Wyner-Ziv Coding for Video: Applications to Compression and Error Resilience

Anne Aaron, Shantanu Rane, Rui Zhang and Bernd Girod
Information Systems Laboratory,
Stanford University, Stanford, CA 94305
{amaaron,srane,rui,bgirod}@stanford.edu

Abstract

In this paper we consider two separate applications of Wyner-Ziv coding of motion video. The Wyner-Ziv Theorem on source coding with side information available only at the decoder suggests that an asymmetric video codec, where individual frames are encoded separately, but decoded conditionally (given temporally adjacent frames) could achieve efficiency comparable to current interframe video compression systems. We report first results on a Wyner-Ziv coding scheme for motion video that uses intraframe encoding, but interframe decoding. Secondly, we apply Wyner-Ziv coding to the transmission of compressed video over an error-prone channel. We use Wyner-Ziv Coding to generate a supplementary bitstream which contains a coarsely quantized representation of the transmitted video signal. Using the conventionally decoded, error-concealed video signal as side information to decode the Wyner-Ziv bits, the transmission errors in the decoded video waveform are corrected up to a certain residual distortion, significantly improving the visual quality of the decoded video.

1 Introduction

Lossy compression exploiting side information at the decoder is only in its infancy today. Information theoretic results [1, 2] suggest the existence of efficient codes, but a general method to design practical codes is not known. In this paper we present first results on two separate applications of Wyner-Ziv coding for video waveforms.

First, we propose a highly asymmetric video compression system where the encoder is very simple but the decoder can be complex. Current video compression standards perform interframe predictive coding to exploit the similarities among successive frames. Since predictive coding makes use of motion estimation, the video encoder is typically 5 to 10 times more complex than the decoder. This asymmetry is desirable for broadcasting or for streaming video-on-demand systems where video is compressed once and decoded many times. However, some future systems, such as mobile video cameras may require the dual scenario. Compression must be implemented at the camera where memory and computation are scarce. For this type of system we need a low-complexity encoder, possibly at the expense of a high-complexity decoder, that nevertheless compresses efficiently. To achieve low-complexity encoding, we propose an asymmetric video

1068-0314/03 $17.00

compression scheme, which uses Wyner-Ziv coding, where individual frames are encoded independently (*intraframe encoding*) but decoded conditionally (*interframe decoding*)

In the second thread of this paper, we use the systematic lossy source-channel framework, and show first results on the application of Wyner-Ziv coding for error resilient MPEG broadcast. A video sequence is compressed using MPEG, and transmitted unprotected over a lossy channel. A conventional MPEG decoder performs error-concealment on the received error-prone sequence, but the errors in some portions of the signal are unacceptably large. To correct these remaining errors, up to a certain minimum distortion, we use Wyner-Ziv coding to generate a supplementary bitstream which represents a coarsely quantized version of the transmitted video sequence. A turbo decoder optimally combines the coarse representation and the error-concealed finer side information to output an improved video signal.

2 Related Work

Two results from information theory suggest that an intraframe encoder - interframe decoder system can come close to the efficiency of an interframe encoder-decoder system. Consider two statistically dependent discrete signals, X and Y, which are compressed using two independent encoders but are decoded by a joint decoder. The Slepian-Wolf Theorem on distributed source coding states that even if the encoders are independent, the achievable rate region for probability of decoding error to approach zero is $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$ and $R_x + R_y \geq H(X,Y)$ [1]. The counterpart of this theorem for lossy source coding is Wyner and Ziv's work on source coding with side information [2]. Let X and Y be statistically dependent Gaussian random processes, and let Y be known as side information for encoding X. Wyner and Ziv showed that the conditional Rate-MSE Distortion function for X is the same whether the side information Y is available only at the decoder, or both at the encoder and the decoder. We refer to lossless distributed source coding as Slepian-Wolf coding and lossy source coding with side information at the decoder as Wyner-Ziv coding.

It has only been recently that practical coding techniques for Slepian-Wolf and Wyner-Ziv coding have been studied. Pradhan and Ramchandran presented a practical framework based on sending the syndrome of the codeword coset to compress the source [3]. Since then similar concepts have been extended to more advanced channel codes. Garcia-Frias and Zhao [4][5], Bajcsy and Mitran [6][7] and Aaron and Girod [8] showed that using turbo codes for compression can come close to the Slepian-Wolf bound. Liveris et al. argued that low-density parity-check codes, another form of iterative channel coding, are also suitable for this problem [9]. In [10], Jagmohan et al. suggest that Wyner-Ziv codes could be used to prevent prediction mismatch or drift in video systems but do not present an actual implementation.

Pradhan and Ramchandran applied their syndrome idea to a system where a digital stream provides enhancement to a noisy analog image transmission [11]. The digital stream contains the syndromes representing the codewords of the wavelet coefficients for the original image, and the syndromes are decoded using the analog signal as side information. Similarly, Liveris et al. used turbo codes to encode the pixels of an image which has a noisy version at the decoder [12].

In this paper we apply Wyner-Ziv coding to compress a real-world video signal. We take X as the even frames and Y as the odd frames of the sequence. X is compressed by an intraframe encoder that does not know Y. The compressed stream is sent to a decoder which uses Y as side information to conditionally decode X. Note that we do not force a given correlation between X and Y but instead use the inherent temporal similarities between adjacent frames of a video sequence.

The Wyner-Ziv problem is closely related to the problem of systematic lossy source-channel coding [13]. In this configuration, an analog source X is transmitted over an analog Channel A without coding. A second encoded version of X is sent over a digital channel D as enhancement information. The noisy version Y of the original serves as side information to decode the output of Channel D and produce the enhanced version Y^*. The term "systematic coding" has been introduced in extension of systematic error-correcting channel codes to refer to a partially uncoded transmission. Shamai, Verdu, and Zamir establish information theoretic bounds and condition for optimality of such a configuration in [13].

In this paper, we apply the "systematic coding" framework to error resilient MPEG video broadcast. In particular, we show that a decoder having access to a Wyner-Ziv coded version of the transmitted video sequence, in addition to the error-concealed conventionally decoded sequence as side information, can provide an output with superior visual quality and average PSNR, in the presence of channel errors.

The remainder of this paper is organized as follows. In Section 3, we describe the building blocks of our Wyner-Ziv video codec and describe how it is used for the two proposed applications. In Section 4, we compare the performance of the intraframe encoder- interframe decoder system to that of a conventional standard H263+ video coder. We also report test cases to demonstrate the error-resilience capability of our lossy forward error protection scheme.

3 Wyner-Ziv Video Codec

The Wyner-Ziv video codec is composed of an inner turbo code-based Slepian-Wolf codec and an outer quantization-reconstruction pair. The side information available at the decoder is used at the turbo decoder and the reconstruction block of the receiver. We discuss in more detail below how this basic structure is utilized in the two proposed video applications.

3.1 Intraframe video coding with interframe decoding

We propose an intraframe encoder and interframe decoder system for video compression as shown in Fig. 1. Let $X_1, X_2, ..., X_N$ be the frames of a video sequence. The odd frames, X_{2i+1}, where $i \in \{0, 1, ..., \frac{N-1}{2}\}$, are the key frames which are available as side information at the decoder. To simplify the problem, we do not consider the compression of the key frames and assume they are known perfectly at the decoder. Each even frame, X_{2i}, is encoded independent of the key frames and the other even frames.

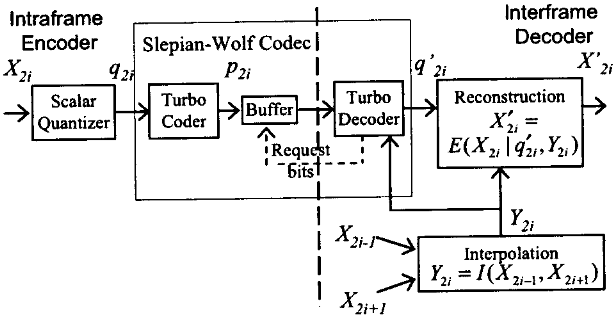


Figure 1: Wyner-Ziv video codec with intraframe encoding and interframe decoding.

X_{2i} is encoded as follows: First, we scan the frame row by row and quantize each pixel value using use a uniform scalar quantizer with 2^M levels to form the quantized symbol stream q_{2i}. For a given frame, we take the symbols and form a long symbol block which is then sent to the Slepian-Wolf encoder. The Slepian-Wolf coder is implemented using a rate compatible punctured turbo (RCPT) code [14]. The RCPT, combined with feedback, provides rate flexibility which is essential in adapting to the changing statistics between the side information and the frame to be encoded. q_{2i} is fed into the two constituent convolutional encoders of a turbo encoder. Before passing the symbols to the second convolutional encoder, interleaving is performed on the symbol level. The parity bits, p_{2i}, produced by the turbo encoder are stored in a buffer. The buffer transmits a subset of these parity bits to the decoder upon request.

For each frame X_{2i}, the decoder takes the adjacent key frames $X_{2i-1}$and $X_{2i+1}$and performs temporal interpolation $Y_{2i} = I(X_{2i-1}, X_{2i+1})$ to form a good estimate of X_{2i}. In our system we implement two types of interpolation techniques. The first technique, Average Interpolation, takes the adjacent frames and performs a simple averaging of the pixels to generate the interpolated frame. In the second technique we perform motion compensated interpolation by assuming symmetric motion vectors (SMV) between the current frame and the two adjacent frames. To be able to use Y_{2i} as side information at the turbo decoder and the reconstruction block, we need to model the statistical dependency between Y_{2i} and X_{2i}. For the two interpolation techniques we implemented, we observed that if we take a pixel from the current frame and subtract from it the corresponding side information, the resulting statistics is very close to that of a Laplacian random variable.

The turbo decoder uses the side information Y_{2i} and the received subset of p_{2i} to form the decoded symbol stream q'_{2i}. If the decoder cannot reliably decode the symbols, it requests additional parity bits from the encoder buffer through feedback. The request and decode process is repeated until an acceptable probability of symbol error is guaranteed. By using the side information, the decoder needs to request $k \leq M$ bits to decode which of the 2^M bins a pixel belongs to and so compression is achieved.

After the receiver decodes q'_{2i} it calculates a reconstruction of the frame X'_{2i} where $X'_{2i} = E(X_{2i}|q'_{2i}, Y_{2i})$. With this reconstruction function, if the side information is within the reconstructed bin then the reconstructed pixel will take the value of the side information. If the side information is outside the bin, the function clips the reconstruction

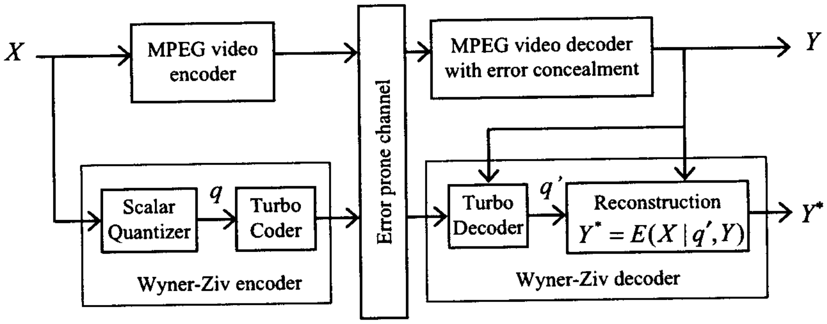


Figure 2: Wyner-Ziv bitstream uses decoded error concealed video waveform as side information in a systematic lossy source-channel setup.

towards the boundary of the bin closest to the side information. This kind of reconstruction function has the advantage of limiting the magnitude of the reconstruction distortion to a maximum value, determined by the quantizer coarseness. Perceptually, this property is desirable since it eliminates the large positive or negative errors which may be very annoying to the viewer.

In areas where the side information is not close to the original signal (i.e. high motion frames, occlusions), the reconstruction scheme can only rely on the quantized symbol for reconstruction and quantizes towards the bin boundary. Since the quantization is coarse, this could lead to contouring which is visually unpleasant. To remedy this we perform subtractive dithering by shifting the quantizer partitions for every pixel using a pseudo-random pattern. This leads to better subjective quality in the reconstruction.

3.2 Lossy forward error protection for digital video broadcast

We now propose a systematic lossy source-channel coding scheme to mitigate the effect of channel errors in digital video broadcast. This scheme is shown in Fig. 2, using MPEG video compression as an example. At the transmitter, the input video signal X is compressed independently by an MPEG video coder and a Wyner-Ziv coder. Since the MPEG video bitstream is generated without consideration of the error resilience provided by the Wyner-Ziv coder, we refer to the overall scheme as systematic source-channel coding. The video signal compressed by MPEG and transmitted (unprotected) over an error-prone channel constitutes the systematic portion of the transmission which is augmented by the Wyner-Ziv bitstream. As before, the Wyner-Ziv coder consists of a coarse scalar quantizer followed by a turbo encoder.

At the receiver, the MPEG bitstream is decoded and possible transmission errors are concealed, resulting in the decoded video Y. Despite concealment, Y contains some portions of the signal that are degraded by unacceptably large errors. These errors are corrected, up to a residual distortion, by the Wyner-Ziv decoder. The Wyner-Ziv code can be thought of as a second, independent description of the input video Y with coarser quantization. Without transmission errors, this description is fully redundant, i.e., it can be regenerated bit-by-bit at the decoder, using the decoded video Y. But with transmission errors, Wyner-Ziv bits must be sent to allow an error-free reconstruction of the coarser second description, employing the decoded video signal Y as side information.

The error correction capability of the turbo coder can be simultaneously used to protect the Wyner-Ziv bits against transmission errors. The coarser second description and finer, but error-prone side information Y are optimally combined to yield an improved decoded video signal $Y^* = E(X|q', Y)$. In portions where the waveform Y is not affected by transmission errors, Y^* will be essentially identical to Y. However, in portions of the waveform where Y is substantially degraded by transmission errors, the second coarser representation transmitted at very low bit-rate in the Wyner-Ziv bit-stream limits the maximum degradation that can occur.

4 Results

4.1 Intraframe video coding with interframe decoding

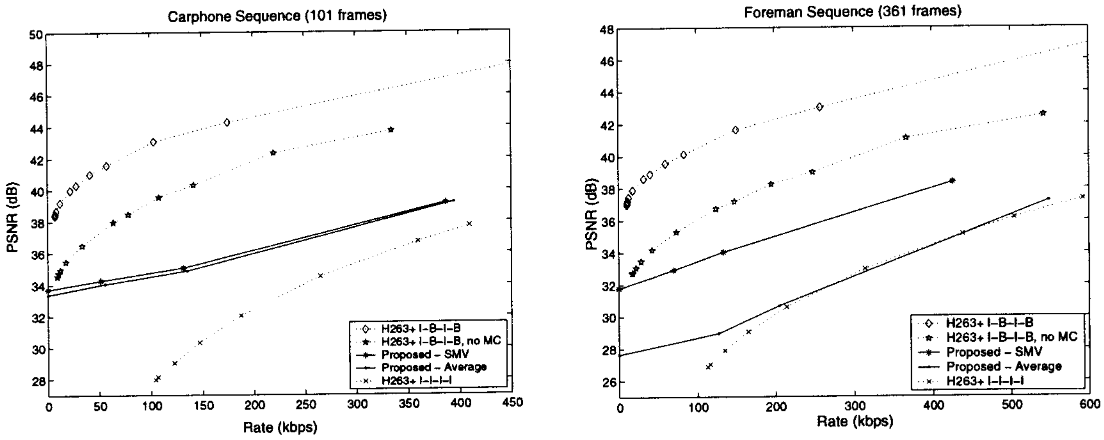
We implemented the system proposed in Section 3.1 and assessed the performance on sample QCIF video sequences. To change the rate (and correspondingly, the distortion) we varied the number of quantization levels, where $2^M \in \{2, 4, 16\}$. For every even frame of the sequence, we gathered the quantized symbols to form an input block of length $L = 144 \times 176 = 25344$.

The turbo encoder was composed of two constituent convolutional encoders of rate $\frac{4}{5}$, identical to those used in [8]. To achieve the rate compatibility for the turbo code, we devised an embedded puncturing scheme, with a puncturing pattern period of 8 parity bits. The simulation set-up assumed ideal error detection at the decoder - we assumed that the decoder can determine whether the current symbol error rate, P_e, is greater than or less than 10^{-3}. If $P_e \leq 10^{-3}$ it requests for additional parity bits. In practical systems, the error rate can be estimated by jointly observing the statistics of the decoded stream and the convergence of the turbo decoder. We implemented Average and SMV interpolation.

We compared the Rate-PSNR performance of our system to three set-ups of the H263+:

1. Intraframe (*I-I-I-I*) - The even frames are encoded as *I* frames.
2. Interframe, no motion compensation (*I-B-I-B, No MC*)- The even frames are encoded as *B* frames (predicted from the previous and next frame) but the motion vectors are set to zero. For fair comparison, we assume that the key frames are perfectly reconstructed at the decoder.
3. Interframe, with motion compensation (*I-B-I-B*) - Same as Set-up 2 but we allow motion compensation.

The results for the *Carphone* and *Foreman* QCIF sequences are shown in Fig. 3. For the plots, we only count the rate and distortion of the luminance of the even frames and consider the even frame rate as 15 frames per second. The zero-rate point in our scheme corresponds to using the interpolated frame as the decoded frame.



(a) (b)

Figure 3: Rate vs. PSNR for (a) *Carphone* and (b) *Foreman*

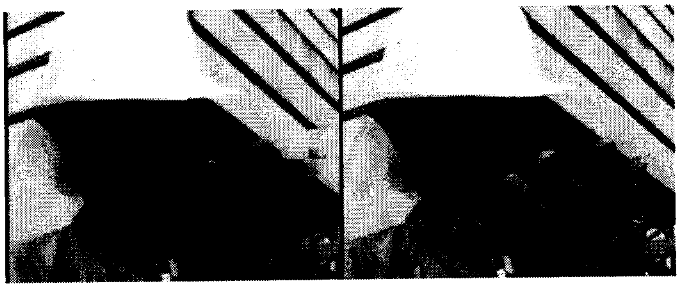
As seen in Fig. 3(a), the interpolation scheme does not significantly change the performance for the *Carphone* sequence. This is due to the fact that most of the new information in the sequence is caused by the changing scenery in the car window and not by high motion. On the other hand, for the *Foreman* sequence in Fig. 3(a), using SMV interpolation gives 3 to 4 dB improvement over Average interpolation. For this sequence, simple averaging was not effective since there was high motion throughout the frame.

With good interpolation at the decoder, our system performs much better than H263+ intraframe coding. For the *Carphone* sequence, the gain compared to H263+ intraframe coding ranged from 2 to 6 dB. For the *Foreman* sequence with SMV interpolation, the gain above intraframe coding was about 4 to 7 dB.

As expected the performance of our Wyner-Ziv codec is below that of H263+ interframe coding. For *Carphone*, the gap from the corresponding interframe plots ranges from 1 to 8 dB, with a smaller gap in lower bit rates. For *Foreman* with SMV, our system performance is about 5 to 7 dB lower than interframe coding. This is partly due to the fact that the H263+ coder exploits both the spatial and temporal redundancy in the signal. For our codec, the spatial correlation has not yet been incorporated into the decoding process.

Even if a sophisticated MC interpolation scheme is implemented, the content of the video may be such that it is difficult to have a good estimate of the current frame from the adjacent frames. In Fig. 4, we see how our coding scheme can fix the MC-interpolation artifacts in cases of occlusions and high motion. As we can see, the encoding sharpens the image and closely reconstructs the hand even if the interpolation is bad. The dithering of the quantizer also improves the visual quality in the areas where motion compensation fails and coarse quantization dominates. Comparing this sequence to that of H263+ intraframe coding with the same sequence bit rate, we observe that the intraframe decoded sequence has obvious blocking artifacts which are not present in our system.

One artifact introduced by our scheme is the presence of residual errors from the Slepian-Wolf decoder which result in isolated blinking pixels at random locations or clustered error specks in a part of the image where the side information is not reliable.



(a) Interpolated frame (b) 16-level encoded

Figure 4: Interpolated and Wyner-Ziv encoded frame from *Foreman*

In our simulations we fixed the maximum error to be less than 10^{-3} or about 25 pixels per frame. Determining a visually acceptable error rate is left to future investigation.

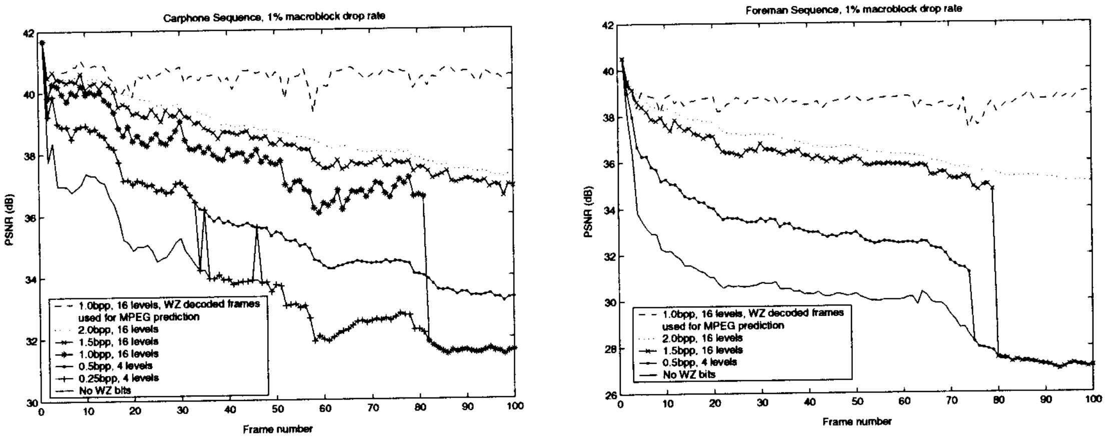
4.2 Lossy forward error protection for digital video broadcast

We implemented the scheme proposed in Section 3.2 for standard MPEG-compressed CIF sequences encoded at 1 Mbps. We simulated a random loss so that 1% of the macroblocks are lost. The decoder performs previous frame error-concealment on the lost macroblock to generate the sequence Y, which serves as side information for the Wyner-Ziv decoder.

The CIF video format produces for every frame a symbol block of length $288 \times 352 = 101376$ which is the input to the turbo encoder. To vary the level of error protection, we change the number of quantization levels ($2^M \in \{4, 16\}$) and control the amount of parity bits sent to the decoder. Note that for this application there is no feedback in the system; so for a given simulation set-up the Wyner-Ziv transmission rate is fixed.

As explained in Section 3.2, the Wyner-Ziv decoder optimally combines the side information Y, and the Wyner-Ziv bits to generate a signal Y^* of improved visual quality. Figs. 5(a) and (b) show that our scheme limits the error propagation caused by imperfect concealment of channel errors, and achieves acceptable PSNR levels. For low rates of the Wyner-Ziv bitstream, the error protection capability breaks down when the errors are too severe. As expected, the average PSNR is maximum when the rate of the Wyner-Ziv bitstream is maximum (2 bpp in our simulations). We note, as before, that this rate is high at present because our pixel domain Wyner-Ziv coder does not exploit any spatial correlation in the input video sequence. Fig. 6 shows that the reconstructed frame quality is improved significantly by using the Wyner-Ziv bits as an enhancement for the systematic unprotected bitstream. In the reconstructed frame, with Wyner-Ziv transmission rate of 1.5 bits per pixel, the distortion caused by the macroblock errors and error propagation is greatly reduced. Upon close inspection, a few isolated pixel errors can be observed. For high resolution video, which is typical for video broadcast, isolated pixel errors are not very evident to the viewer.

The trade-off between distortion in the case of transmission errors and Wyner-Ziv bit-rate can be exploited to construct an embedded Wyner-Ziv code that achieves graceful degradation of the decoded video when the error rate of the channel increases. The advantage of this scheme is that it can achieve graceful degradation without the need



(a) (b)

Figure 5: PSNR at the decoder is better if sequence is protected by Wyner-Ziv bits (a) *Carphone* and (b) *Foreman*

(a) (b)

Figure 6: Decoded video quality improves with rate of Wyner-Ziv bitstream (a) Error concealment only (31.5dB) (b) With Wyner-Ziv bits @ 1.5 bpp (37.5dB)

for layered video representations, which are known to incur a substantial rate-distortion penalty. Morever, the average decoded PSNR can be further improved if the MPEG decoder uses the Wyner-Ziv decoded frame, instead of the conventional error-concealed frame, when it decodes the next frame. This corresponds to the best reconstruction PSNR curve in Figs. 5(a) and 5(b). These significant enhancements will be discussed in detail in [15].

5 Conclusions

In this paper we proposed a Wyner-Ziv video codec which uses intraframe encoding and interframe decoding. This type of codec is useful for systems which require simple encoders but can handle more complex decoders. The encoder is composed of a scalar quantizer and a rate compatible turbo encoder. The decoder performs turbo decoding using an interpolated frame as side information. We showed that our proposed scheme performs 2 to 7 dB better than H263+ intraframe encoding and decoding. The scheme has not yet reached the compression efficiency of a H263+ interframe coder but this gap could be reduced in the future by exploiting spatial correlation in the proposed codec.

We also proposed a novel application of Wyner-Ziv video coding in error-resilient video broadcast. Using a Wyner-Ziv encoder, we generate a supplementary bitstream that mitigates transmission errors in the decoded video waveform. In the event of transmission errors, the turbo decoder uses the conventionally decoded, error-concealed video signal as side information to decode the Wyner-Ziv bits. We showed in our first results that the proposed scheme corrects transmission errors up to a certain minimum distortion, substantially improving visual quality. Our future aim is to refine our Wyner-Ziv coder in order to reduce the rate overhead of the supplementary Wyner-Ziv bitstream.

References

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 471–480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[3] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Proc. IEEE Data Compression Conference*, Snowbird, Utah, Mar. 1999, pp. 158 –167.

[4] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, no. 10, pp. 417–419, Oct. 2001.

[5] Y. Zhao and J. Garcia-Frias, "Data compression of correlated non-binary sources using Punctured Turbo Codes," in *Proc. IEEE Data Compression Conference*, Snowbird, Utah, Apr. 2002, pp. 242–251.

[6] J. Bajcsy and P. Mitran, "Coding for the slepian-wolf problem using turbo codes," in *Proc. IEEE Global Communications Symposium*, San Antonio, Texas, Nov. 2001, pp. 1400–1404.

[7] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," in *Proc. IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002, p. 91.

[8] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conference*, Snowbird, Utah, Apr. 2002, pp. 252 –261.

[9] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440–442, Oct. 2002.

[10] A. Jagmohan, A. Sehgal, and N. Ahuja, "Predictive encoding using coset codes," in *Proc. International Conference on Image Processing*, Rochester, New York, Sept. 2002, pp. 29–32.

[11] S.S. Pradhan and K. Ramchandran, "Enhancing analog image transmission systems using digital side information: a new wavelet based image coding paradigm," in *Proc. IEEE Data Compression Conference*, Snowbird, Utah, Mar. 2001, pp. 305 –309.

[12] A. Liveris, Z. Xiong, and C. Georghiades, "A distributed source coding technique for correlated images using Turbo codes," *IEEE Communications Letters*, vol. 6, pp. 379–381, Sept. 2002.

[13] S. Shamai, S. Verdu, and R. Zamir, "Systematic Lossy Source/Channel Coding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 564–579, mar 1998.

[14] D. Rowitch and L. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo codes," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 948–959, June 2000.

[15] D. Rebollo A. Aaron, S. Rane and B. Girod, "Systematic Lossy Forward Error Protection of Video Waveforms," in *Proc. IEEE International Conference on Image Processing, Invited Paper*, Barcelona, Spain, Sept. 2003.