

IMPROVING WYNER-ZIV VIDEO CODING BY BLOCK-BASED DISTORTION ESTIMATION

Zouhair M. Belkoura and Thomas Sikora

Communication Systems Group, Technische Universität Berlin
Einsteinufer 17, 10587 Berlin, Germany

phone: + (49) 3031428917, fax: + (49) 3031422514, email: belkoura,sikora@nue.tu-berlin.de
web: www.nue.tu-berlin.de

ABSTRACT

Conventional video coding uses motion estimation to perform adaptive linear predictive encoding. Wyner-Ziv coding does not use predictive coding but performs motion estimation at the decoder. Recent work uses a difference signal at the encoder to estimate the prediction quality at the decoder. In this paper, we recognise that this operation constitutes a step in the motion estimation process. We exploit this information by omitting suitable blocks, effectively implementing linear predictive coding with deadzone quantisation for parts of the input signal. This modified Wyner-Ziv coding results in large bitrate reductions as well as significant decoding complexity decrease. At certain bitrates, our modified Wyner-Ziv codec outperforms conventional hybrid coding in an I-B-I-B setup.

1. INTRODUCTION

The past few years have seen a rapid increase in research work on Distributed Video Coding (DVC) techniques, also known as Wyner-Ziv (WZ) coding of video. A non-exhaustive overview of the achievements so far is given in [1]. Groundbreaking work was also done by Ramchandran et al [2].

In the treatment of single source video data, development of these techniques are mainly motivated by the search for encoders that use few computations. This is different from the classical, predictive coding approach like H.264 (called conventional in the following), where the encoder performs most operations. The most computationally intensive part of conventional coding is the block-based motion estimation procedure, consisting of an iterative search algorithm to find an optimal prediction for an input signal block. Replacing a conventional codec by a turbo coder based WZ codec (e.g [3]) shifts the bulk workload from encoder to decoder. Temporal video signal correlation is not explicitly reduced during encoding but exploited at the decoder to produce a prediction for the coded signal. This prediction, again, is found using motion estimation algorithms. The turbo decoding process used to correct the prediction errors at the decoder adds workload to the decoding process. In [4], we show that the overall workload in a Wyner-Ziv video codec can exceed conventional coding complexity by an order of magnitude. Decoder complexity reduction is therefore an important aspect in WZ codec design.

Based on the approach introduced in [3], numerous results have been presented which improve the rate-distortion performance of WZ coding. Many of these works concentrate on the decoder side. Artigas and Torres [5] propose an improvement by iteratively improving the side information.

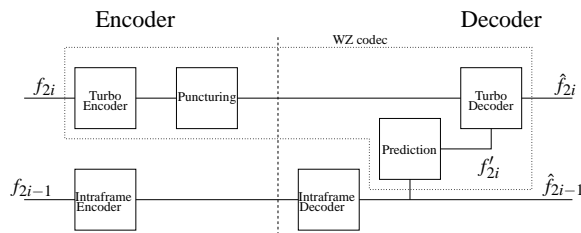


Figure 1: Wyner-Ziv codec using a turbo coder

Nataro et al. [6] generate side information using extrapolation to improve performance while Ascenso et al. propose better frame interpolation algorithms [7], [8] for side information generation.

Trapanese et al. [9] focus on the encoding process and introduce a block-based mode decision in the Wyner-Ziv frames. Performance gains are achieved by compressing a subset of blocks using conventional transform coding instead of the WZ approach.

In this paper, we reverse the idea of [9]. Instead of performing transform coding for a subset of blocks, we omit suitable blocks from the encoding process altogether. The mode decision process itself produces a prediction for the input video blocks. Omitting suitable blocks from encoding is equivalent to predictive coding with deadzone quantisation. We exploit this to reduce the required transmission bitrate for WZ coding of the frame. At the same time, we achieve substantial decoder workload reductions.

We continue by presenting the underlying WZ codec architecture in section 2 in detail. Sections 3 and 4 discuss the modifications proposed. Experimental results are presented in 5, giving an evaluation of the bitrate performance and the workload reduction. Section 6 concludes the paper.

2. TURBO CODER BASED WYNER-ZIV VIDEO CODEC

The WZ codec presented below is an implementation of [7]. Figure 1 shows the functional blocks of the WZ codec. The odd frames f_{2i-1} (also called key-frames) of a video sequence f_1, f_2, \dots, f_M are coded using conventional intraframe coding. The even frames f_{2i} (also called W-frames in the following) of size N pixels are split into bitplanes which are then turbo encoded. The two constituent codes of the $(3N, N)$ systematic turbo code are $(2N, N, K)$ systematic recursive convolutional codes, with K being memory length. At the turbo coder output, the N systematic bits

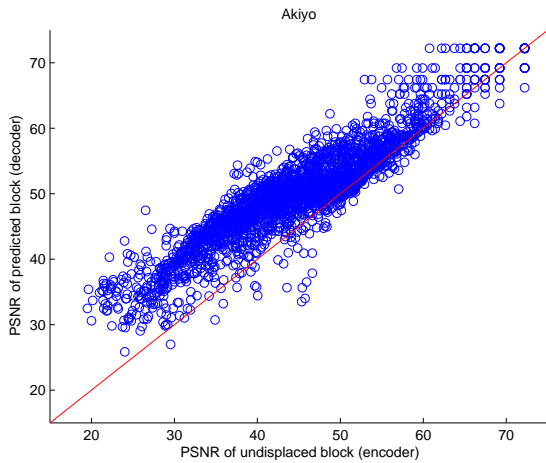


Figure 2: blockwise DPCM quality against decoder prediction quality (lossless keyframes). The line illustrates where both quantities are equal.

are discarded. In addition, the remaining $2N$ parity bits are punctured according to a puncturing rule: for every p_p bits, $p_p - n_p$ bits at known positions are deleted. Thus, only

$$\text{WZ-rate} = \frac{n_p}{p_p} 2N$$

bits are transmitted to the decoder. At the decoder, the intra-coded odd frames f_{2i-1} and f_{2i+1} are decoded and produce a prediction f'_{2i} of f_{2i} . The f'_{2i} bitplanes are used as systematic input for turbo decoding. In combination with the received parity bits, the turbo decoder corrects the prediction errors in f'_{2i} and returns a reconstructed version \hat{f}_{2i} as its output.

The prediction f'_{2i} is produced using motion compensated interpolation. Displacement vectors are calculated for each block using the reconstructed key-frames \hat{f}_{2i-1} and \hat{f}_{2i+1} . Assuming continuous motion for the blocks, interpolation is performed along the motion trajectories defined by the displacement vectors [7].

If the reconstructed odd frames \hat{f}_{2i-1} are not identical to f_{2i-1} , the prediction f'_{2i} will have more errors with respect to f_{2i} . With this setup, variation in puncturing varies the WZ-rate. Rate can be traded for reconstruction quality by omitting bitplanes. This omission can be seen as equivalent to a quantisation operation. With bitplanes missing, the reconstruction \hat{f}_{2i} will differ from f_{2i} .

3. MODIFIED PREDICTIVE CODING OF IMAGE SEQUENCES

Some 20 years ago, Jayant and Noll [10] explain hybrid video coding as a combination of transform coding in the spatial direction and differential (DPCM) coding in the temporal direction. In [10], differential coding requires a linear prediction filter producing a prediction $\hat{X}(n)$ for the current input $X(n)$ using previous inputs:

$$\hat{X}(n) = \sum_{i=1}^M \alpha_i X(n-i)$$

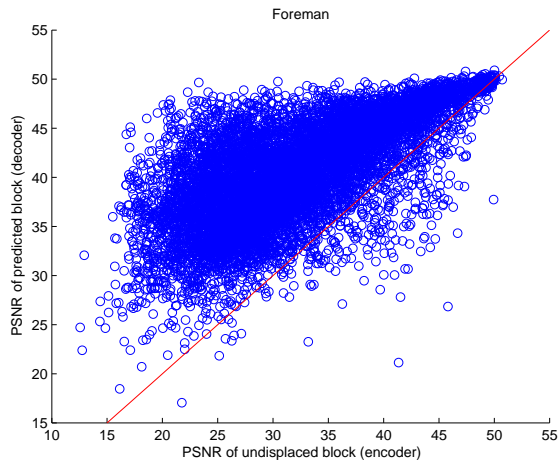


Figure 3: blockwise DPCM quality against decoder prediction quality (lossless keyframes). The line illustrates where both quantities are equal.

For a video signal, this can be formulated as:

$$\hat{f}(m, n, t) = \sum_{i=1}^M \alpha_i f(\beta_i, \gamma_i, t - i)$$

Using this vocabulary, motion compensation in conventional hybrid video simply implements adaptive linear prediction with the motion vector field being the prediction filter coefficients β_i, γ_i , which are optimally chosen using a motion estimation search algorithm.

We propose here to incorporate into the Wyner-Ziv codec: nonadaptive predictive coding and deadzone quantisation on a blocklevel. This is implemented by computing a difference signal for each block as described in section 4. If the difference signal energy falls below a given threshold, nothing is sent to the decoder, i.e. the block is skipped and reconstructed at the decoder by simply inserting the corresponding block from the previous frame.

3.1 Further Improvements

Computation of the block difference between adjacent frames gives a good estimate of the motion compensated prediction computed at the Wyner-Ziv decoder. Figures 2 and 3 show scatter plots comparing the quality of the nonadaptive prediction for each block computed at the encoder (x-axis) to the prediction at the decoder (y-axis) for the Akiyo and Foreman test sequences. Clearly, there is a correlation between these values. More importantly, however, the decoder prediction quality is higher than the encoder prediction in almost all cases. It is therefore sensible to replace skipped blocks not by the blocks in the previous frame, but by the predicted blocks computed at the decoder. This implements a modified differential decoding. The next section details how this is incorporated into the Wyner-Ziv codec.

4. MODIFIED WYNER-ZIV CODEC

For each even input frame f_{2i} , forward and backward difference signals d_f, d_b are produced using

$$d_f = f_{2i} - f_{2i+1}$$

and

$$d_b = f_{2i} = f_{2i-1}$$

The difference signals are then subdivided into square blocks $b_{f,m,n}$ and $b_{b,m,n}$ containing 256 pixels. For each position (m,n) , the mean squared error

$$e_{f,m,n} = \frac{1}{256} \sum_{i=1}^{16} \sum_{j=1}^{16} b_{f,m,n}^2(i,j)$$

is computed. Similarly, an MSE $e_{b,m,n}$ is obtained. Depending on the target quality, a threshold E is set. If $e_{f,m,n} < E$ or $e_{b,m,n} < E$, the block at position (m,n) will not be Wyner-Ziv coded. As detailed above, not WZ coding these blocks in effect corresponds to a quantisation operation of a difference signal in predictive coding, with the quantiser threshold for deadzone quantisation set to E . In this work, the mode decision is stored in a binary vector s corresponding to the block scan order of the frame. This binary vector needs to be transmitted to the decoder. This overhead information can be compressed using run-length coding [9] and is of negligible size.

This modified codec is called Skip-WZ codec in the following. The Skip-WZ (SWZ) encoder operates as follows:

- Encode key frames f_{2i-1} and f_{2i+1}
- Generate a mode decision vector using f_{2i-1} , f_{2i} and f_{2i+1}
- Form a row of input blocks from f_{2i} , skipping the ones where s is true
- Split the row of input blocks into bitplanes and perform turbo coding and puncturing

At the decoder side, the processing is given by:

- Decode key frames f_{2i-1} and f_{2i+1}
- Generate prediction f'_{2i}
- Perform turbo decoding for non-skipped blocks only

5. EXPERIMENTAL RESULTS

The first 51 frames of the Akiyo test sequence were coded using the Wyner-Ziv codec, the H.264 codec and the proposed SWZ codec. The coding structure was I-B-I-B for the conventional codec and I-W-I-W for the other two. Each sequence contained 25 WZ coded frames. The video size was 176x144 pixels, with frame rate 30Hz. The turbo coders in the WZ and SWZ codec used a puncturing period of $p_p = 32$. Both turbo coder generator polynomials were set to $\frac{33}{23}$, given in octal form, resulting in a coder memory of $K = 4$. Naturally, for each frame in the SWZ codec, the turbo coder blocklength and interleaver was varied, because the number of input blocks varied.

Two experiments were run, one with lossy keyframes and another one with lossless keyframes. For each experiment, four WZ-rate points were simulated. This bitrate variation was achieved by transmitting between one and four bitplanes for each f_{2i} .

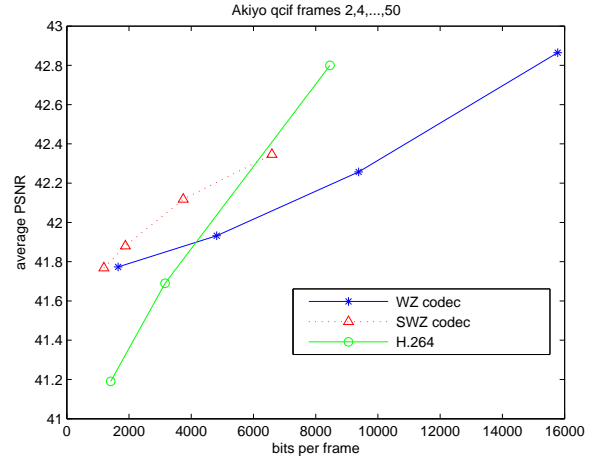


Figure 4: average PSNR against bitrate for W/B frames. Lossy keyframes. On average 50% of blocks are skipped.

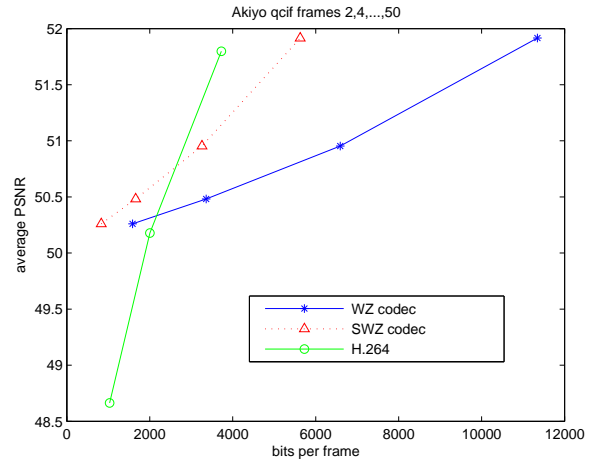


Figure 5: average PSNR against bitrate for W/B frames. Lossless keyframes. On average 50% of blocks are skipped.

Conventional video coding was simulated with varying quantisation parameter ($qp(B)$) for B-Frames. For the lossless keyframe case, $qp(B)$ was selected from $\{16, 20, 24\}$. For the lossy keyframe case, $qp(B)$ was selected from $\{20, 22, 24\}$. Motion estimation in H.264, WZ and SWZ codecs was performed using full-pel precision and 16 pel search range. Lossy keyframe transmission was simulated using H.264 coding with quantisation parameter $qp(I) = 24$, resulting in an average distortion of ≈ 41 dB PSNR.

Figures 4 and 5 show the coding performance of the three simulated codecs. Using block-based mode decision (SWZ codec), up to 50% of the rate can be saved compared to WZ coding. Especially in the low rate case, the SWZ codec outperforms conventional hybrid coding.

Not only rate-distortion gains can be observed. Figures 6 and 7 plot the average workload per frame against average quality for the WZ codec and the SWZ codec. The decoder complexity calculation was done as presented in [4]. Like with the bitrate, a substantial decrease in required workload can be seen.

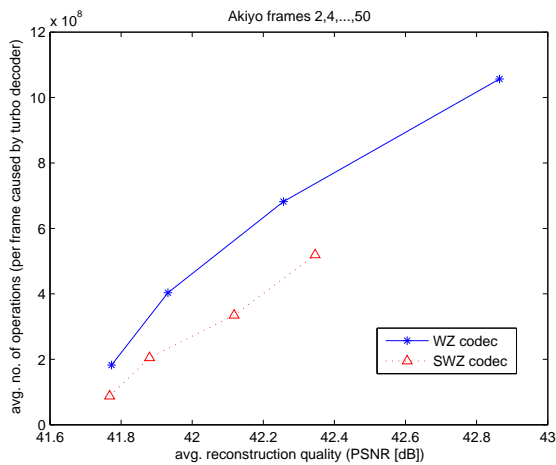


Figure 6: average decoding workload against average PSNR for W frames. Lossy keyframes.

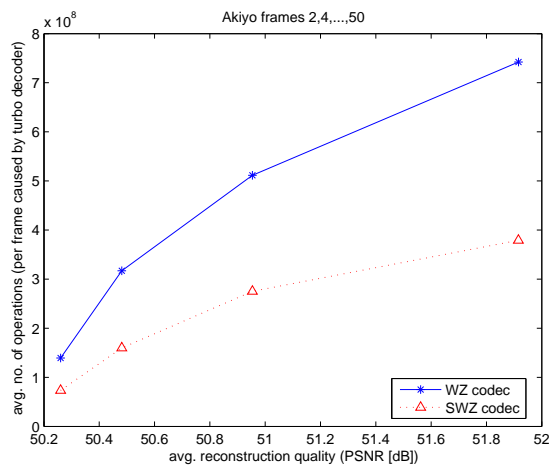


Figure 7: average decoding workload against average PSNR for W frames. Lossless keyframes.

6. CONCLUSIONS AND OUTLOOK

We have combined turbo coder based Wyner-Ziv video coding with modified predictive coding on a block level. It is shown that large reductions in bitrate and decoding complexity can be achieved. A further step would see the combination of the coding modes presented here with the intra-modes of [9]. Correlation between the decoder prediction signal and encoder prediction signal has been shown. Current work examines how this fact can be exploited to vary the puncturing rates on a block level.

REFERENCES

- [1] Bernd Girod, Anne Aaron, Shantanu Rane, and David Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71 – 83, January 2005.
- [2] R. Puri and K. Ramchandran, "Prism: A 'reversed' multimedia coding paradigm," in *Proceedings of the*

IEEE International Conference on Image Processing, Barcelona, Spain, 2003.

- [3] A. Aaron, E. Setton, and B Girod, "Towards practical wyner-ziv coding of video," in *Proceedings of the IEEE International Conference on Image Processing, Barcelona, Spain, Sept. 2003.*
- [4] Zouhair M. Belkoura and Thomas Sikora, "Toward rate-decoder complexity optimisation in turbo-coder based distributed video coding," in *Picture Coding Symposium, Beijing, China, 2006.*
- [5] Xavi Artigas and Luis Torres, "Iterative generation of motion-compensated side information for distributed video coding," in *Proceedings of the IEEE International Conference on Image Processing, Genoa, Italy, 2005.*
- [6] L. Natario, C. Brites, J. Ascenso, and F. Pereira, "Extrapolating side information for low-delay pixel-domain distributed video coding," in *Int. Workshop on Very Low Bitrate Video Coding, Sardinia, Italy,, September 2005.*
- [7] J. Ascenso, C. Brites, and F. Pereira, "Motion compensated refinement for low complexity pixel based distributed video coding," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance, Como - Italy, September 2005.*
- [8] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, June 2005.*
- [9] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Embedding a block-based intra mode in frame-based pixel domain wyner-ziv video coding," in *Int. Workshop on Very Low Bitrate Video Coding, Sardinia, Italy, September 2005.*
- [10] N.S. Jayant and Peter Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.