TOWARDS PRACTICAL WYNER-ZIV CODING OF VIDEO

Anne Aaron, Eric Setton, and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305
{amaaron,esetton,bgirod}@stanford.edu

ABSTRACT

In current interframe video compression systems, the encoder performs predictive coding to exploit the similarities of successive frames. 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 similar efficiency. We report results on a Wyner-Ziv coding scheme for motion video that uses intraframe encoding, but interframe decoding. In the proposed system, key frames are compressed by a conventional intraframe codec and in-between frames are encoded using a Wyner-Ziv intraframe coder. The decoder uses previously reconstructed frames to generate side information for interframe decoding of the Wyner-Ziv frames.

1. INTRODUCTION

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 in complexity is desirable for broadcasting or for streaming video-on-demand systems where video is compressed once and decoded many times. However, some future systems may require the dual scenario. For example, we may be interested in compression for mobile wireless cameras uploading video to a fixed base station. Compression must be implemented at the camera where memory and computation are scarce. For this type of system what we desire is 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 compression scheme where individual frames are encoded independently (*intraframe encoding*) but decoded conditionally (*interframe decoding*). 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-Mean Squared Error 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.

Although these information theoretic results present significant insights on compression, there are few examples where they have been considered for practical compression applications. Pradhan and Ramchandran applied distributed source coding to a system where a digital stream enhances the quality of a noisy analog image transmission [3]. Similarly, Liveris et al. used turbo codes to encode the pixels of an image with a noisy version of the image available at the decoder [4]. In these systems [3, 4], Wyner-Ziv coding is applied to natural images, and the side information is defined to be a version of the image corrupted by additive Gaussian noise. In [5], Jagmohan et al. discuss how a predictive coding scheme with multiple predictors can be seen as a Wyner-Ziv problem, and thus, can be solved using coset codes. Specifically, they suggest that Wyner-Ziv codes could be used to prevent prediction mismatch or drift in video systems.

In [6] we apply Wyner-Ziv coding to a real-world video signal. We take X as the even frames and Y as the odd frames of the video 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. A similar video compression system using distributed source coding principles was proposed independently by Puri et al. in [7].

In this work, we extend the Wyner-Ziv video codec, first outlined in our paper [6], to a more general and practical framework. The *key frames* of the video sequence are compressed using a conventional intraframe codec. The remaining frames, the *Wyner-Ziv frames*, are intraframe encoded using a Wyner-Ziv encoder. To decode a Wyner-Ziv frame, previously decoded frames (both key frames and Wyner-Ziv frames) are used to generate side information. Interframe decoding of the Wyner-Ziv frames is performed by exploiting the inherent similarities between the Wyner-Ziv frame and the side information.

In Section 2, we describe the proposed Wyner-Ziv video codec. In Section 3, we present different frame dependency arrangements and discuss the decoder flexibility in generating side information. Finally, in Section 4, we compare the performance of the proposed coder to conventional intraframe coding, using a standard H263+ video coder.

2. WYNER-ZIV VIDEO CODEC

We propose an intraframe encoder and interframe decoder system for video compression as shown in Fig. 1. A subset of frames from

the sequence are designated as key frames. The key frames, K, are encoded and decoded using a conventional intraframe codec. In between the key frames are Wyner-Ziv frames which are intraframe encoded but interframe decoded.
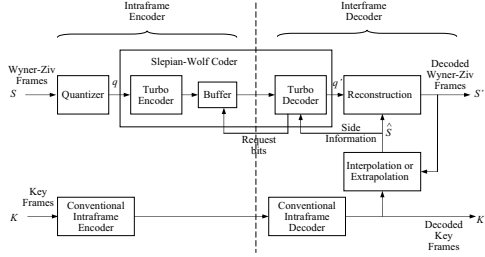
Fig. 1. Wyner-Ziv video codec with intraframe encoding and interframe decoding.

A Wyner-Ziv frame, S, is encoded as follows: We quantize each pixel value of the frame using a uniform scalar quantizer with 2^M levels to form the quantized symbol stream q. We take these 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 code (RCPT) [8]. 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 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 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 Wyner-Ziv frame, the decoder takes adjacent previously decoded key frames and, possibly, previously decoded Wyner-Ziv frames to form the side information, $\hat{S}$, which is an estimate of S. Ways to generate the side information is discussed in more detail in Section 3. To be able to exploit the side information, the decoder assumes a statistical dependency model between S and $\hat{S}$.

The turbo decoder uses the side information $\hat{S}$ and the received subset of parity bits to form the decoded symbol stream q'. 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' it calculates a reconstruction of the frame S', where $S' = E(S|q', \hat{S})$. With this reconstruction function, if the side information is within the reconstructed bin, the reconstructed pixel will take a value very close to the side information. If the side information is outside the bin, the function clips the reconstruction 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 frame (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. FLEXIBLE DECODER SIDE INFORMATION

Analogous to the idea of varying the ratio of I frames, P frames, and B frames in conventional video coding, we can vary the number of Wyner-Ziv frames between key frames to achieve different rate-distortion points for the proposed system. Many high quality key frames in the sequence lead to better side information for the Wyner-Ziv frames. For example, if there is one key frame for every Wyner-Ziv frame, the decoder can perform sophisticated motion compensated (MC) interpolation on the two adjacent key frames to generate a very good estimate of the Wyner-Ziv frame. For this case, reconstruction errors from other decoded Wyner-Ziv frames need not corrupt the side information of the current Wyner-Ziv frame. Better side information translates to improved rate-distortion performance for the Wyner-Ziv encoded frame. However, since the key frames are intraframe encoded and decoded, they require more rate than the Wyner-Ziv frames, so the over-all rate of the system increases. Finding a good trade-off between the number of key frames and the degradation of the side information is a significant aspect in optimizing the compression performance. Aside from the number of key frames, the quality of their reconstruction also affects the side information and is an important consideration in the design.

The proposed Wyner-Ziv video coder employs feedback from the decoder to the encoder to send the proper number of bits. This is advantageous since the required bit-rate depends on the side information which is unknown to the encoder. Because of this feedback, the decoder has great flexibility to choose what side information to use. In fact, given the same Wyner-Ziv video encoder, there can be decoders of different sophistication and with different statistical models. For example, a "smart" decoder might use sophisticated motion compensated interpolation and request fewer bits, while a "dumb" decoder might use no motion compensation at all (simply takes a reconstructed adjacent frame as the side information) and request more bits for successful decoding.

Fig. 2 and 3 illustrate two hierarchical frame dependency arrangements. In the diagrams, both schemes have the same number of Wyner-Ziv frames in between the key frames, but the corresponding side information techniques applied at the decoder are different. In Fig. 2, the decoded previous frame (whether a key frame or a Wyner-Ziv frame) is extrapolated to generate the side information for the current Wyner-Ziv frame. This technique requires a minimum of memory at the encoder. A more complex arrangement, shown in Fig. 3, increases the temporal resolution 2:1 in each step of the hierarchy with bidirectional interpolation. This technique requires larger memory in the encoder. For N hierarchy levels, the punctured symbols of $2^N - 1$ frames have to be stored. Besides the different memory requirements, identical encoders can potentially be used by both schemes. The different decoders would, however, request different numbers of bits for the individual frames and, hence, choose a different bit allocation. Furthermore, the characteristics of the error propagation into dependent frames for the two schemes would be different.
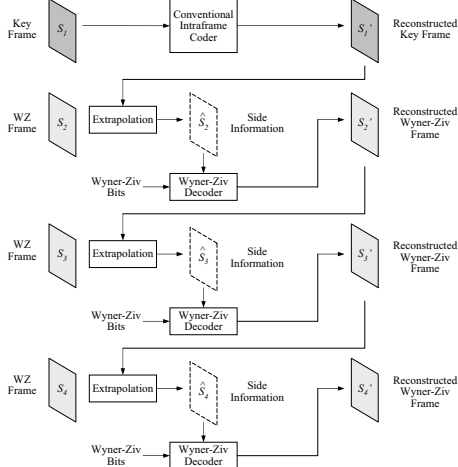
Fig. 2. Reconstructed previous frames are extrapolated to form the side information for the Wyner-Ziv decoding of the next frame.
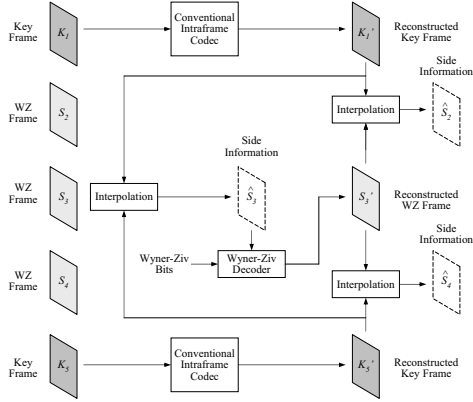
Fig. 3. A hierarchical frame dependency structure where previously reconstructed key frames and Wyner-Ziv frames are used to interpolate a frame.

4. SIMULATION RESULTS

We implemented the intraframe encoder - interframe decoder system proposed in Section 2 and assessed the performance for QCIF video sequences.

The key frames were encoded as I frames, with a fixed quantization parameter, using a standard H263+ codec. We varied the number of Wyner-Ziv frames in between the key frames and changed the frame dependency structure for each case. For these simulations, we applied MC interpolation, based on the assumption of symmetric motion vectors, to generate the side information. Let $MCI(A, B, d)$ be the result of MC interpolation between frames A and B at d fractional distance from A. Four frame dependency arrangements were simulated with the side information derived as follows:

- 1 WZ frame: K_1-S_2-K_3
    - $\hat{S}_2 = MCI(K_1', K_3', \frac{1}{2})$
- 2 WZ frame: K_1-S_2-S_3-K_4

- $\hat{S}_2 = MCI(K_1', K_3', \frac{1}{3})$
- $\hat{S}_3 = MCI(K_1', K_3', \frac{2}{3})$
- 3 WZ frames: K_1-S_2-S_3-S_4-K_5 (Illustrated in Fig. 3)
    - $\hat{S}_3 = MCI(K_1', K_5', \frac{1}{2})$
    - $\hat{S}_2 = MCI(K_1', S_3', \frac{1}{2})$
   - $\hat{S}_4 = MCI(S_3', K_5', \frac{1}{2})$
- 4 WZ frames: K_1-S_2-S_3-S_4-S_5-K_6
    - $\hat{S}_3 = MCI(K_1', K_6', \frac{2}{5})$
    - $\hat{S}_4 = MCI(K_1', K_6', \frac{3}{5})$
    - $\hat{S}_2 = MCI(K_1', S_3', \frac{1}{2})$
    - $\hat{S}_5 = MCI(S_4', K_6', \frac{1}{2})$

Given a fixed frame arrangement, we varied the rate of the Wyner-Ziv frames by changing the number of levels of the scalar quantizer ($2^M \in \{2, 4, 16\}$).

The turbo encoder was composed of two constituent convolutional encoders of rate $\frac{4}{5}$, identical to those used in [9]. 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 $P_e \geq 10^{-3}$ it requests for additional parity bits.

The results for the *Foreman* and *Carphone* QCIF sequences are shown in Fig. 4 and Fig. 5. The rate and PSNR shown in the plots are the average values for all the frames (both key frames and Wyner-Ziv frames). The frame rate is 30 frames per second. We compare the rate-PSNR performance of the proposed system to that of conventional intraframe coding using H263+.
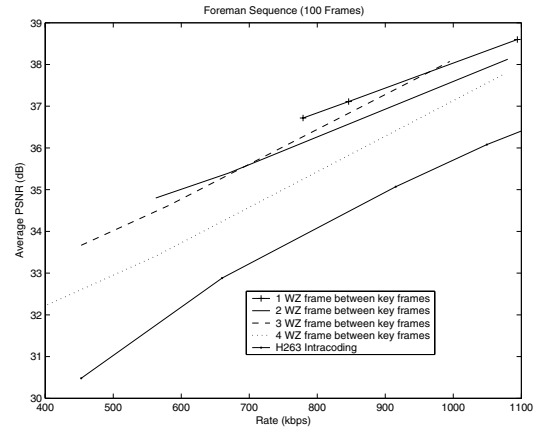
Fig. 4. Bit rate vs. Average Frame PSNR for *Foreman* sequence

As it can be seen from Fig. 4 and 5, at moderate to high bit rates, spacing the key frames too far apart results in a lower PSNR compared to the case with closer key frames at the same bit rate. This is expected since MC interpolation is less accurate, thus, degrading the side information, and error propagation is more severe. However, using fewer key frames is necessary to operate at the lower bit rates. If the quality of the key frames are to be fixed at a reasonable level to allow effective interpolation, which was the case in the experiments, these low rates cannot be achieved by setups with many key frames because of the rate overhead associated with them.
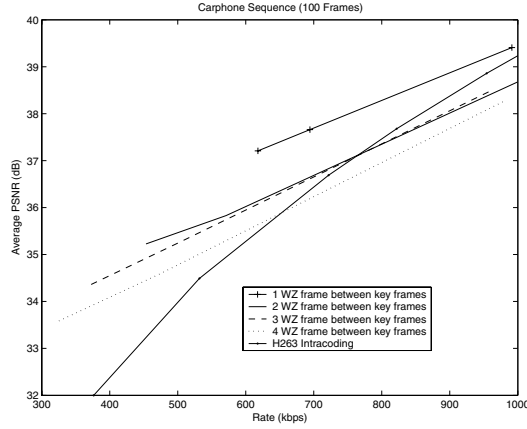
Fig. 5. Bit rate vs. Average Frame PSNR for *Carphone* sequence

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

Fig. 6. Interpolated and encoded frame from *Foreman*. Three Wyner-Ziv frames in between key frames.

In Fig. 6 we see that the Wyner-Ziv coding scheme can fix the MC interpolation artifacts in cases of occlusions and high motion and produce visually acceptable PSNR values. The image on the left is a frame $\hat{S}_3$ generated by interpolating between $K_1{}'$ and $K_5{}'$. The frame on the right is encoded with $2^M = 16$ levels (average sequence bit-rate = 1000 kbps). As we can see, the encoding sharpens the image and closely reconstructs the face even if the interpolation fails. 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 average bit-rate, we observe that the intraframe decoded sequence has obvious blocking artifacts which are not present in our system. The H263+ average PSNR is 2.5 dB lower than that of the proposed scheme.

For the *Foreman* sequence, the proposed Wyner-Ziv video codec performed 1.5 to 3 dB better than H263+ intraframe coding. The gain above the H263+ codec is smaller for the *Carphone* sequence. This can be attributed to the fact that most of the new information in the sequence is caused by occlusions – changing scenery in the car window, mouth opening and closing. This new information cannot be easily predicted by MC interpolation, especially with key frames spaced farther apart. The side information is less useful and high bit-rates are necessary to decode the Wyner-Ziv frames.

A new type of 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 re-

liable. In the experiments, we allow a maximum error rate of less than 10^{-3} (25 pixels per frame).

5. CONCLUSIONS

In this paper we propose a Wyner-Ziv video codec which uses intraframe encoding but interframe decoding. This type of codec is useful for systems which require simple encoders but can handle more complex decoders. The video sequence is divided into key frames which are coded using a conventional intraframe codec and the Wyner-Ziv frames which are intraframe encoded and interframe decoded.

We simulated a practical Wyner-Ziv video compression system and investigated different frame dependency arrangements involving motion compensated interpolation. We showed that our proposed scheme can perform up to 3 dB better than H263+ intraframe coding. The current system does not exploit any spatial correlation of the Wyner-Ziv frames. Future work will focus on incorporating the spatial statistics to further reduce the bit-rate.

6. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[2] A. 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, "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. 63–72.

[4] A. Liveris, Z. Xiong, and C. Georghiades, "A distributed source coding technique for correlated images using turbo codes," *IEEE Communications Letters*, vol. 6, no. 9, pp. 379–381, Sept. 2002.

[5] A. Jagmohan, A. Sehgal, and N. Ahuja, "Predictive encoding using coset codes," in *Proc. International Conference on Image Processing*, Rochester, New York, Sept. 2002, vol. 2, pp. 29–32.

[6] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, California, Nov. 2002.

[7] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, Oct. 2002.

[8] 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.

[9] 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.