Compressive Video Coding:
A Review of the State-Of-The-Art

Muhammad Yousuf Baig, Edmund M-K. Lai and Amal Punchihewa
Massey University
School of Engineering and Advanced Technology
Palmerston North
New Zealand

1. Introduction

Video coding and its related applications have advanced quite substantially in recent years. Major coding standards such as MPEG [1] and H.26x [2] are well developed and widely deployed. These standards are developed mainly for applications such as DVDs where the compressed video is played over many times by the consumer. Since compression only needs to be performed once while decompression (playback) is performed many times, it is desirable that the decoding/decompression process can be done as simply and quickly as possible. Therefore, essentially all current video compression schemes, such as the various MPEG standards as well as H.264 [1, 2] involve a complex encoder and a simple decoder. The exploitation of spatial and temporal redundancies for data compression at the encoder causes the encoding process to be typically 5 to 10 times more complex computationally than the decoder [3]. In order that video encoding can be performed in real time at frame rates of 30 frames per second or more, the encoding process has to be performed by specially designed hardware, thus increasing the cost of cameras.

In the past ten years, we have seen substantial research and development of large sensor networks where a large number of sensors are deployed. For some applications such as video surveillance and sports broadcasting, these sensors are in fact video cameras. For such systems, there is a need to re-evaluate conventional strategies for video coding. If the encoders are made simpler, then the cost of a system involving tens or hundreds of cameras can be substantially reduced in comparison with deploying current camera systems. Typically, data from these cameras can be sent to a single decoder and aggregated. Since some of the scenes captured may be correlated, computational gain can potentially be achieved by decoding these scenes together rather than separately. . Decoding can be simple reconstruction of the video frames or it can be combined with detection algorithms specific to the application at hand. Thus there are benefits in combing reduced complexity cameras with flexible decoding processes to deliver modern applications which are not anticipated when the various video coding standards are developed.

Recently, a new theory called Compressed Sensing (CS) [4, 5, 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 amounts 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. Even though practical cameras that make use of CS are still in their very early days, the concept can be applied to video coding. A lower sampling rate implies less energy required for data processing, leading to lower power requirements for the camera. Furthermore, the complexity of the encoder can be further simplified by making use of distributed source coding [21, 22]. The distributed approach provides ways to encode video frames without exploiting any redundancy or correlation between video frames captured by the camera. The combined use of CS and distributed source coding can therefore serve as the basis for the development of camera systems where the encoder is less complex than the decoder.

We shall first provide a brief introduction to Compressed Sensing in the next Section. This is followed by a review of current research in video coding using CS.

2. Compressed sensing

Shannon's uniform sampling theorem [7, 8] provides a lower bound on the rate by which an analog signal needs to be sampled in order that the sampled signal fully represents the original. If a signal $f(t)$ contains no frequencies higher than ω_{max} radians per second, then it can be completely determined by samples that are spaced $T = 2\pi/\omega_{max}$ seconds apart. $f(x)$ can be reconstructed perfectly using the these samples $f(nT)$ by

$$f(t) = \sum_{\{k \in z\}} f(kT) sinc\ (t/T - k)$$

(1.1)

The uniform samples $f(nT)$ of $f(t)$ may be interpreted as coefficients of basis functions obtained by shifting and scaling of the sinc function. For high bandwidth signals such as video, the amount of data generated based on a sampling rate of at least twice the bandwidth is very high. Fortunately, most of the raw data can be thrown away with almost no perceptual loss. This is the result of lossy compression techniques based on orthogonal transforms. In image and video compression, the discrete cosine transform (DCT) and wavelet transform have been found to be most useful. The standard procedure goes as follows. The orthogonal transform is applied to the raw image data, giving a set of transform coefficients. Those coefficients that have values smaller than a certain threshold are discarded. Only the remaining significant coefficients, typically only a small subset of the original, are encoded, reducing the amount of data that represents the image. This means that if there is a way to acquire only the significant transform coefficients directly by sampling, then the sampling rate can be much lower than that required by Shannon's theorem.

Emmanuel Candes, together with Justin Romberg and Terry Tao, came up with a theory of Compressed Sensing (CS) [9] that can be applied to signals, such as audio, image and video

that are *sparse* in some domain. This theory provides a way, at least theoretically, to acquire signals at a rate potentially much lower than the Nyquist rate given by Shannon's sampling theorem. CS has already inspired more than a thousand papers from 2006 to 2010 [9].

2.1 Key Elements of compressed sensing

Compressed Sensing [4-6, 10] is applicable to signals that are sparse in some domain. Sparsity is a general concept and it expresses the idea that the information rate or the signal significant content may be much smaller than what is suggested by its bandwidth. Most natural signals are redundant and therefore compressible in some suitable domain. We shall first define the two principles, sparsity and incoherence, on which the theory of CS depends.

2.1.1 Sparsity

Sparsity is important in Compressed Sensing as it determines how efficient one can acquire signals non-adaptively. The most common definition of sparsity used in compressed sensing is as follows. Let $f \in R^n$ be a vector which represents a signal which can be expanded in an orthonormal basis $\psi = [\psi_1 \psi_2 \cdots \psi_n]$ as

$$f(t) = \sum_{i-1}^{n} x_i \psi_i \tag{1.2}$$

Here, the coefficients $x_i = \langle f, \psi_i \rangle$. In matrix form, (1.2) becomes

$$\boldsymbol{f} = \boldsymbol{\Psi x} \tag{1.3}$$

When all but a few of the coefficients x_i are zero, we say that f is sparse in a strict sense. If S denotes the number of non-zero coefficients with $S \ll n$, then f is said to be S-sparse. In practice, most compressible signals have only a few significant coefficients while the rest have relatively small magnitudes. If we set these small coefficients to zero in the way that it is done in lossy compression, then we have a sparse signal.

2.1.2 Incoherence

We start by considering two different orthonormal bases, Φ and Ψ, of R^n. The coherence between these two bases is defined in [10] by

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \le k, j \le n} |\langle \phi_k, \psi_k \rangle| \tag{1.4}$$

which gives us the largest correlation between any two elements of the two bases. It can be shown that

$$\mu(\Phi, \Psi) \in [1, \sqrt{n}] \tag{1.5}$$

Sparsity and incoherence together quantify the compressibility of a signal. 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. Interestingly, random matrices are largely incoherent with any fixed basis [18].

2.1.3 Random sampling

Let X be a discrete time random process and let $x = \{x[1], \dots x[N]\}$ be a vector of N real-valued samples of X. If the representation of x in a transform domain Ψ given by s, then

$$x = \Psi s = \sum_{i-1}^{N} s_i \psi_i \tag{1.6}$$

Here $\Psi = [\psi_1, \psi_2 \dots \psi_N]$ is the transform basis matrix and $s = [s_1 \dots s_N]$ is the vector of weighted coefficients where $s_i = <x, \psi_i>$.

Consider a general linear measurement process that computes $M < N$ inner products of $y_i = \langle x, \phi_j \rangle$ between x and a collection of vectors $\{\phi_j\}_1^M$. Let Φ denote the $M \times N$ matrix with the measurement vectors ϕ_j as rows. Then y is given by

$$y = \Phi x = \Phi \Psi x = \Theta s \tag{1.7}$$

where $\Theta = \Phi \Psi$. If Φ is fixed, then the measurements are not adaptive or depend on the structure of the signal x [6]. The minimum number of measurements needed to reconstruct the original signal depends on the matrices Φ and Ψ.

Theorem 1 [11]. Let $f \in R^N$ has a discrete coefficient sequence x in the basis Ψ. Let x be S-sparse. Select M measurements in the domain Φ uniformly at random. Then if

$$M \geq C \cdot \mu^2(\Phi, \Psi) \cdot S \log N \tag{1.8}$$

for some positive constant C, then with high probability, x can be reconstructed using the following convex optimization program:

$$\min_{\tilde{x}} \left| \left| \tilde{x} \right| \right|_{l1} subject\ to\ y_k = \langle \phi_k, \Psi \tilde{x} \rangle, \forall\ k \in J \tag{1.9}$$

where J denotes the index set of the M randomly chosen measurements.

This is an important result and provides the requirement for successful reconstruction. It has the following three implications [10]:

i.   The role of the coherence in above equation is transparent – the smaller the coherence between the sensing and basis matrices, the fewer the number of measurements needed.
ii.   It provides support that there will be no information loss by measuring any set of m coefficients, which may be far less than the original signal size.
iii. The signal f can be exactly recovered without assuming any knowledge about the non-zero coordinates of x or their amplitudes.

2.1.4 CS Reconstruction

The reconstruction problem in CS involves taking the M measurements y to reconstruct the length-N signal x that is K-sparse, given the random measurement matrix Φ and the basis matrix Ψ. Since $M < N$, this is an ill-conditioned problem. The classical approach to solving ill-conditioned problems of this kind is to minimize the l_2 norm. The general problem is given by

$$\hat{s} = \arg\min \|s'\|_2 \; such \; that \; \mathbf{\Theta}s' = \boldsymbol{y} \tag{1.10}$$

However, it has been proven that this l_2 minimization will never return a K-sparse solution. Instead, it can only produce a non-sparse solution [6]. The reason is that the l_2 norm measures the energy of the signal and signal sparsity properties could not be incorporated in this measure.

The l_0 norm counts the number of non-zero entries and therefore allows us to specify the sparsity requirement. The optimization problem using this norm can be stated as

$$\hat{s} = \arg\min \|s'\|_0 \; such \; that \; \mathbf{\Theta}s' = \boldsymbol{y} \tag{1.11}$$

There is a high probability of obtaining a solution using only $M = K + 1$ i.i.d Gaussian measurements [10]. However, the solution produced is numerically unstable [6]. It turns out that optimization based on the l_1 norm is able to exactly recover K sparse signals with high probability using only $M \geq cK \log N/K$ i.i.d Gaussian measurements [4, 5]. The convex optimization problem is given by

$$\hat{s} = \arg\min \|s'\|_1 \; such \; that \; \Theta s' = y \tag{1.12}$$

Which can be reduced to a linear program. Algorithms based on Basis Pursuit [12] can be used to solve this problem with a computational complexity of $O(N^3)$ [4].

3. Compressed Video Sensing (CVS)

Research into the use of CS in video applications has only started very recently. We shall now briefly review what has been reported in the open literature.

The first use of CS in video processing is proposed in [13]. Their approach is based on the single pixel camera [14]. The camera architecture employs a digital micro mirror array to perform optical calculations of linear projections of an image onto pseudo-random binary patterns. It directly acquires random projections. They have assumed that the image changes slowly enough across a sequence of snapshots which constitutes one frame. They acquired the video sequence using a total of M measurements, which are either 2D or 3D random measurements. For 2D frame-by-frame reconstruction, 2D wavelets are used as the sparsity-inducing basis. For 3D joint reconstruction, 3D wavelets are used. The Matching Pursuit reconstruction algorithm [15] is used for reconstruction.

Another implementation of CS video coding is proposed in [16]. In this implementation, each video frame classified as a reference or non-reference frame. A reference frame (or key frame) is sampled in the conventional manner while non-reference frames are sampled by CS techniques. The sampled reference frame is divided into B non-overlapping blocks each of size $n \times n = N$ pixels whereby discrete cosine transform (DCT) is applied. A compressed sensing test is applied to the DCT coefficients of each block to identify the sparse blocks in the non-reference frame. This test basically involves comparing the number of significant DCT coefficients against a threshold T. If the number of significant coefficients is small, then the block concerned is a candidate for CS to be applied. The sparse blocks are compressively sampled using an i.i.d. Gaussian measurement matrix and an inverse DCT sensing matrix. The remaining blocks are sampled in the traditional way. A block diagram of the encoder is shown in Figure 1.

Signal recovery is performed by the OMP algorithm [17]. In reconstructing compressively sampled blocks, all sampled coefficients with an absolute value less than some constant C are set to zero. Theoretically, if there are $N - K$ non significant DCT coefficients, then at least $M = K + 1$ samples are needed for signal reconstruction [10]. Therefore the threshold is set to $T < N - K$. The choice of values for M, T, and C depends on the video sequence and the size of the blocks. They have proved experimentally that up to 50% of savings in video acquisition is possible with good reconstruction quality.
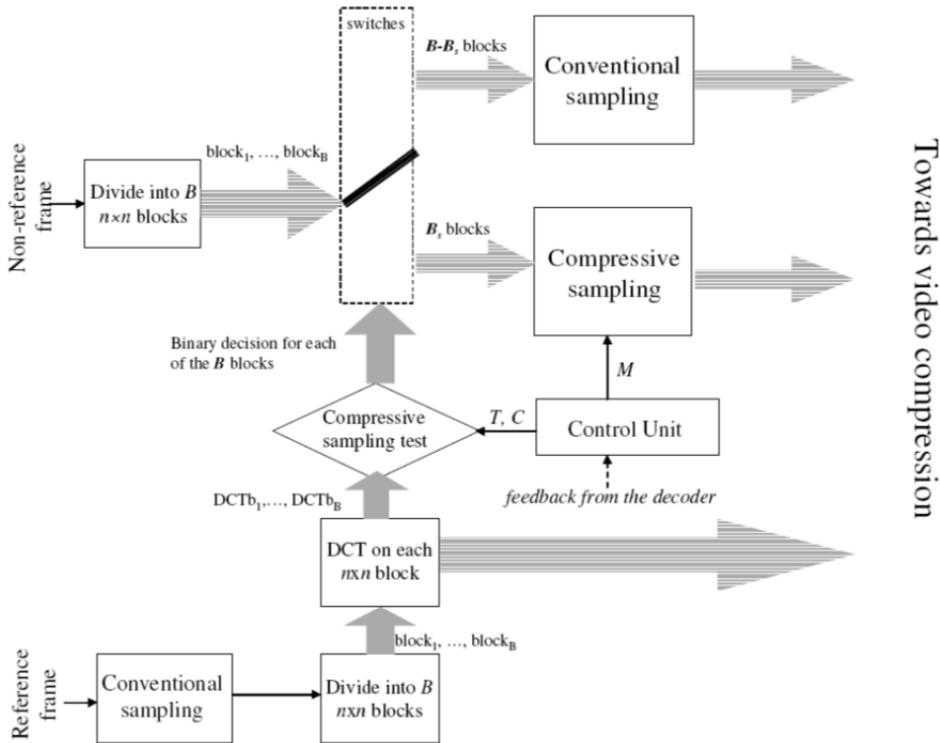
Fig. 1. System Block Diagram of Video Coding Scheme Proposed in [16]

Another technique which uses motion compensation and estimation at the decoder is presented in [18]. At the encoder, only random CS measurements were taken independently from each frame with no additional compression. A multi-scale framework has been proposed for reconstruction which iterates between motion estimation and sparsity-based reconstruction of the frames. It is built around the LIMAT method for standard video compression [19].

LIMAT [19] uses a second generation wavelets to build a fully invertible transform. To incorporate temporal redundancy, LIMAT adaptively apply motion-compensated lifting steps. Let k-th frame of the n frame video sequence is given by x_k, where $k \in \{1,2, \dots\}$. The lifting transform partitions the video into even frames $\{x_{2k}\}$ and odd frames $\{x_{2k+1}\}$ and attempts to predict the odd frames from the even ones using a forward motion compensation operator. Suppose $\{x_{2k}\}$ and $\{x_{2k+1}\}$ differ by a 3-pixel shift that is captured precisely by a motion vector v_f, then it is given by $\{x_{2k+1}\} = F(x_{2k}, v_f)$ exactly.

The proposed algorithm in [18] uses block matching (BM) to estimate motion between a pair of frames. Their BM algorithm divides the reference frame into non-overlapping blocks. For each block in the reference frame the most similar block of equal size in the destination frame is found and the relative location is stored as a motion vector. This approach overcomes previous approaches such as [13] where the reconstruction of a frame depends only on the individual frame's sparsity without taking into account any temporal motion. It is also better than using inter-frame difference [20] which is insufficient for removing temporal redundancies.

3.2 Distributed Compressed Video Sensing (DCVS)

Another video coding approach that makes use of CS is based on the distributed source coding theory of Slepian and Wolf [21], and Wyner and Ziv [22]. Source statistics, partially or totally, is only exploited at the decoder, not at the encoder as it is done conventionally. Two or more statistically dependent source data are encoded by independent encoders. Each encoder sends a separate bit-stream to a common decoder which decodes all incoming bit streams jointly, exploiting statistical dependencies between them.

In [23], a framework called Distributed Compressed Video Sensing (DISCOS) is introduced. Video frames are divided into key frames and non-key frames at the encoder. A video sequence consists of several GOPs (group of pictures) where a GOP consists of a key frame followed by some non-key frames. Key frames are coded using conventional MPEG intra-coding. Every frame is both block-wise and frame-wise compressively sampled using structurally random matrices [25]. In this way, more efficient frame based measurements are supplemented by block measurement to take advantage of temporal block motion.

At the decoder, key frames are decoded using a conventional MPEG decoder. For the decoding of non-key frames, the block-based measurements of a CS frame along with the two neighboring key frames are used for generating sparsity-constraint block prediction. The temporal correlation between frames is efficiently exploited through the inter-frame sparsity model, which assumes that a block can be sparsely represented by a linear combination of few temporal neighboring blocks. This prediction scheme is more powerful than conventional block-matching as it enables a block to be adaptively predicted from an optimal number of neighboring blocks, given its compressed measurements. The block-based prediction frame is then used as the side information (SI) to recover the input frame from its measurements. The measurement vector of the prediction frame is subtracted from that of the input frame to form a new measurement vector of the prediction error, which is sparse if the prediction is sufficiently accurate. Thus, the prediction error can be faithfully recovered. The reconstructed frame is then simply the sum of the prediction error and the prediction frame.

Another DCVS scheme is proposed in [24]. The main difference from [23] is that both key and non-key frames are compressively sampled and no conventional MPEG/H.26x codec is required. However, key frames have a higher measurement rate than non-key frames.

The measurement matrix Φ is the scrambled block Hadamard ensemble (SBHE) matrix [28]. SBHE is essentially a partial block Hadamard transform, followed by a random permutation of its columns. It provides near optimal performance, fast computation, and memory efficiency. It outperforms several existing measurement matrices including the Gaussian i.i.d matrix and the binary sparse matrix [28]. The sparsifying matrix used is derived from the discrete wavelet transform (DWT) basis.

At the decoder, the key frames are reconstructed using the standard Gradient Projection for Sparse Reconstruction (GPSR) algorithm. For the non-key frames, in order to compensate for lower measurement rates, side information is first generated to aid in the reconstruction. Side information can be generated from motion-compensated interpolation from neighboring key frames. In order to incorporate side information, GPSR is modified with a special initialization procedure and stopping criteria are incorporated (see Figure 3). The convergence speed of the modified GPSR has been shown to be faster and the reconstructed video quality is better than using original GPSR, two-step iterative shrinkage/thresholding (TwIST) [29], and orthogonal matching pursuit (OMP) [30].
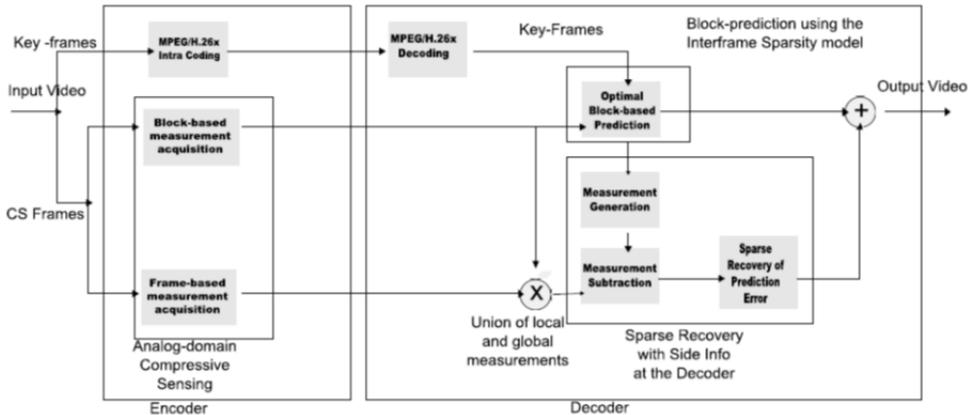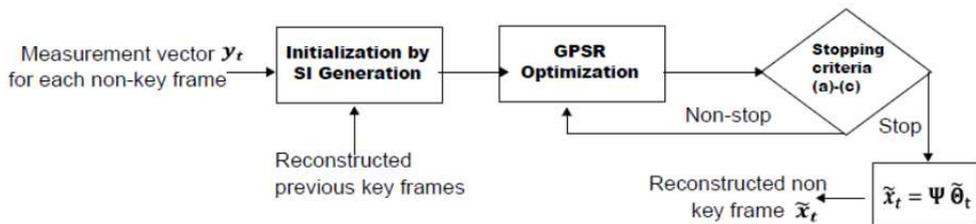

Fig. 2. Architecture of DISCOS [23]

Fig. 3. Distributed CS Decoder [24]

3.3 Dictionary based compressed video sensing

In dictionary based techniques, a dictionary (basis) is created at the decoder from neighbouring frames for successful reconstruction of CS frames.

A dictionary based distributed approach to CVS is reported in [32]. Video frames are divided into key frames and non-key frames. Key frames are encoded and decoded using conventional MPEG/H.264 techniques. Non-key frames are divided into non-overlapping blocks of n pixels. Each block is then compressively sampled and quantized. At the decoder, key frames are MPEG/H.264 decoded while the non-key frames are dequantized and recovered using a CS reconstruction algorithm with the aid of a dictionary. The dictionary is constructed from the decoded key frame. The architecture of this system is shown in Figure 4.
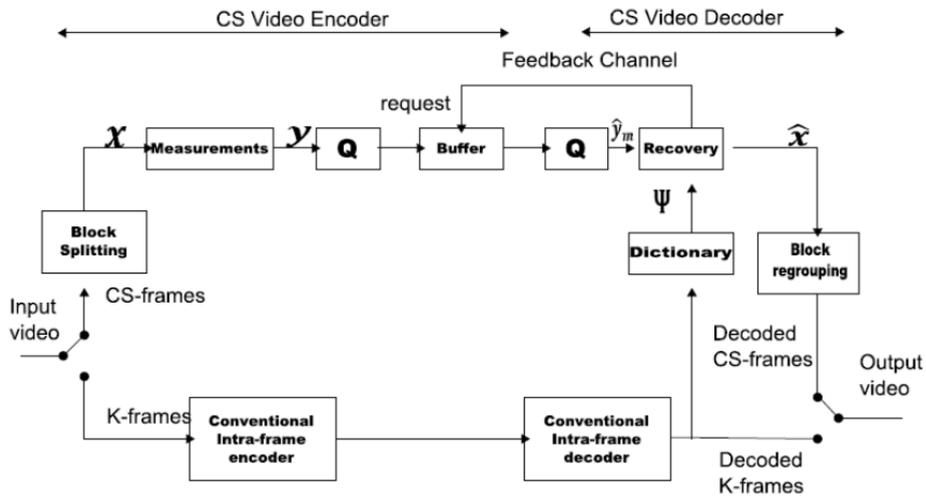
Fig. 4. A Dictionary-based CVS System [32]

Two different coding modes are defined. The first one is called the SKIP mode. This mode is used when a block in a current non-key frame does not change much from the co-located decoded key frame. Such a block is skipped for decoding. This is achieved by increasing the complexity at the encoder since the encoder has to estimate the mean squared error (MSE) between decoded key frame block and current CS frame block. If the MSE is smaller than some threshold, the same decoded block is simply copied to current frame and hence the decoding complexity is very minimal. The other coding mode is called 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 pre-determined threshold, then the block is marked as a decoded block. Dictionary is created from a set of spatially neighboring blocks of previous decoded neighboring key frames. 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.

Another dictionary based approach is presented in [33]. The authors proposed the idea of using an adaptive dictionary. The dictionary is learned from a set of blocks globally extracted from the previous reconstructed neighboring frames together with the side information generated from them is used as the basis of each block in a frame. In their encoder, frame are divided as Key-frames and CS frames. For Key-frames, frame based CS measurements are taken and for CS frames, block based CS measurements are taken. At the decoder, the reconstruction of a frame or a block can be formulated as an l_1-minimization problem. It is solved by using the sparse reconstruction by separable approximation (SpaRSA) algorithm [34]. Block diagram of this system is shown in Figure 5.

Adjacent frames in the same scene of a video are similar, therefore a frame can be predicted by its side information which can be generated from the interpolation of its neighboring reconstructed frames. at decoder in [33], for a CS frame x_t , its side information I_t can be generated from the motion-compensated interpolation (MCI) of its previous x_{t-1} and next reconstructed key frames x_{t+1}, respectively. To learn the dictionary from x_{t-j} , I_t and x_{t+j} , Q training patches were extracted. For each block in the three frames, 9 training patches
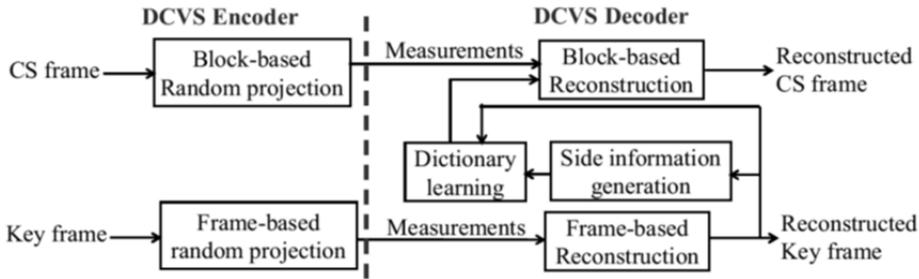
Fig. 5. Distributed Compressed Video Sensing with Dictionary Learning

including the nearest 8 blocks overlapping this block and this block itself are extracted. After that, the K-SVD algorithm [35] is applied to Q training patches to learn the dictionary $D_t \in R^{N_b \times P}$, $N_b \leq P$. D_t is an overcomplete dictionary containing P atoms. By using the learned dictionary D_t, each block b_{ti} in x_t can be sparsely represented as a sparse coefficient vector $\alpha_{ti} \in R^{P \times 1}$. This learned dictionary provides sparser representation for the frame than using the fixed basis dictionary. Same authors have extended their work in [36] for dynamic measurement rate allocation by incorporating feedback channel in their dictionary based distributed video codec.

4. Summary

CS is a new field and its application to video systems is even more recent. There are many avenues for further research and thorough quantitative analyses are still lacking. Key encoding strategies adopted so far includes:

- Applying CS measurements to all frames (both key frames and non-key frames) as suggested by [24].
- Applying conventional coding schemes (MPEG/H.264) to key frames and acquire local block-based and global frame-based CS measurements for non-key frames as suggested in [23, 32].
- Split frames into non-overlapping blocks of equal size. Reference frames are sampled fully. After sampling, a compressive sampling test is carried out to identify which blocks are sparse [16].

Similarly, key decoding strategies includes:

- Reconstructing the key frames by applying CS recovery algorithms such as GPSR and reconstruct the non-key frames by incorporating side information generated by recovered key frames [24].
- Decoding key frames using conventional image or video decompression algorithms and perform sparse recovery with decoder side information for prediction error reconstruction. Add reconstructed prediction error to the block-based prediction frame for final frame reconstruction [23].
- Using a dictionary for decoding [32] where a dictionary is used for comparison and prediction of non-key frames. Similarly, a dictionary can be learned from neighboring frames for reconstruction of non-key frames [33].

These observations suggest that there are many different approaches to encode videos using CS. In order to achieve a simple encoder design, conventional MPEG type of encoding

process should not be adopted. Otherwise, there is no point in using CS as an extra overhead. We believe that the distributed approach in which each key-frame and non-key-frame is encoded by CS is able to utilise CS more effectively. While spatial domain compression is performed by CS, temporal domain compression is not exploited fully since there is no motion compensation and estimation performed. Therefore, a simple but effective inter-frame compression will need to be devised. In the distributed approach, this is equivalent to generating effective side information for the non-key frames.

5. 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 Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[4] D. Donoho, "Compressed sensing," *IEEE Transactions on Information 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 Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[7] C. Shannon, "Communication in the presence of noise," *Proceedings of IRE*, vol. 37, pp. 10–21, Jan. 1949.

[8] ——, "Classic paper: Communication in the presence of noise," *Proceedings of the IEEE*, vol. 86, no. 2, pp. 447–457, Feb. 1998.

[9] J. Ellenberg, "Fill in the blanks: Using math to turn lo-res datasets into hi-res samples," *Wired Magazine*, vol. 18, no. 3, Mar. 2010.

[10] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21–30, Mar 2008.

[11] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[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] 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 *Proceedings of Picture Coding Symposium*, Beijing, China, 24-26 April 2006.

[14] 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 *Proceedings of SPIE Symposium on Electronic Imaging: Computational Imaging*, vol. 6065, 2006.

[15] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 3397–3415, Dec. 1993.

[16] V. Stankovic, L. Stankovic, and S. Chencg, "Compressive video sampling," in *Proceedings of 16th European Signal Processing Conference*, Lausanne, Switzerland, Aug. 2008.

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

[18] J.Y.Park and M. B. Wakin, "A Multiscale Framework for Compressive Sensing of Video," Picture Coding Symposium Chicago, Illinois, 2009.

[19] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *Image Processing, IEEE Transactions on,* vol. 12, pp. 1530-1542, 2003.

[20] R. Marcia and R. Willett, "Compressive Coded Aperture Video Reconstruction," in *16th European Signal Processing Conference, EUSIPCO-2008* Lausanne, Switzerland, 2008.

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

[22] A. Wyner, "Recent results in the Shannon theory," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.

[23] T. T. Do, C. Yi, D. T. Nguyen, N. Nguyen, G. Lu, and T. D. Tran, "Distributed Compressed Video Sensing," in *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, 2009, pp. 1-2.

[24] K. Li-Wei and L. Chun-Shien, "Distributed compressive video sensing," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 1169-1172.

[25] T.T.Do, L.Gan, and T.D.Tran, "Fast and efficient compressive sampling using structural Random Matrices," *To be submitted to IEEE Trans. of Information Theory, 2008,* 2008.

[26] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Aslimore Conference on Signals, Systems and Computers* Pacific Grove, California, 2008.

[27] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *Selected Topics in Signal Processing, IEEE Journal of,* vol. 1, pp. 586-597, 2007.

[28] L. Gan, T.T.Do, and T.D.Tran, "Fast compressive imaging using scrambled hadamard block ensemble," in *16th Eurpoian Signal Processing Conference*, Lausanne, Switzerland, 2008.

[29] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," *Image Processing, IEEE Transactions on,* vol. 16, pp. 2992-3004, 2007.

[30] T. Blumensath and M. E. Davies, "Gradient Pursuits," *Signal Processing, IEEE Transactions on,* vol. 56, pp. 2370-2382, 2008.

[31] K. Simonyan and S. Grishin, "AviSynth MSU frame rate conversion filter." http://www. compression.ru/video/frame_rate_conversion/index_en_msu.html.

[32] 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.

[33] Hung-Wei Chen, K. Li-Wei and L. Chun-Shien, "Dictionary Leraning-Based Distributed Compressive Video Sensing," in 28th Picture Coding Symposium. PCS 2010. *Dec 8-10, 2010* , pp. 210-213.

[34] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. on Signal Processing*, vol. 57, no. 7, pp. 2479-2493, July 2009.

[35] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.

[36] Hung-Wei Chen, Li-Wei Kang, and Chun-Shien Lu, "Dynamic Measurement Rate Allocation for Distributed Compressive Video Sensing," *Proc. IEEE/SPIE Visual Communications and Image Processing (VCIP): special session on Random Projection and Compressive Sensing*, July 2010
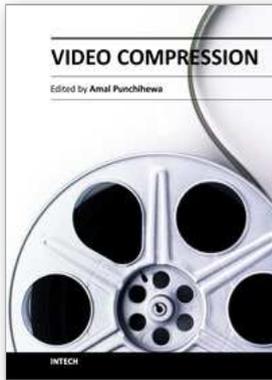
Video Compression

Edited by Dr. Amal Punchihewa

Even though video compression has become a mature field, a lot of research is still ongoing. Indeed, as the quality of the compressed video for a given size or bit rate increases, so does users' level of expectations and their intolerance to artefacts. The development of compression technology has enabled number of applications; key applications in television broadcast field. Compression technology is the basis for digital television. The "Video Compression" book was written for scientists and development engineers. The aim of the book is to showcase the state of the art in the wider field of compression beyond encoder centric approach and to appreciate the need for video quality assurance. It covers compressive video coding, distributed video coding, motion estimation and video quality.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Muhammad Yousuf Baig, Edmund M-K. Lai and Amal Punchihewa (2012). Compressive Video Coding: A Review of the State-Of-The-Art, Video Compression, Dr. Amal Punchihewa (Ed.), ISBN: 978-953-51-0422-3, InTech, Available from: http://www.intechopen.com/books/video-compression/compressive-video-coding-a-review-of-the-state-of-the-art

INTECH
open science | open minds