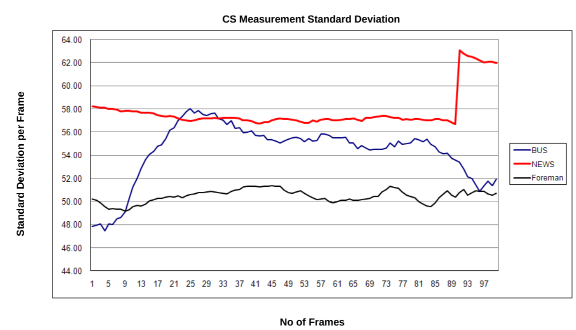



Fig. 3. Standard Deviation Graph

property as described in Section II, the i.i.d. Gaussian matrix is often used. Thus we would expect the distribution of CS coefficients to be Gaussian. Figure 1(a) and Figure 1(b) show the histograms of the DCT and CS coefficients, respectively, for one frame of video taken from the "news" sequence. As expected, the DCT coefficient values concentrate on the lower end of the frequency spectrum. The majority of the DCT coefficients are zero or close to zero. On the other hand, the CS measurement values follow a more or less normal (Gaussian) distribution. Histograms of other frames show a similar pattern. This indicates that the quantizer for CVS should be different from those used in the current video coding

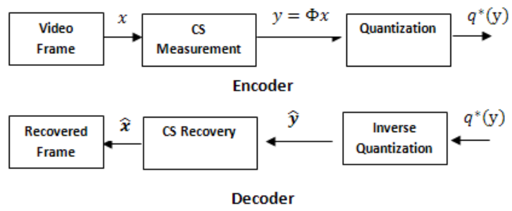



Fig. 4. System Block Diagram

standards.

To design a quantization matrix for video CS measurements, we computed the mean and standard deviation of first 100 frames of three video sequences: "news", "bus" and "foreman". Figures 2 and 3 show the variations of the mean and standard deviation, respectively, from frame to frame. While these parameters do not vary much for some sequences like "news" and "foreman", the variation is more significant for others. One question that we seek to answer is whether a fixed quantizer is robust to these variations.

In our experiments, a quantization matrix which follows the random Gaussian distribution is used. We take the average value of the means and standard deviations of the first 100 frames of the "news" sequence and generate a random quantization matrix using (5).

$$Q = rnd(x) * sqrt(var) + mean \quad (5)$$

One problem in designing a matrix this way is that due to Gaussian distribution, few matrix values are close to zero or 1 or too high. To make quantization matrix weights appropriate, we change these values to 16, which is suitable as our results shows in next section. For too much higher values, the square root can be taken to make them appropriate.

### *B. Simulation Model*

In order to isolate the effects of quantization on CVS, we have chosen to use a model that does not incorporate any side information for reconstruction. The system model shown in Figure 4. At the encoder, the CS measurement process is applied to each individual frame independently. This process is performed using structurally random matrices [12] with DCT as the transform. Blocks of $8 \times 8$ CS measurements are then quantized using the quantizer generated as described above. It should be emphasized that the same quantization matrix is used for all video frames of all the sequences being tested. In other words, the quantizer is not adaptive.

At the decoder, inverse quantization is followed by the CS recovery process to reconstruct the signal. We used the GPSR algorithm [13] for CS recovery.

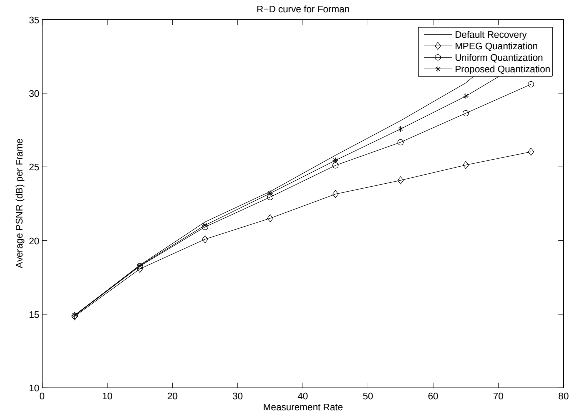



Fig. 5. Average PSNR(dB) vs. CS measurement rate for various quantization schemes using the Foreman Sequence.

Note that this framework is computationally expensive, but our aim is to analyze the effects of quantization only.

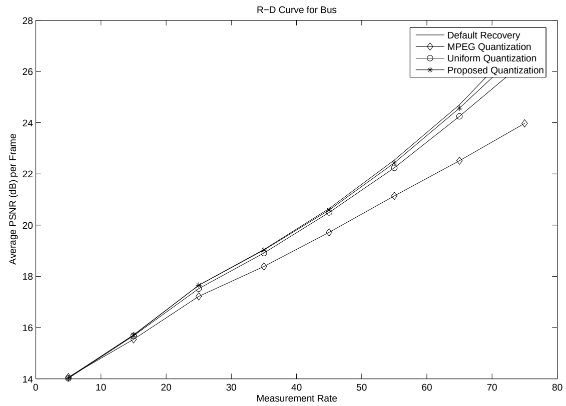
### *C. Experimental Results*

Three CIF video sequences, "foreman", "news" and "bus", with a frame size of $352 \times 288$ pixels are used for our experiments. Three different quantization schemes are compared. The first one is the MPEG quantization matrix as defined in the MPEG standard for intra-frame coding [1]. The second one is a uniform quantization matrix [21]. The third one is our CVS quantization matrix as described in Section IV-A.

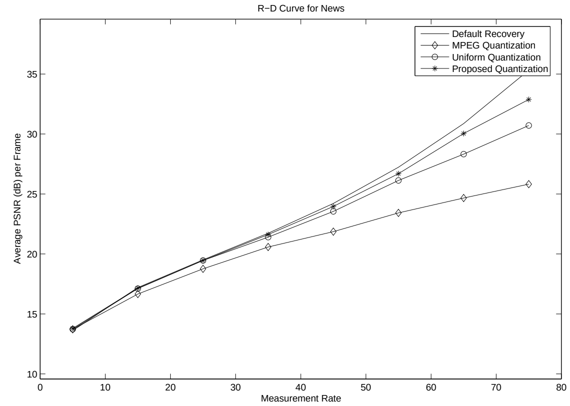
Figure 5 shows the average PSNR against different CS measurements rates for the "foreman" video sequence using the three quantization schemes as well as no quantization. Our proposed quantization method produces better results than both the other two quantization schemes for all measurement rates. Similar results are obtained for the other two video sequences as shown in Figure 6.

Our results show that even though the proposed Gaussian quantization matrix is generated based on the statistics of a part of one particular video sequence, it works well for other sequences with different statistics. This shows that a fixed quantization matrix based on the Gaussian distribution can be used in general. This is confirmed when the same quantization matrix is applied to a number of other video test sequences obtained from [22]. These results are shown in Table I. For example, a 33% gain on MPEG and 9% over uniform quantization is obtained when the Gaussian quantization matrix is used.

Figure 7 shows one frame of foreman video sequence. The visual quality obtained with and without quantization can be compared with the original. A 75% CS measurement rate is used. It shows that our proposed quantization scheme produces better results visually than uniform and MPEG quantization.




(a) Bus sequence




(b) News sequence

Fig. 6. Average PSNR(dB) vs. CS measurement rate for various quantization schemes.

TABLE I
PERFORMANCE COMPARISON OF 50TH FRAME, PSNR(DB)

| Sequence | No quantization | MPEG | Uniform | Proposed |
|---|---|---|---|---|
| Foreman | 40.77 | 27.17 | 33.33 | 36.41 |
| News | 42.48 | 27.21 | 33.58 | 37.22 |
| Bus | 33.25 | 25.87 | 30.11 | 31.76 |
| City | 35.31 | 26.73 | 31.48 | 33.44 |
| Coastguard | 34.68 | 24.41 | 31.14 | 32.94 |
| Crew | 41.64 | 27.35 | 33.32 | 36.55 |
| mobile | 29.04 | 24.49 | 27.44 | 28.34 |
| paris | 36.66 | 26.16 | 31.28 | 33.79 |
| stefan | 34.07 | 25.80 | 30.26 | 32.10 |

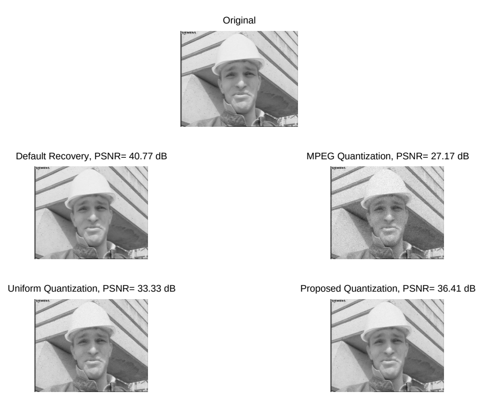



Fig. 7. Reconstruction visual quality for 50th frame of foreman sequence

In particular, the visual quality is particularly poor for MPEG quantization. The proposed quantization matrix produces better results than Uniform and MPEG quantization.

## V. CONCLUSIONS

We studied the effects of quantization on compressed sensing video. Our results show that a quantization matrix for CS coefficients can be designed using random Gaussian distribution. This type of quantization matrix has been shown to be robust with respect to the mean and standard deviation of the individual video frames. Our results open up a way to design practical CVS codecs with quantization incorporated. While theoretical analysis of quantization has been reported for compressed sensing, our work is the first, as far as we know, to report on experimental results on quantized CS video.

## REFERENCES

[1] P. Symes, *Digital Video Compression*. McGraw-Hill, 2004.
[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[4] T. Do, Y. Chen, D. Nguyen, N. Nguyen, L. Gan, and T. Tran, "Distributed compressed video sensing," in *43rd Annual Conf. on Information Sciences and Systems (CISS)*, March 2009, pp. 1–2.
[5] L.-W. Kang and C.-S. Lu, "Distributed compressive video sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 1169–1172.
[6] O. Milenkovic, W. Dai, and H. V. Pham, "A comparative study of quantized compressive sensing schemes," in *IEEE Int. Symp. on Information Theory (ISIT)*, 28 June-3 July 2009, pp. 11–15.
[7] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *IEEE Int. Symp. on Information Theory (ISIT)*, 28 June-3 July 2009, pp. 6–10.
[8] S. Chen and D. Donoho, "Basis pursuit," in *28th Annual Asilomar Conference on Signals, Systems, and Computers*, vol. 1, Nov. 1994, pp. 41–44.
[9] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
[10] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 3397–3415, Dec. 1993.
[11] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[12] T. Do, L. Gan., and T. D. Tran, "Fast compressive sampling with structurally random matrices," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 30 March - 4 April 2008, pp. 3369–3372, journal paper submitted. (Matlab code available from http://thanglong.ece.jhu.edu/cs).

[13] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007, (Matlab code available from http://www.lx.it.pt/ mtf/GPSR).

[14] 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 *Picture Coding Symposium (PCS)*, May 2006.

[15] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new camera architecture based on optical- domain compression," in *SPIE Symposium on Electronic Imaging: Computational Imaging*, vol. 6065, 2006.

[16] F. Pereira, "Distributed video coding: Basics, main solutions and trends," in *IEEE Int. Conf. on Multimedia and Expo (ICME)*, 28 June-3 July 2009, pp. 1592–1595.

[17] J. Y. Park and M. Wakin, "A multiscale framework for compressive sensing of video," in *Picture Coding Symposium (PCS)*, May 2009, pp. 1–4.

[18] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (limat) framework for highly scalable video compression," *IEEE Trans. Image Processing*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[19] V. Stankovic, L. Stankovic, and S. Chencg, "Compressive video sampling," in *16th European Signal Processing Conf., Lausanne, Switzerland*, aug 2008.

[20] R. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

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

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