DISTRIBUTED COMPRESSED VIDEO SENSING

*Thong T. Do†, Yi Chen†, Dzung T. Nguyen†, Nam Nguyen†, Lu Gan‡ and Trac D. Tran† **

† Department of Electrical and Computer Engineering
The Johns Hopkins University
‡School of Engineering and Design
Brunel University, UK

ABSTRACT

This paper proposes a novel framework called Distributed Compressed Video Sensing (DISCOS) – a solution for Distributed Video Coding (DVC) based on the recently emerging Compressed Sensing theory. The DISCOS framework compressively samples each video frame independently at the encoder. However, it recovers video frames jointly at the decoder by exploiting an *interframe sparsity model* and by performing *sparse recovery with side information*. In particular, along with global frame-based measurements, the DISCOS encoder also acquires local block-based measurements for block prediction at the decoder. Our interframe sparsity model mimics state-of-the-art video codecs: the sparsest representation of a block is a linear combination of a few temporal neighboring blocks that are in previously reconstructed frames or in nearby key frames. This model enables a block to be optimally predicted from its local measurements by l_1-minimization. The DISCOS decoder also employs a sparse recovery with side information to jointly reconstruct a frame from its global measurements and its local block-based prediction. Simulation results show that the proposed framework outperforms the baseline compressed sensing-based scheme of intraframe-coding and intraframe-decoding by $8 - 10dB$. Finally, unlike conventional DVC schemes, our DISCOS framework can perform most encoding operations in the analog domain with very low-complexity, making it be a promising candidate for real-time, practical applications where the analog to digital conversion is expensive, e.g., in Terahertz imaging.

Index Terms— distributed video coding, Wyner-Ziv coding, compressed sensing, compressive sensing, sparse recovery with decoder side information, structurally random matrices.

1. INTRODUCTION

Distributed Video Coding (DVC) [1] refers to a special coding scheme that encodes correlated samples (e.g. frames) of a video sequence independently and decode them jointly to obtain higher quality performance. The theoretical foundation of DVC originates from the Slepian-Wolf and Wyner-Ziv (WZ) distributed source coding theorems [2] [3], which assert that dependent random sequences can be coded independently with minimal loss of performance if they are decoded jointly. The DVC framework implies that we can avoid computational and time-consuming operations such as motion estimation or prediction search at the encoder by simply shifting them to the decoder, making DVC very attractive to various distributed applications such as: wireless video cameras, wireless low-power surveillance, video conferencing with mobile devices, visual sensor networks and distributed video streaming.

More recently, Compressed Sensing(CS) theory [4] has become widely popular in the signal processing community. CS theory asserts that a K-sparse (or compressible) signal can be faithfully recovered from only $\mathcal{O}(K \log N)$ incoherent measurements (e.g. random projections) via linear programming, suggesting a significant cost reduction of digital data acquisition. In addition, due to the simplicity of the measurement acquisition at the encoder, the CS framework is a natural fit for distributed applications.

Motivated by the fact that the two aforementioned theories share a common principle of a complexity shift from encoder to decoder, we propose a novel framework called Distributed Compressed Video Sensing (DISCOS), a marriage of the DVC and the CS theory. Our framework not only retains the best features of both theories but also provides additional functions to current state-of-the-art DVC and CS-based coding schemes, introducing a few novel techniques along the way: interframe sparsity model, sparsity-constraint motion estimation and compensation, and sparse recovery with side-information at the receiver.

2. RELATED WORKS AND OUR CONTRIBUTIONS

Conventional DVC systems are often realized by Wyner-Ziv coding, a coding technique that generates parity information to correct errors of the virtual correlation channel between the source signal and the side information. For example, the PRISM framework in [5] is based on block coding mode selection and rate control at the encoder. Each block of transformed coefficients of a frame difference is classified into one of the three modes: not coded, intra-coded or WZ-coded with a set of predefined rates. Cyclic Redundancy Check (CRC) of a WZ block is also sent to help motion estimation and compensation at the decoder. Another practical DVC scheme is proposed in [1] by exploiting frame-based WZ coding and a feedback channel to control the bit-rate. While key frames are intra-encoded by conventional video compression standard such as MPEG/H.26x (with intra-coding mode), the WZ-frames are quantized and encoded using turbo coder and their parity bits are stored in the buffer and gradually transmitted based on the feedback request from the decoder. The additional hash code, a small set of quantized low frequency DCT coefficients of blocks, is also transmitted to help motion estimation at the decoder.

The theory of Distributed Compressed Sensing (DCS) is introduced and analyzed in [6], based on a concept of joint sparsity of a signal ensemble. This work mainly focuses on characterizing the fundamental performance limits of DCS recovery for three modes of jointly sparse signals rather than attempting to design practical algorithms for real-time signals like video sequences. Kalman-Filtered

*This work has been supported in part by the National Science Foundation under Grant CCF-0728893.

Compressed Sensing (KF-CS) is proposed in [7] for reconstructing a time sequence of spatially sparse signals (e.g. video frames). The KF-CS algorithm is based on two assumptions: *(i)* the sparsity pattern of frames' transform coefficients changes slowly over time; and *(ii)* a simple prior model on the temporal dynamics of its support set is available. Unfortunately, both assumptions are quite restrictive and not always true in many practical applications.

2.1. Our Contributions

While current CS-based approaches only focus on modifying the sparse recovery algorithms, our approach aims at a complete design of both measurement acquisition and recovery processes as in the aforementioned practical Wyner-Ziv based DVC schemes. Our main original contributions in this work include

- A novel model of interframe sparsity for video sequences;
- Novel algorithms of sparsity-constraint block prediction, motion estimation and motion compensation;
- Sparse recovery with decoder side information;
- A practical, real-time system design of distributed video coding based on a compressed sensing acquisition method of both local block-based and global frame-based measurements.

3. DISTRIBUTED COMPRESSED VIDEO SENSING

3.1. Architecture

The architecture of our proposed DISCOS framework is depicted in Fig. 1.

1. *Description of the Encoder*

Frames of a video sequence are divided into two categories: key frames (also called I-frames) and non-key frames (also called CS-frames). Key frames are intra-coded by conventional video compression standards such as the MPEG/H.26x (with intra-coding mode) and are sent periodically after a certain number of CS-frames. This is similar to the Group-of-Pictures (GOP) structure in many video codecs. CS-frames are compressively sampled by a common measurement ensemble. Both block-based measurements and frame-based measurements of each CS-frame are rounded to integers and transmit sequentially to the decoder.

2. *Description of the Decoder*

Key frames are obviously decoded by conventional video compression standards. The bitstream of measurements is first inverse-quantized. Block-based measurements, along with preceding and following key frames, are used for generating sparsity-constraint block prediction as described in the Algorithm 1. Instead of using a fixed linear transform (e.g. the block DCT), we use a dictionary of temporal neighboring blocks as the sparsifying matrix for a block in CS-frames. In our opinion, this is by far the sparsest representation of a block in a typical video sequence where temporal correlation dominates spatial correlation. In this algorithm, it is also possible to use only neighboring blocks in previously reconstructed frames rather than those in preceding and following key frames to minimize latency. Unfortunately, this would result in error propagation in subsequent CS frames in the same GOP since there is always a mismatch between the original frame and the reconstructed one at the decoder.

The process of block prediction generates side information at the decoder that is similar to the philosophy of motion estimation at the decoder side in other conventional DVC schemes mentioned above although our l_1-minimization based approach is more powerful because it contains the conventional block matching motion estimation as its special case. The DISCOS decoder regards the block-based prediction frame as the side information to recover the input frame from global measurements by employing Algorithm 2.

3.2. Key Underlying Principles

1. *Block-based measurements vs. frame-based measurements*

Unlike all other existing CS-based schemes in which compressed measurements are taken globally at the frame level, DISCOS acquires a mixture of measurements at both frame level and block level. Our motivation comes from the observation that the block motion model is very efficient for exploiting temporal correlation among frames, serving as the key factor for the success of current video compression standards.

From the perspective of incoherence principle in CS theory, block-based measurements seem to be less efficient than frame-based measurements as the sensing matrix of block-based measurements is block-diagonal while that of frame-based measurements is dense, implying that the former is less incoherent than the latter (in some sparsifying domain). However, by sacrificing a part of incoherence, block-based measurements can preserve *local information* that helps the decoder construct more accurate SI based on the interframe sparsity model and sparsity-constraint block prediction. While conventional CS trades off structure for randomness to get maximal incoherence, our approach attempts to balance between randomness and structure to gain better SI at the decoder.

2. *Interframe sparsity model and sparsity-constraint block prediction*

The interframe sparsity model, which is depicted in the Fig. 2, assumes that a block can be sparsely represented by a linear combination of a few temporal neighboring blocks that were available at the decoder. The neighboring blocks can be the ones in previously reconstructed frames or in nearby key frames (I-frames). While the block prediction algorithm in conventional video coding standards seeks the best-matching block, the sparsity-constraint block prediction algorithm in DISCOS finds the one that can be linearly represented by the fewest number of temporal neighboring blocks, given its compressed measurements. Our sparsity-constraint block prediction scheme is more powerful than the block-matching as it enables a block to be adaptively predicted from the optimal number of neighboring blocks, given its compressed measurements.

3. *Sparse Recovery with Decoder SI*

DISCOS employs a very simple but efficient algorithm of sparse recovery with decoder SI: to substract the measurement vector of an input frame from that of a block-based prediction frame to form a new measurement vector of the prediction error. As soon as the prediction is sufficiently close to the original frame, the prediction error should be sparse (in the spatial domain or in some transform domain such as DCT or DWT) and thus, it can be faithfully recovered from its compressed measurements. The approximation of an input frame is then simply recovered by adding the prediction error to the prediction frame.

4. SIMULATION RESULTS

This section compares the performance between our proposed DISCOS scheme and that of the baseline CS-based scheme of intraframe-coding and intraframe-decoding, which simply recovers frames from its frame-based measurements independently without exploiting temporal correlation among frames.

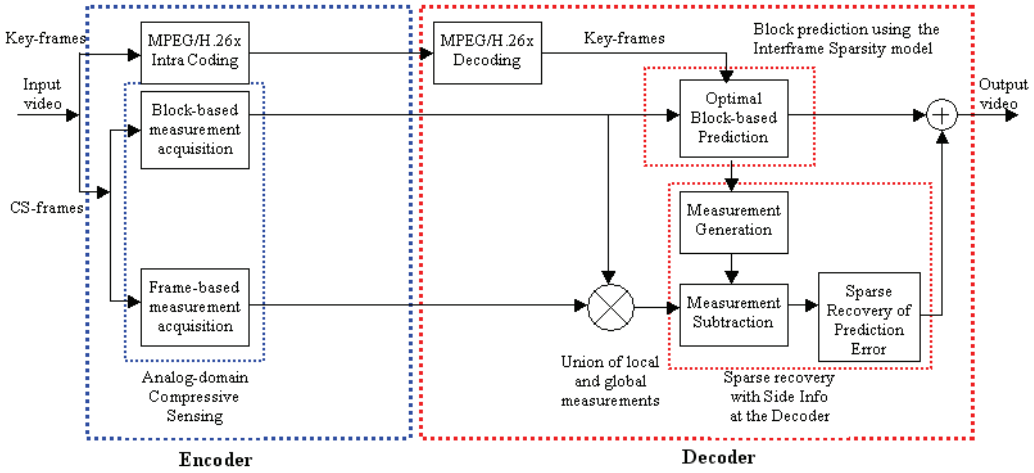


Fig. 1. Architecture of Distributed Compressed Video Sensing.

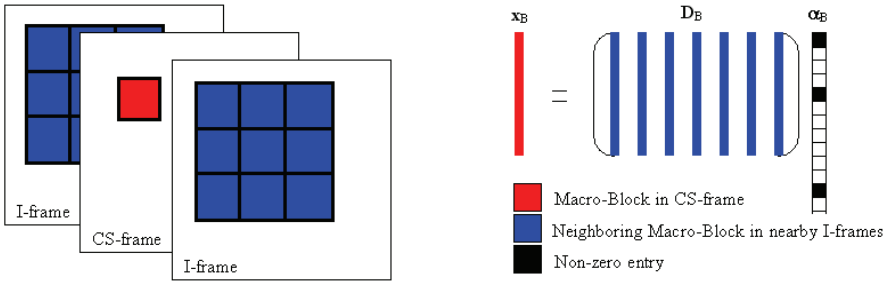


Fig. 2. The interframe sparsity model assumes a (vectorized) block in a CS-frame can be sparsely represented as a linear combination of (vectorized) temporal neighboring blocks in preceding and following key frames.

Input: block-based sensing matrix $\mathbf{\Phi}_B$, compressed measurement vector of an input macro-block (MB) $\mathbf{y}_B$, coordinates of the center of the MB (u, v), preceding and following key frames $\mathbf{I}_1$, $\mathbf{I}_2$, diameter of the correlated areas in two key frames $2s$

Output: l_1-minimization prediction of the input MB

Let $\mathbf{D}_B$ be a matrix whose columns are all vectorized MBs inside the correlated area of the preceding key frame $\mathbf{I}_1(u-s:u+s, v-s:v+s)$ and inside that of the following key frame $\mathbf{I}_2(u-s:u+s, v-s:v+s)$. Solve the following optimization:

$$\hat{\alpha}_B = \text{argmin}\|\alpha_B\|_1 \quad \text{s.t.} \quad \mathbf{y}_B = \mathbf{\Phi}_B\mathbf{D}_B\alpha_B$$

Output: $\hat{\mathbf{x}}_B = \mathbf{D}_B\hat{\alpha}_B$, a prediction of the input MB;

Algorithm 1: Sparsity-Constraint Block Prediction Algorithm

Input: (vectorized) prediction frame $\tilde{\mathbf{x}}$, measurements of an input frame $\mathbf{y}$, its corresponding sensing matrix $\mathbf{\Phi}$ and its sparsifying matrix $\mathbf{\Psi}$

Output: Approximation of the (vectorized) input frame.

$\tilde{\mathbf{y}} = \mathbf{\Phi}\tilde{\mathbf{x}}$ { Measurement vector of a prediction frame}
$\mathbf{z} = \mathbf{y} - \tilde{\mathbf{y}}$ { Measurement vector of a prediction error}
Solve the following optimization:

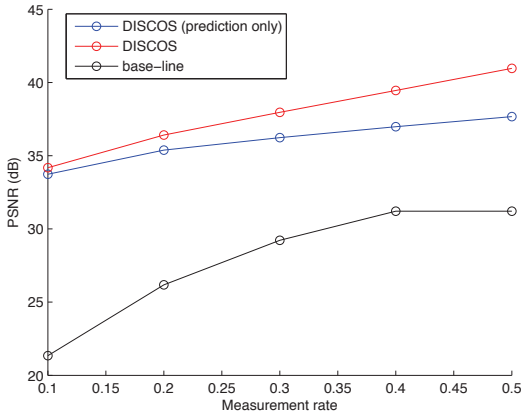
$$\hat{\beta} = \text{argmin}\|\beta\|_1 \quad \text{s.t.} \quad \mathbf{z} = \mathbf{\Phi}\mathbf{\Psi}\beta$$

Output: $\tilde{\mathbf{x}} + \mathbf{\Psi}\hat{\beta}$, an approximation of the (vectorized) input frame

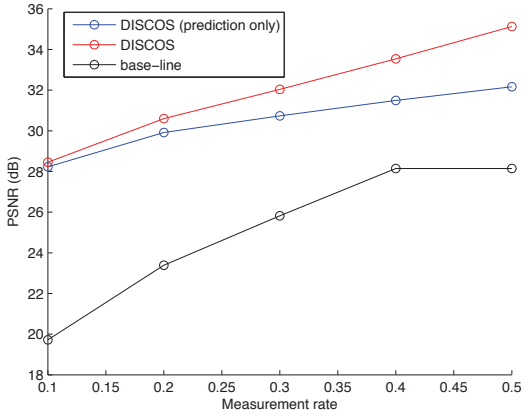
Algorithm 2: Algorithm of Sparse Recovery with Decoder SI

In both schemes, we use the method of Structurally Random Matrices (SRMs) [8] for acquiring measurements. SRM naturally fits for real-time, practical distributed applications as it requires very low-complexity at the encoder, supports fast reconstruction at the decoder and is easy to implement in analog/optical domain. At the decoder-side, Sparsity Adaptive Matching Pursuit (SAMP) [9] reconstruction algorithm is used in sparsity-constraint block prediction (Algorithm 1) to solve the l_1-minimization since it has high reconstruction accuracy while retaining a low-complexity implementation. The GPSR algorithm [10] is used in both the algorithm of sparse recovery with decoder SI (Algorithm 2) and the sparse recovery in the baseline scheme.

The test signals are the first 100 frames of two CIF video sequences: *Foreman* and *CoastGuard*. In the simulation, DISCOS uses the GOP size of 4 and block size of 16×16, the budget of measurements is equally divided into block-based and frame-based measurements. Both schemes use Daubechies wavelet $9/7$ as their spatially sparsifying matrices at the decoder side.



(a)



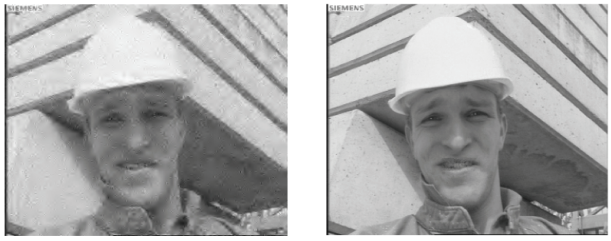
(b)

Fig. 3. Performance comparison between DISCOS and base-line CS: average reconstruction quality of the first 100 frames of (a) *Foreman* sequence; (b) *Coastguard* sequence.

Fig. 3(a) and Fig. 3(b) are performance comparison between the proposed DISCOS framework and the baseline scheme of these two sequences *Foreman* and *CoastGuard*, respectively. The numerical values on the x-axis denotes the percentage of measurements while those on the y-axis represents the average reconstruction quality (PSNR in dB) of *CS-frames*. Visual reconstructions of the frame 41 of the *Foreman* from 25 percent measurements are depicted in Fig. 4. One can clearly see that DISCOS outperforms the base-line scheme by a significant margin in both PSNR and visual reconstruction quality. Even block-based prediction frames that employ local measurements only are far more accurate than reconstructed ones in the base-line.

5. CONCLUSION AND FUTURE WORKS

Our DISCOS framework incorporates successful features of block motion model in conventional video compression standards into a CS-based approach. It is built on four innovations: *(i)* the acquisition of both local and global measurements; *(ii)* the interframe sparsity model; *(iii)* the sparsity-constraint block prediction (motion estimation and motion compensation); and *(iv)* sparse recovery with



(a) PSNR:27.9 dB (b) PSNR:38.7 dB

Fig. 4. Reconstruction of frame 41 from 25% measurements using (a) baseline and (b) DISCOS.

decoder sifde information. Simulation shows that DISCOS outperforms conventional CS scheme by a significant margin $(8 - 10dB)$. A fair performance comparison of DISCOS and other state-of-the-art Wyner-Ziv based schemes [5] [1] is beyond the scope of this paper; it will be presented in the near future. We emphasize that DISCOS is designed for special distributed applications where analog to digital conversion is expensive (the cost of digital acquisition is high). Moreover, to obtain optimal RD performance, it is necessary to design the optimal strategy for local-vs-global measurement allocation, especially when a feedback channel is available to help the encoder in rate control. We leave these open issues for our future works.

6. REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, pp. 71–83, Jan 2005.

[2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on IT*, pp. 471–480, Jul 1973.

[3] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on IT*, vol. 22, pp. 1–10, Jan 1976.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on IT*, vol. 52, pp. 489 – 509, Feb. 2006.

[5] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Trans. on IP*, vol. 16, pp. 2436–2448, Oct 2007.

[6] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," *Submitted to IEEE Trans. on IT*, Jan 2009.

[7] N. Vaswani, "Kalman filtered compressed sensing," *IEEE Intl. Conf. Image Proc.*, 2008.

[8] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Fast and efficient compressive sampling using structurally random matrices," *Submitted*, Feb 2009.

[9] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit for practical compressed sensing," *Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA*, Oct 2008.

[10] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, Dec 2007.