Wyner-Ziv Residual Coding of Video

Anne Aaron, David Varodayan, and Bernd Girod*

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

Abstract. In Wyner-Ziv coding of video, a frame is encoded independent of other frames, but decoded using adjacent frames as side information. In this work we extend our system by Wyner-Ziv encoding the residual of a frame with respect to a known reference at the encoder. At the decoder, better side information is generated through motion estimation and is used to reconstruct the frame. We show through experimental results that by allowing the encoder this additional complexity of frame store and frame subtraction, we achieve better rate-distortion performance compared to previous Wyner-Ziv coding schemes.
Index Terms Wyner-Ziv coding, distributed source coding, video coding

1 INTRODUCTION

Implementations of current video compression standards, such as the ISO MPEG schemes or the ITU-T recommendations H.263 and H.264 require much more computation for the encoder than for the decoder; typically the encoder is 5 to 10 times more complex. This asymmetry is well-suited for broadcasting or for streaming video-on-demand systems where video is compressed once and decoded many times. However, some applications may require the dual system, i.e., low-complexity encoders, possibly at the expense of high-complexity decoders. Examples of such systems include wireless video sensors for surveillance, wireless PC cameras, mobile camera phones, and future networked camcorders. For these applications, compression must be implemented at the camera where memory and computation are scarce.

To achieve low-complexity encoding, we have developed a *Wyner-Ziv video codec* – an asymmetric video compression scheme where individual frames are encoded independently (*intraframe encoding*) but decoded conditionally (*interframe decoding*) [1, 2]. A similar video compression system was first suggested by Witsenhausen and Wyner in a 1980 United States Patent [3], and was also developed independently by Puri and Ramchandran in recent years [4–6]. Sehgal et al. also propose Wyner-Ziv coding for a state-free causal video encoder [7].

Two results from information theory suggest that an intraframe encoder - interframe decoder system can approach 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 $R_x + R_y \geq H(X, Y)$ [8]. The partial extension of this theorem for lossy source coding are Wyner and Ziv's results on source coding with side information [9]. 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.

In this work we extend our Wyner-Ziv video codec by Wyner-Ziv encoding the residual of a frame with respect to an available reference frame at the encoder. Since we aim for low encoder complexity, we apply Wyner-Ziv coding simply to the frame difference obtained without motion compensation. In particular, we apply Wyner-Ziv residual coding to the pixels of a frame and simply use the previous frame pixels as encoder reference. The decoder, which is less complexity constrained, can generate better side information using compute-intensive motion estimation techniques. With this scheme, the encoder can exploit some of the similarities between the current frame and the previous frame, while the decoder can use both the previous frame and the more sophisticated, motion-compensated side information for conditional decoding.

* This work is supported in part by the National Science Foundation Grant No. CCR-0310376.

We show through experimental results that by allowing the encoder this additional complexity of frame store and frame subtraction, the pixel-domain Wyner-Ziv residual coder achieves better rate-distortion performance compared to our previous pixel-domain schemes and similar performance as transform-domain Wyner-Ziv video codecs.

In Section 2, we describe in detail the proposed Wyner-Ziv residual video codec. In Section 3, we compare the performance of the proposed codec to conventional intraframe coding, conventional interframe predictive coding and non-residual Wyner-Ziv coding schemes.

2 WYNER-ZIV RESIDUAL VIDEO CODEC

The Wyner-Ziv residual video codec is depicted in Fig. 1. At the encoder, the difference between the current frame, X, and the encoder reference frame X_{er} is fed into a Wyner-Ziv encoder. X_{er} can be any frame that is easy to generate at the encoder – for example, the reconstructed previous frame or the average of two or more adjacent reconstructed frames. To avoid drift, X_{er} should be replicable at the decoder. The decoder generates side information, Y, using more sophisticated, motion-based techniques. Since the decoder takes into account motion, Y is expected to be a better estimate of frame X than X_{er}. The Wyner-Ziv decoder uses both Y and X_{er} to calculate the reconstruction X'.

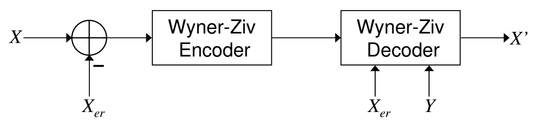


Fig. 1. Wyner-Ziv residual video codec. Residual of a frame with respect to an encoder reference frame is fed into a Wyner-Ziv encoder.

In Fig. 2, we show the details of the pixel-domain Wyner-Ziv residual codec which we implemented for the simulation results in Sec. 3. The pixels of the residual frame, $D = X - X_{er}$, are quantized using a deadzone uniform scalar quantizer to generate D_Q. The bit-planes of the quantized symbols, D_Q, are encoded by a low-density parity check (LDPC) code and the accumulated syndrome bits are stored in an encoder buffer [10]. As described in [10], the accumulation of the syndrome bits allows rate-adaptivity which is necessary for varying frame statistics. The encoder transmits a subset of these bits to the decoder upon request. The encoder can also calculate and send hash information to aid the decoder in performing motion estimation [11].

The decoder generates the side information Y by applying motion-compensated interpolation, motion-compensated extrapolation or hash-based motion estimation on previously reconstructed adjacent frames. It also reconstructs the encoder reference frame X_{er}. The LDPC iterative decoder successively decodes the bit-planes starting with the most significant bit-plane. It takes the received subset of accumulated syndrome bits corresponding to the bit-plane and the residual side information $Y - X_{er}$ to decode the current bit-plane. Note that using the side information residual $Y - X_{er}$ instead of the joint information (Y, X_{er}) makes decoding simpler. In general, the Wyner-Ziv decoding blocks can optimally use Y and X_{er} and the corresponding probability distribution $f(X|Y, X_{er})$. If the decoder cannot reliably decode the bits, it requests additional accumulated syndrome bits from the encoder buffer through feedback. The request-and-decode process is repeated until an acceptable probability of symbol error is guaranteed. The probabilities generated for the current bit-plane are used for decoding the less significant bit-planes. When all the bit-planes are decoded, the decoded symbols and the residual side information $Y - X_{er}$ are used by the reconstruction block to reconstruct D', which is added back to X_{er} to generate X'.

With the addition of the encoder reference frame generation, frame store and frame subtraction, the pixel-domain Wyner-Ziv residual encoder has slightly higher complexity than the original pixel-domain Wyner-Ziv video coder described in [1]. The new residual scheme exploits some of the similarities of the frame with respect to the known reference at the encoder, but still benefits from conditional decoding with respect to a better side information frame at the decoder. This allows the system to achieve better rate-distortion performance than the original pixel-domain Wyner-Ziv codec.

The rate-distortion improvement of Wyner-Ziv residual coding over non-residual coding can be attributed more specifically to two things. First, quantizing the residual $D = X - X_{er}$ can be interpreted as conditionally designing quantizers for X given X_{er}. At high rates, shifting the Wyner-Ziv quantizers does not significantly impact the performance of the system [12]. At low rates, the histogram of the quantized symbols D_Q shows high probability around the zero bin and this results in lower con-

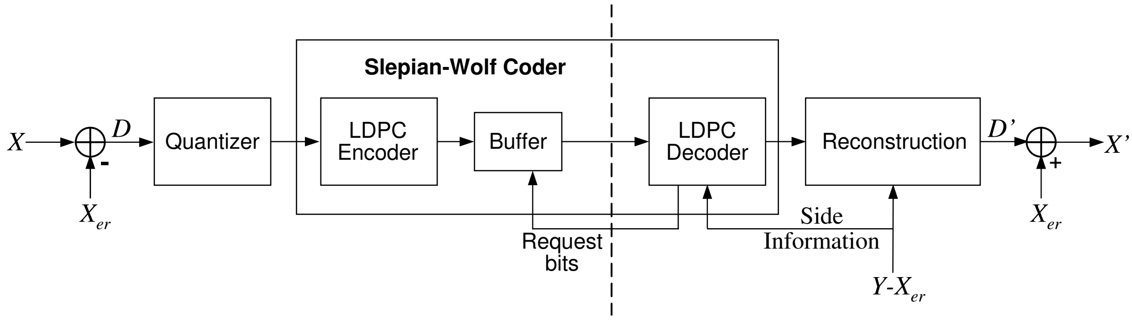


Fig. 2. Wyner-Ziv residual video codec with LDPC-based Slepian-Wolf codec.

ditional entropy $H(D_Q|Y - X_{er})$ compared to the conditional entropy of the quantized X given the side information Y, for the same distortion. Secondly, in the non-residual Wyner-Ziv coding case, the encoder reference frame is not used at all. In the residual coding case, X_{er} can be seen as a second side information available both at the encoder and at the decoder. If X_{er} and X are not conditionally independent given Y, then using X_{er} at the encoder and decoder can reduce the encoding rate of the system.

3 SIMULATION RESULTS

To investigate the performance of the Wyner-Ziv residual codec, we implement two simulation set-ups and test on the first 100 frames of the *Salesman* and *Foreman* QCIF sequences at 15 fps. For the pixel-domain Wyner-Ziv schemes, we use a rate-adaptive LDPC accumulate code of block length 25344 bits. For the DCT-domain systems, the LDPC accumulate code applied to each coefficient band has a block length of 396 bits.

3.1 Side Information from Motion-Compensated Interpolation

In the first simulation set-up we assume alternating *key frames* and *Wyner-Ziv frames*. Key frames are simply intracoded as I frames using a standard H.263+ codec. The Wyner-Ziv frames are Wyner-Ziv encoded and the side information at the decoder is generated using motion-compensated interpolation of the previous and next key frame. For the simulation results shown in Fig. 3, we compare three Wyner-Ziv coding schemes:

1. **Pixel-domain Wyner-Ziv residual coding**. The frame is encoded using the system described in Fig. 2. We use the previous reconstructed key frame as the encoder reference X_{er}.

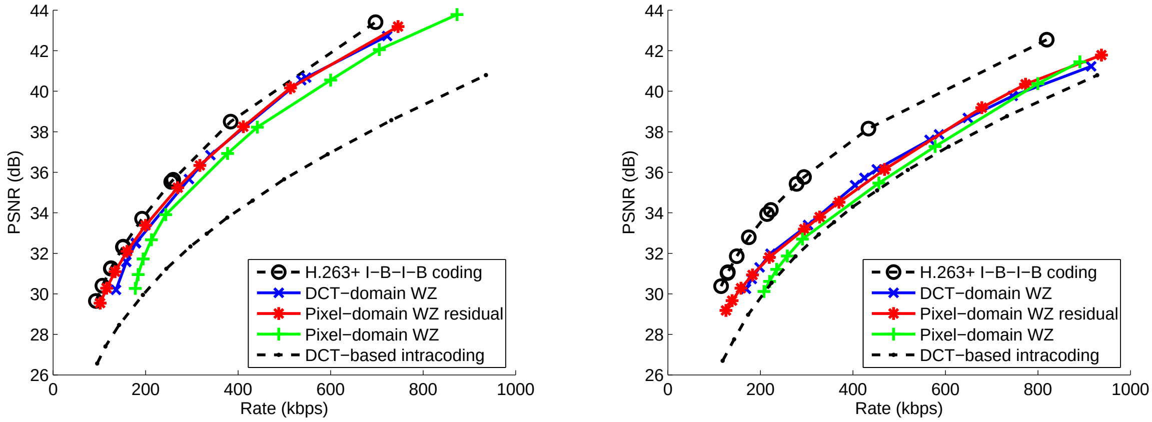
2. **Pixel-domain Wyner-Ziv coding**. Wyner-Ziv coding is applied to the pixels of the frame itself, similar to the scheme described in [1].

3. **DCT-domain Wyner-Ziv coding**. An 8x8 DCT is applied to the frame and all the coefficients are quantized using deadzone uniform scalar quantizers with the same step size. The quantized symbols of the ten lowest frequency coefficient bands are independently compressed using Wyner-Ziv coding [13]. The remaining high frequencies are quantized and compressed using H.263+ zero run-length coding and entropy coding.

We compare the rate-distortion performance of these schemes to H.263+ intraframe coding (all I frames) and H.263+ interframe coding with an I-B-I-B predictive structure. The rate and PSNR values in Fig. 3 are averaged over both key frames and Wyner-Ziv frames. As it can be seen from the plots, the pixel-domain Wyner-Ziv residual coding scheme has slightly better compression performance (less than 1 dB in most cases) than the non-residual pixel-domain system. For the *Salesman* sequence, all Wyner-Ziv coding schemes have significantly better rate-distortion performance than intraframe coding. This is not the case for the *Foreman* sequence which has high motion throughout the frame.

3.2 Hash-based Side Information

For the second set of experiments, every 8th frame is a key frame and the remaining frames are Wyner-Ziv frames. For the Wyner-Ziv frames, the encoder



(a) (b)

Fig. 3. Rate vs. PSNR for (a) *Salesman* and (b) *Foreman* QCIF sequences at 15 fps. Every other frame is encoded as an I frame and the side information is generated by motion-compensated interpolation.

generates a hash for each block [11]. The encoder determines whether to send the hash or not depending on the distance of the current hash from the hash of the previous co-located block. If the distance is smaller than a threshold, a "no hash bits" codeword is sent. If the distance exceeds the threshold, the hash is sent. At the decoder, for a given block of the current frame, if no hash bits are sent, the co-located block from the previous frame is used as the side information. If the hash is sent, the decoder uses it in a motion search to generate the best side information block from the previous frame. The following hash-based Wyner-Ziv coding schemes are compared in Fig. 4:

1. **Pixel-domain Wyner-Ziv residual coding with pixel residual hash**. The hash for each block is a quantized subsample of the pixels in $Y - X_{er}$. We calculate the entropy of the sent hash symbols and add this ideal rate to the rate shown in Fig. 4. We apply the system in Fig. 2 to encode the Wyner-Ziv frames, with the previous frame utilized as encoder reference.

2. **Pixel-domain Wyner-Ziv coding with pixel hash**. A quantized subsample of X is used as a hash. Similar to the pixel residual hash, we calculate the entropy of the pixel hash sent and include it in the plot. Each Wyner-Ziv frame is encoded using the simple pixel-domain Wyner-Ziv system [1].

3. **DCT-domain Wyner-Ziv coding with DCT hash**. We apply the Wyner-Ziv scheme described in [14]. An 8x8 DCT is applied to the frame and the ten lowest frequency coefficient bands are independently compressed using Wyner-Ziv coding. The remaining high frequencies are quantized and compressed using H.263+ zero run-length coding and entropy coding. These high frequencies, if sent, serve as the hash at the decoder.

In Fig. 4, we compare the Wyner-Ziv results to H.263+ DCT-based intracoding and H.263+ interframe coding with an I-P-P predictive structure and GOP size of 8. It is evident from Fig. 4 that the pixel-domain Wyner-Ziv residual scheme has significantly better performance (2 to 3 dB) than the original pixel-domain system. The improved performance is due to both the reduction of the hash bitrate, as well as the reduction of the Wyner-Ziv bits sent for a given quantizer step size. Since the quantizers of the residual system are "aligned" to the encoder reference frame, the conditional entropy of the quantized symbols given the side information is usually less than that of the non-residual scheme.

For both simulation set-ups, pixel-domain Wyner-Ziv residual coding exhibits similar compression performance as DCT-domain Wyner-Ziv coding. The advantage of the pixel residual system is that there is no need to optimally allocate rate between different coefficient bands. In transform-domain Wyner-Ziv codecs, bit allocation between frequency bands tends to be non-trivial since the rate and distortion are affected by the side information known only at the decoder.

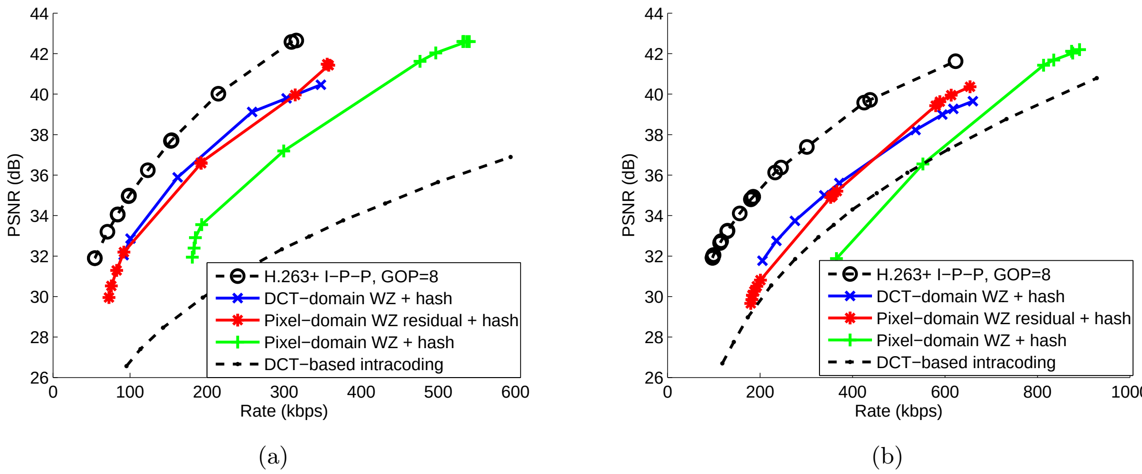


Fig. 4. Rate vs. PSNR for (a) *Salesman* and (b) *Foreman* QCIF sequences at 15 fps. Every 8th frame is encoded as an I frame and the side information is generated through hash-based motion estimation at the decoder.

4 CONCLUSIONS

In this work we perform Wyner-Ziv coding on the residual pixels of a frame with respect to a simple encoder reference. In this new Wyner-Ziv residual coding scheme, the encoder can exploit part of the redundancies among adjacent frames, while the decoder uses more sophisticated motion estimation techniques to conditionally decode the frames using better side information. The scheme requires the additional complexity of frame store and frame subtraction at the encoder, but demonstrates an improvement in compression efficiency (up to 3 dB) compared to simply Wyner-Ziv encoding the pixels of a frame.

References

1. Aaron, A., Zhang, R., Girod, B.: Wyner-Ziv coding of motion video. In: Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA (2002)
2. Girod, B., Aaron, A., Rane, S., Rebollo-Monedero, D.: Distributed video coding. Proc. of the IEEE **93** (2005) 71–83
3. Witsenhausen, H., Wyner, A.: Interframe coder for video signals. United States Patent 4191970 (1980)
4. Puri, R., Ramchandran, K.: PRISM: A new robust video coding architecture based on distributed compression principles. In: Proc. Allerton Conference on Communication, Control, and Computing, Allerton, IL (2002)
5. Puri, R., Ramchandran, K.: PRISM: An uplink-friendly multimedia coding paradigm. In: Proc. International Conference on Acoustics, Speech, and Signal Processing, Hong Kong (2003)
6. Puri, R., Ramchandran, K.: PRISM: A 'reversed' multimedia coding paradigm. In: Proc. IEEE International Conference on Image Processing, Barcelona, Spain (2003)
7. Sehgal, A., Jagmohan, A., Ahuja, N.: A causal state-free video encoding paradigm. In: Proc. IEEE International Conference on Image Processing, Barcelona, Spain (2003)
8. Slepian, D., Wolf, J.K.: Noiseless coding of correlated information sources. IEEE Transactions on Information Theory **IT-19** (1973) 471–480
9. Wyner, A., Ziv, J.: The rate-distortion function for source coding with side information at the decoder. IEEE Transactions on Information Theory **IT-22** (1976) 1–10
10. Varodayan, D., Aaron, A., Girod, B.: Rate-adaptive distributed source coding using Low-Density Parity-Check codes. In: Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA (2005)
11. Aaron, A., Rane, S., Girod, B.: Wyner-Ziv video coding with hash-based motion compensation at the receiver. In: Proc. IEEE International Conference on Image Processing, Singapore (2004)
12. Rebollo-Monedero, D., Aaron, A., Girod, B.: Transforms for high-rate distributed source coding. In: Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA (2003)
13. Aaron, A., Rane, S., Setton, E., Girod, B.: Transform-domain Wyner-Ziv codec for video. In: Proc. SPIE Visual Communications and Image Processing, San Jose, CA (2004)
14. Aaron, A., Girod, B.: Wyner-Ziv video coding with low encoder complexity. In: Proc. IEEE International Conference on Image Processing, San Francisco, CA (2004)