ROBUST VIDEO TRANSMISSION USING LAYERED COMPRESSED 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

We propose a novel Layered Compressed Sensing (CS) approach for robust transmission of video signals over packet loss channels. In our proposed method, the encoder consists of a base layer and an enhancement layer. The base layer is a conventionally encoded bitstream and transmitted without any error protection. The additional enhancement layer is a stream of compressed measurements taken across slices of video signals for error-resilience. The decoder regards the corrupted base layer as the side information (SI) and employs a sparse recovery with SI to recover approximation of lost packets. By exploiting the SI at the decoder, the enhancement layer is required to transmit a minimal amount of compressed measurements for error protection that is only proportional to the amount of lost packets. Simulation results show that both compression efficiency and error-resilience capacity of the proposed scheme are competitive with those of other state-of-the-art robust transmission methods, in which Wyner-Ziv (WZ) coders often generate an enhancement layer. Thanks to the soft-decoding feature of sparse recovery algorithms, our CS-based scheme can avoid the cliff effect that often occurs with other Wyner-Ziv based schemes when the error rate is over the error correction capacity of the channel code. In addition, our result suggests that compressed sensing is actually closer to source coding with decoder side information than to conventional source coding.

***Index Terms*—** Distributed video coding, systematic lossy source/channel coding, Wyner-Ziv coding, side information, compressed sensing, sparse recovery.

1. INTRODUCTION

When video signals are transmitted over channels such as wireless networks or the Internet, they are often protected by some form of Forward Error Correction codes (FEC) to combat potential corruptions on the transmission channels such as fading effect or packet loss. In theory, we can protect the video packets perfectly use FEC by sending a sufficient amount of protection bits to recover lost packets. However, the error rate often varies with time and unknown precisely at the encoding time. When the error rate is over the error correction capacity of the channel code, quality of reconstructed video is rapidly degraded as error concealment techniques alone are insufficient for an acceptable video quality, resulting in the commonly known cliff effect. Although we can avoid the cliff effect by sending a large amount of protection bits corresponding to the worst error rate, this method result in rate inefficiency when the error rate is indeed smaller than the estimation. Thus, the key issue of robust video transmission methods involves a tradeoff between the redundant bitrate used for error protection and the quality of video reconstruction.

Recently, the Compressed Sensing theory [1] [2] has become widely popular in signal processing community. Suppose that $\mathbf{x}$ is a length-N signal. It is said to be K-sparse (or compressible) if $\mathbf{x}$ can be well approximated using $K \ll N$ coefficients under some linear transform (e.g. the DCT or the DWT). According to the CS theory, such a signal can be acquired through the following linear random projections:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{e},$$

where $\mathbf{y}$ is the measurement vector with $M \ll N$, Φ represents an $M \times N$ random projection matrix and $\mathbf{e}$ is the acquisition noise. The CS theory asserts that $\mathbf{x}$ can be faithfully recovered from only $M = \mathcal{O}(K \log N)$ measurements.

More remarkably, in [3], Candès and Tao shown the connection between the Compressed Sensing and error correction coding theory. They show that noisy observation of random linear codes can be successfully decoded using linear programming. Motivated by this theoretical result, we propose a CS-based approach for robust transmission of video signals over packet loss channels. Our method can achieve gradual degradation of the reconstructed video quality while keeping the rate efficiency competitive with other state-of-the-art systems. From our best knowledge, this work is one of the first efforts to realize a practical robust communication system for video signals based on principles of the CS theory.

2. RELATED WORKS AND OUR CONTRIBUTIONS

2.1. WZ-based Approaches

Wyner-Ziv coding refers to the coding technique that generates parity information to correct errors of the virtual correlation channel between the source signal and the side information. Many robust video transmission systems are based on the Wyner-Ziv coding technique as the parity information can also be used to correct the errors introduced by physical channels as well. In [4], a Systematic Lossy Error Protection (SLEP) is proposed based on the idea of systematic lossy source channel coding [5]. The SLEP consists of a base layer that is compressed independently by some conventional video compression standards such as MPEG/H.26x and a Wyner-Ziv enhancement layer. The Wyner-Ziv layer is regarded as the coarser second description of the base layer. The decoder uses the corrupted base layer as its SI and jointly recover the signal from the Wyner-Ziv bit

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

978-1-4244-4464-9/09/$25.00

stream and the SI. By using a coarser representation at the Wyner-Ziv layer, the SLEP system can provide more robustness than the FEC at the same bit rate of the enhancement layer. In [6] a layered Wyner-Ziv video codec is proposed using the DCT, nested scalar quantization and irregular LDPC code based Slepian-Wolf coding. As the Wyner-Ziv layer generated from the original video sequence without using the base layer, this codec is shown to be more robust than conventional codec H.26x with Fine Granular Scalability (FGS) coding. In [7], the scalable PRISM codec is proposed based on the principle of distributed source coding, i.e. encoding frames independently and decoding them jointly. As the PRISM does not use predictive coding, there is no interframe error propagation. Experiments in [7] report a substantial gain for video multicast over packet loss channels compared to standard codecs.

2.2. The CS-based Approach

Theoretical result in [3] asserts that a corrupted observation of a random linear code can be decoded successfully via linear programming, i.e. it is possible to recover a message word $\mathbf{u} \in \mathbf{R}^m$ from its corrupted code word $\mathbf{v} \in \mathbf{R}^n$:

$$\mathbf{v} = \mathbf{A} * \mathbf{u} + \alpha$$

where $\mathbf{A}$ is $n \times m$ (random linear) coding matrix ($n > m$), α is the unknown error vector, provided that the number of nonzero (or sparsity number) of the error vector α is on the order of $m/\log m$. In fact, $\mathbf{u}$ is shown to be the solution of the following optimization:

$$\min_{\mathbf{u} \in \mathbf{R}^m} \|\mathbf{v} - \mathbf{A}\mathbf{u}\|_1 \tag{1}$$

Our proposed Layered Compressed Sensing codec (LACOS) is a marriage of the above principle and the model of systematic lossy source-channel coding [5]. In LACOS, the coding matrix $\mathbf{A}$ has the form $[\mathbf{I}|\mathbf{\Phi}]$, where $\mathbf{I}$ is $m \times m$ identity matrix corresponding to the systematic part and $\mathbf{\Phi}$ is $(n-m) \times m$ random linear matrix corresponding to the parity part. Instead of directly solving the optimization (1), LACOS employs a simpler but efficient algorithm of sparse recovery with receiver SI to recover lost packets. To be suitable for real-time applications, LACOS also takes compressed measurements across slices of macro-blocks rather than across an entire frame as this would reduce significantly the complexity at both encoder and decoder.

WZ-based approaches are all based on some well-known channel coding over a finite field (e.g. Reed-Solomon or LDPC). The main drawback of coding on a finite field is that the decoder often requires to receive a sufficient number of parity bits before it can start decoding. However, the amount of parity bits necessary for successful decoding depends on the rate of packet loss which might vary with time and is often unknown at the encoding time. Traditional channel coding does not allow the decoder to exploit partial information from a subset of its parity bits. In other words, traditional channel coding employs "get all or nothing" principle. On the contrary, our CS-based approach enables the decoder to obtain a coarse signal reconstruction from an arbitrary subset of compressed measurements by employing *a sparse recovery algorithm with decoder SI*. The less parity measurements the decoder receives, the worse reconstructed signal is but there is no breaking point. In principle, the proposed framework can also be viewed as a novel Wyner-Ziv coding technique over the *real* field.

3. LAYERED COMPRESSED SENSING CODEC

The proposed architecture is illustrated in the Fig. 1

3.1. Description of the Encoder

The encoder consists of a base layer and an enhancement layer. The base layer is conventionally encoded by some video compression standard MPEG/H.26x and is transmitted over error-prone channels without any forward error correcting codes. A frame of prediction error is divided into slices of macro-blocks and each of which is packetized and transmitted independently. The enhancement layer takes compressed measurements across these slices as depicted in the Fig. 2. The stream of quantized compressed measurements, along with side information of motion vectors and mode decision from the base layer, is entropy-coded and transmitted to the decoder.

3.2. Description of the Decoder

The corrupted base layer is regarded as the SI at the decoder. LACOS employs a simple but very efficient algorithm of sparse recovery with decoder SI that is to subtract the measurement vector of cross-slice blocks from that of its corrupted version to form a new measurement vector of block loss. Note that the vector of block loss is zero every where except location of lost blocks. In addition, lost block locations also contain many zero entries because they contain quantized transform coefficients of the prediction error. Thus, nonzeros entries in lost blocks can be faithfully recovered from a relatively small number of measurements via linear programming or greedy-pursuit based sparse recovery. The stream of reconstructed error prediction, along with side information of motion vectors and mode decision, is normally decoded by conventional video compression standards.

Our current framework assumes that the enhancement layer can be transmitted without packet-loss effect. In fact, the raw measurement stream is also well-tolerant with packet loss as measurements are roughly of equal importance, i.e. quality of reconstructed signal is only dependent on the number of measurements received (i.e. independent of which measurements are received), provided that these measurements are generated from an incoherent domain. Thus, the enhancement layer can also be robust to packet-loss if quantized compressed measurements (rather than their entropy coded version) are transmitted to the decoder, in exchange for a small decrease of rate efficiency.

4. SIMULATION RESULTS

We compare the proposed LACOS scheme, traditional FEC method using Reed-Solomon (RS) channel code and a simple error concealment method that simply replaces lost blocks of prediction error by zero entries. This concealment method is equivalent to replacing lost blocks of a frame by their motion compensated blocks from the previously reconstructed frame. In general, this is the best thing we can do without additional side information from the encoder. In the FEC scheme, when failure decoding occurs at the RS-decoder, we use this method of concealment as a replacement.

In all these schemes, the base layer uses the same conventional video compression standard MPEG2 and uses the GOP of 4. The enhancement layer divides frame of quantized, transformed prediction error into slices and each of which contains a row of macro-blocks. Here, we assume that each packet contain one slice and thus, our framework deals with the effect of slice loss. For simplicity, we assume that there is no slice loss in key frames (I-frames). For non-key frames, the pattern of slice loss is generated in a uniformly random fashion.

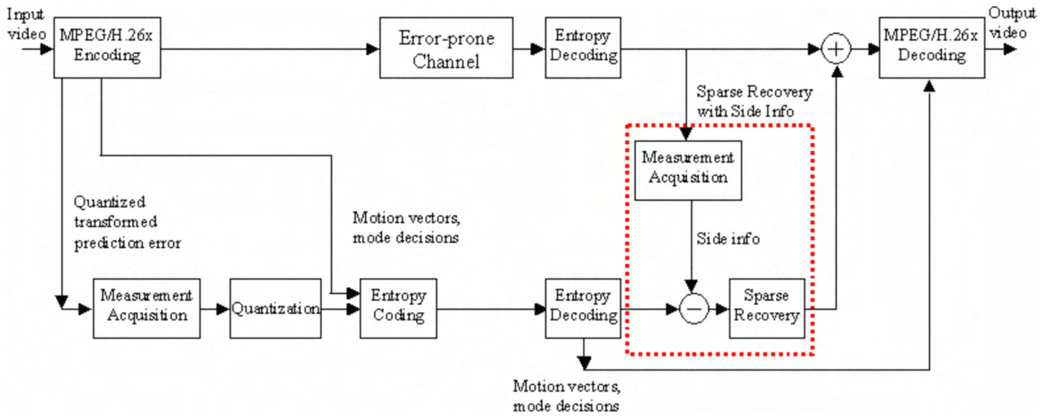


Fig. 1. The architecture of Layered Compressed Sensing Codec.

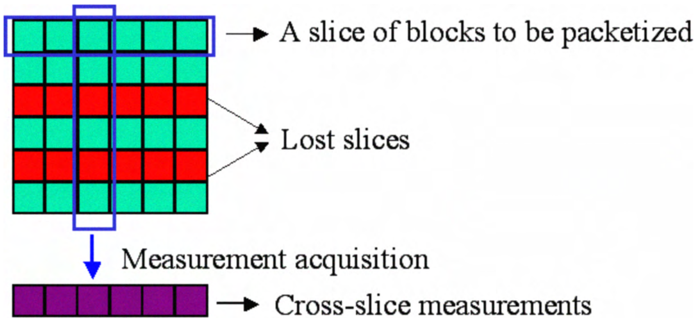


Fig. 2. Acquisition of cross-slice measurements.

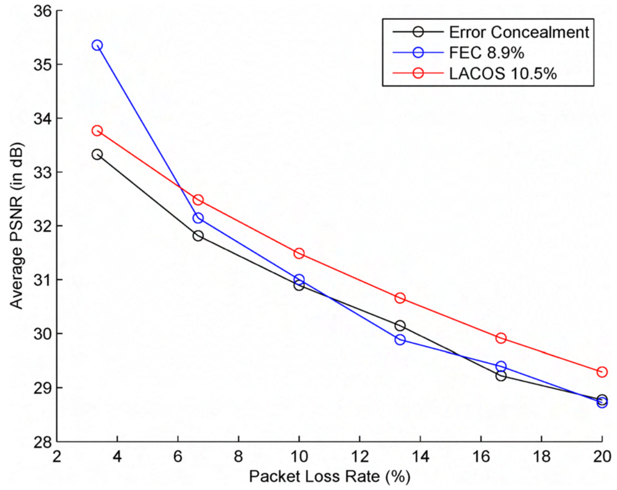
In LACOS, we use the method of Structurally Random Matrices (SRMs) [8] for acquiring measurements. SRM is a product of three matrices: $\mathbf{DFR}$, where $\mathbf{R}$ is a random permutation matrix or a diagonal matrix of Bernoully random variables(± 1); $\mathbf{F}$ can be a Walsh-Hadamard (WH) fast transform ; $\mathbf{D}$ is a random subset of rows of an identity matrix. More remarkably, SRM can be implicitly implemented as serial operators: (i) randomly permuting entries (or randomly flipping signs of entries)of the source signal; (ii) WH transforming the randomized signal and finally (iii) randomly subsampling the transform coefficients. The SRM naturally fits for real-time, practical applications as it requires very low-complexity at the encoder, supports fast reconstruction at the decoder. Note that there is no need to send the measurement matrix over the channel as it always can be regenerated at the decoder using the same pseudorandom seeds at the encoder. The Sparsity Adaptive Matching Pursuit (SAMP) [9] is employed for sparse recovery with decoder SI as it provide high accuracy while keeping a low-complexity computation. In addition, a uniform quantization with unit step-size and Huffman code are employed as its quantizer and entropy-coder, respectively. The number of measurements is chosen appropriately so that the bitrate of the CS-based enhancement layer is roughly similar to the amount of RS-parity bits in the baseline FEC scheme.

In experiments, test signals are the first 100 frames of the SIF-resolution sequence *Football* and of the CIF-resolution sequence *Mobile*. With *Football*, the base layer stream is encoded at the bit rate 2.97 Mbps (30 frames per second). The parity-bit rate and the bit rate of measurement stream are 8.9% (264 Kbps) and 10.6% (314 Kbps) of the base layer rate, respectively. With *Mobile*, the base layer stream is encoded at the bit rate 3.79 Mbps (30 frames per second) while the parity-bit rate and bit rate of measurement stream are 15.2% (577 Kbps) and 14.9% (567 Kbps) of the base layer rate, respectively. The performance curves of regarding methods are depicted in the Fig. 3 and Fig. 4 . The numerical values on the x-axis denotes the amount of packet/slice loss while those on the y-axis represents the average reconstruction quality (PSNR in dB) of *corrupted frames* only.

One can clearly see that the concealment method results in the worst performance as it requires no additional bit for packet protection. The conventional FEC guarantees the best performance when the amount of packet loss is not larger than error correction capacity of RS code. However, when the amount of packet loss increases and over the error correction capacity of RS code, quality of reconstructed signal from FEC scheme is dropped quickly (cliff effect) while those from the proposed method keeps gradual degraded. Thanks to the soft-decoding feature of sparse recovery algorithm, the proposed framework can get rid of the cliff effect quite efficiently.

5. CONCLUSION

This paper proposes a novel and practical system design of error-resilient video transmission over packet loss channels. Its motivation comes from the corresponding theoretical result that noisy random linear code can be correctly decoded via linear programming or greedy pursuit algorithms. Taking into account real-time oper-

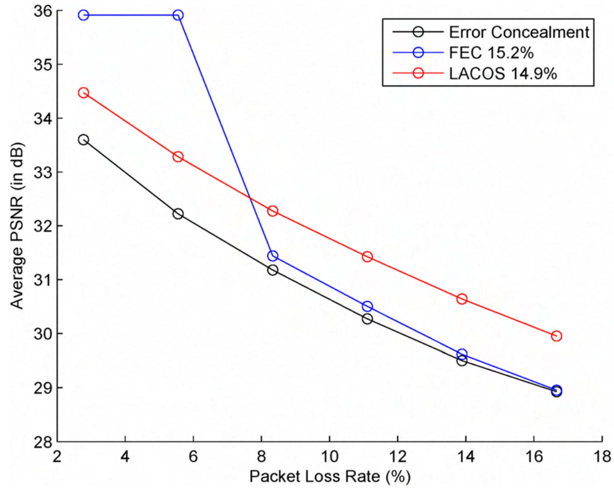


(a)

(b) PSNR:29 dB

(c) PSNR:30.7 dB

Fig. 3. Performance comparison: LACOS, FEC and Error Concealment of SIF-resolution *Football* sequence: (a) PSNR vs. packet loss rate; Reconstruction of the frame 27 with 13.3% packet loss using (b) FEC ; (c) LACOS.



(a)

(b) PSNR:31.3 dB

(c) PSNR:33 dB

Fig. 4. Performance comparison: LACOS, FEC and Error Concealment of CIF-resolution *Mobile* sequence: (a) PSNR vs. packet loss rate; Reconstruction of the frame 34 with 13.9% packet loss using (b) FEC ; (c) LACOS.

ation constraint of a practical communication system, we combine practical features of current state-of-the-art WZ-based design with compressive sensing principles. The proposed framework, LACOS, acquires compressed measurements across slices/packets for error protection and then at the decoder-side employs a novel sparse recovery algorithm with receiver SI to recover a coarse approximation of lost packets. As the reconstruction algorithm is iteratively refinement-based, it provides a flexible trade-off between computational complexity and quality of a reconstructed signal. Simulation results verify that thanks to the soft-decoding feature of sparse recovery algorithm, LACOS obtains a higher robustness to packet loss effect in exchange for a little amount of rate loss and hence, efficiently getting rid of the cliff effect. The amount of rate efficiency loss might be minimized by employing a more complicated sparse recovery algorithm at the decoder-side, e.g. exploiting a statistical structure of nonzero transform coefficients of prediction error [10]. We leave this issue for our future work.

6. REFERENCES

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

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, pp. 1289 – 1306, Apr. 2006.

[3] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[4] S. Rane, P. Baccichet, and B. Girod, "Systematic lossy error protection of video signals," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1347 1360, Oct. 2008.

[5] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Transactions on Information Theory*, vol. 44, pp. 564 579, Mar. 1998.

[6] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 15, pp. 3791 3803, Dec. 2006.

[7] M. Tagliasacchi, A. Majumdar, K. Ramchandran, and S. Tubaro, "Robust wireless video multicast based on a distributed source coding approach," *The Journal of Signal Processing, ScienceDirect*, vol. 86, pp. 3196 3211, 2006.

[8] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Fast and efficient compressive sampling using structurally random matrices," *Submitted*, 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] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Preprint*, 2008.