

Wyner-Ziv Video Coding with Side Matching for Improved Side Information*

Bonghyuck Ko, Hiuk Jae Shim, and Byeungwoo Jeon

Department of Electronic and Electrical Engineering, Sungkyunkwan University
300 Chunchun-dong, Jangan-gu, Suwon, 440-746, Korea
{o3eda@skku.edu, waitnual@ece.skku.ac.kr, bjeon@skku.edu}

Abstract. To make an encoder extremely simple by eliminating motion prediction/compensation from encoder, source coding with side information has been investigated based on the Wyner-Ziv theorem as the basic coding principle. However, the frame interpolation at decoder which is essential for redundancy elimination makes erroneous side information when the basic assumption of linear motion between frames is not satisfied. In this paper, we propose a new Wyner-Ziv video coding scheme featuring side matching in the frame interpolation to improve the side information. In the proposed scheme, Wyner-Ziv decoder compensates wrong blocks by side information using side matching and bi-directional searching. The noise reduction in side information allows the proposed algorithm to achieve coding improvements not only in bit-rate but also in PSNR. Results of our experiments show improvement of PSNR up to 0.4dB.

Keywords: Distributed video coding, Wyner-Ziv coding, Side matching.

1 Introduction

In conventional video coding such as MPEG-1/2/4 and H.26x, the complexity of encoder is much higher than that of decoder. It is because the temporal redundancy among frames is removed at the encoder using motion estimation and compensation (ME/MC) process. The ME process is computationally very intensive, and also so in consuming power, therefore in many applications where saving in power and computation is a premium factor such as in sensor networks, a new encoding algorithm which claims only low complexity is very essential.

It has been known that the distributed source coding (DSC) [1,2] can provide a clue to solutions to this problem. The Slepian-Wolf [1] and the Wyner-Ziv theorems [2] state that it is possible to encode *correlated* sources *independently* while still achieving optimum compression performance, as long as decoding is performed jointly. In the context of video coding, it enlightens that all the processing at an encoder to exploit temporal and spatial correlation in the video frames can be shifted

* This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the National Research Lab. Program funded by the Ministry of Science and Technology (No.M10600000286-06J0000-28610).

to a decoder side. It makes a video encoder utterly simple at the expense of increased complexity at the decoder. It is just an opposite design principle from the traditional broadcasting.

Lately, Girod *et al.* proposed a Wyner-Ziv coding scheme based on these theorems in which encoder performs turbo encoding, and decoder generates side information reflecting temporal correlation of frames [3]. The decoder also does the turbo decoding to reduce noise between original frame and its side information.

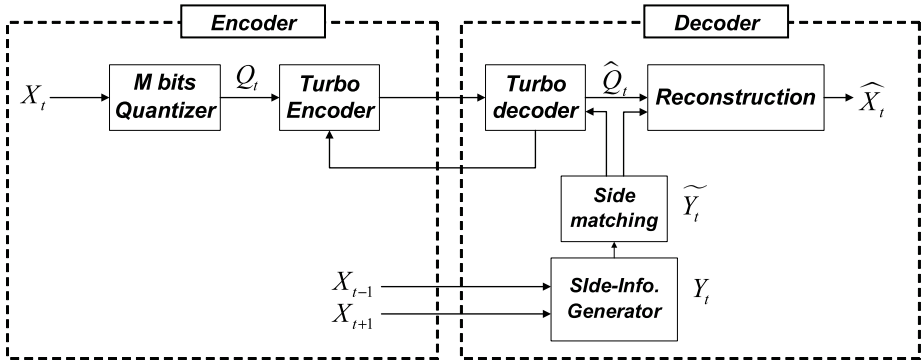


Fig. 1. Proposed Wyner-Ziv Coding Scheme with Side Matching

Basically, this scheme is highly dependant on the accuracy of side information, since the decoder extracts temporal redundancy based on the side information which is generated by frame interpolation using key frames. Note that, the frame interpolation assumes linear motion between frames. Therefore, if there is non-linear motion or occlusion, overall coding performance is bound to be decreased. In this paper, we propose a solution to this problem.

The remainder of this paper is organized as follows. Section 2 presents the coding process of the previous Wyner-Ziv coding [3,4]. In Section 3, we describe the proposed method. Section 4 shows experimental results. Conclusions and future works are drawn in Section 5.

2 Wyner-Ziv Video Coding

Fig. 1 illustrates the proposed architecture of the PDWZ (Pixel Domain Wyner-Ziv) coding. It consists of an M bit quantizer, a turbo-encoder/decoder based on Slepian-Wolf coding, a frame interpolation module generating side-information, a reconstruction module, and a *side matching module* which is newly proposed in this paper.

2.1 Encoding Process

Overall encoding process is as follows. First of all, all frames to be coded are organized into two categories - so called, “key frames” and “Wyner-Ziv frames.” As previously proposed pixel domain Wyner-Ziv coder [3,4], we take the same simple

approach of designating odd frames as key frames and even frames as Wyner-Ziv frames. The key frames are encoded as conventional intra frames by such as H.26x, and then sent to a decoder. In the decoder, these key frames are assumed to be perfectly reconstructed as previously proposed pixel domain Wyner-Ziv coding [3,4].

As for the Wyner-Ziv frames between key frames, the encoder performs turbo encoding. The turbo encoder consists of two RSC's (Recursive Systematic Convolution Coding) of rate 1/2 and an interleaver. Quantized symbol streams from the M bit quantizer are fed into these two RSC's. Input streams to the second RSC are interleaved. After this processing, the parity data produced by the turbo encoder is stored in a buffer, and the systematic data from the turbo encoding are discarded. Once the parity data is generated, puncturing is additionally exercised before parity is transmitted, and the turbo encoder sends the data to the decoder according to the request from the decoder. As already mentioned above, this type of encoder is utterly light-weighted in its complexity since the Wyner-Ziv encoder performs only intra coding, not referring to other frames as inter-coding.

2.2 Decoding Process

The decoding process is as follows. First, the decoder reconstructs the key frames as received from the encoder. By utilizing the reconstructed key frames, the decoder generates the side-information. Typically, it is done through frame interpolation assuming linear motion between key frames. After this step, side information is used by both the turbo decoder and the reconstruction module.

Our turbo decoder is based on MAP (Maximum A Posteriori) algorithm and, unlike the bit-level operation in the previously proposed schemes [3,4], it operates on symbol-level (named as the Hyper-trellis turbo code [5]). In this processing, side information is interpreted as a noisy version of the corresponding original Wyner-Ziv frame and used to decode quantized symbol stream Q_t by referring to the parity data from the encoder. A decision module in the turbo decoder calculates symbol error rate, and based upon this value, if required, it repeats requesting parity bits until the calculated symbol error rate (Pe) decreases under a threshold (for example, $Pe < 10^{-3}$). Once the threshold symbol error rate is achieved, the turbo decoder sends decoded symbol stream to the reconstruction module. The reconstruction module reconstructs the Wyner-Ziv frame using this quantized symbol stream and the side-information. The reconstruction rule is same as the previously proposed PDWZ [3,4].

3 Proposed Wyner-Ziv Video Coding with Side Matching

Our proposed Wyner-Ziv coding differs in two aspects with the previously proposed PDWZ [3,4]. First, the side matching module is added in the side information generator. Second, we use the hyper-trellis turbo coding [5] instead of turbo coding based on bit-plane. We will discuss what differences these two make in the coding performance, especially in subjective quality.

3.1 Importance of Good Side Information

In general, the Wyner-Ziv decoder performs three key functions: side information generation, turbo decoding, and reconstruction. Among them, the side information generation process is the most important in improving coding performance.



Fig. 2. Original 76th frame (left) and side information generated (right) in decoder (assumption of linear motion is satisfied)



Fig. 3. Original 180th frame (left) and side information generated (right) in decoder (assumption of linear motion is *not* satisfied)

One can regard the turbo decoding as a process which corrects, using parity, virtual channel noise originated from side information generation process, and reconstruction as a process which reduces ambiguity in reconstructed frame using correlation between the Wyner-Ziv frame and the side-information. Therefore, the closer the side information is to the original image, *i.e.*, smaller virtual channel noise, the lesser parity bits it will require, and also the better quality of reconstructed image it will get. Usually, the Wyner-Ziv decoder makes side information using frame interpolation assuming linear motion between key frames. Frame interpolation works well in such a circumstance that the motion between frames is well characterized as zero or linear motions. On the other hand, if there is non-linear motion or occlusion between frames, the decoder is bound to generate wrong blocks as side information as shown in Figures 2 and 3.

Although this type of error is desired not to occur so frequently, once it occurs, it is detrimental in overall coding performance, and unfortunately even extremely so in subjective quality. It usually looks like a burst noise. Moreover, it increases the number of requests for parity bits during turbo decoding process as well. Unlike the assumption of stationary processing in turbo decoding, real virtual channel noise is non-stationary. Several researches have been done related to this problem [6,7]. In this paper, however, we focus on reducing this type of non-stationary errors rather than developing a new channel noise model as in [6,7].

Note that although there is a report of a scheme to improve side information [8], our scheme is distinct from it in several aspects. First, our scheme works well with low M (which is same as quantized symbol length) compared to the scheme in [8]. Usually the probability of having such burst-like noise in MSB is very low, therefore it is difficult for the scheme [8] to decide whether errors exist or not. Second, the previous scheme [8] operates on bit-level approach; on the other hand, our proposed scheme can operate not only in bit-level but also in symbol-level. Thirdly, the coding performance of our scheme is affected less by channel coding or reconstruction module while the scheme in [8] is in the opposite way. It is because our scheme is a kind of pre-processing, but [8] is a kind of post-processing.

3.2 Side Matching

The degradation of subjective quality of reconstructed video by the decoder is due to wrong blocks from frame interpolation due to non-linear motion or occlusion. In this case, there are large discrepancies between sides of the block at corresponding positions in the key frames and at the side information.

To quantify the differences, we calculate MAD (Mean Absolute Difference) between sides of the block in side information S_i and average, $(S_{t-1} + S_{t+1})/2$, of corresponding sides at the key frames indicated by motion vectors obtained in the frame interpolation process as in eq. (1).

$$\alpha = MAD\{S_i, (S_{t-1} + S_{t+1})/2\}. \tag{1}$$

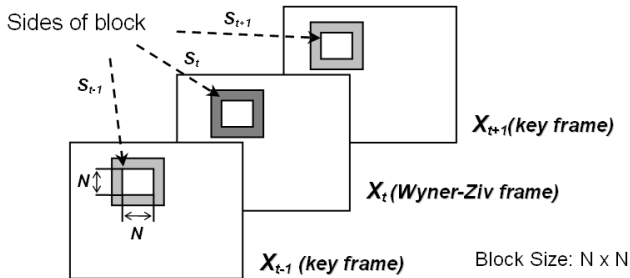


Fig. 4. Proposed Side Matching for Wyner-Ziv Coding

If the calculated MAD α in eq.(1) exceeds a predefined threshold \mathcal{E} , the side matching module decides the block as a wrong one. The threshold value \mathcal{E} is obtained from experiments.



Fig. 5. Wrong blocks in side information are marked using the proposed scheme

If $\alpha > \mathcal{E}$, then $B_t = B_i$

$$i = \underset{k=1,2,3}{\operatorname{argmin}} \{M_k | k=1, 2, 3\}$$

$$\text{where } \begin{cases} M_1 = \operatorname{MAD}\{S_t, S'_{t-1}\} & \text{forward direction} \\ M_2 = \operatorname{MAD}\{S_t, S'_{t+1}\} & \text{backward direction} \\ M_3 = \operatorname{MAD}\{S_t, (S'_{t-1} + S'_{t+1})/2\} & \text{bi-direction} \end{cases} \quad (2)$$

(B means a block, S means sides, and S' means searched sides)

For the blocks determined as wrong ones, the decoder generates side information again using bi-directional motion search. For this, three candidate blocks are generated with forward searching (B_1), backward searching (B_2), and bi-directional



Fig. 6. Original side information of 180th frame (left) and improved one by the proposed scheme (right)

searching (B_3) [9]. Among these three, the proposed decoder consequently selects one for the corresponding block in side-information as in eq.(2).

By choosing the best one among the forward searching, the backward searching, and the bi-directional searching, the decoder can estimate non-linear motion between frames better than before, thus total number of wrong blocks generated from non-linear motion and occlusion is expected to be lesser compared to the conventional side information generator based on frame interpolation assuming linear motion between key frames. Figures 5 and 6 are results of our simulation which shows that the decoder corrects such types of error in many cases.

3.3 Hyper-trellis Turbo Coding

The turbo decoder in the Wyner-Ziv coding corrects the differences between the original Wyner-Ziv frame and the side information inflicted by virtual channel noise. It decodes original quantized symbol stream using parity data from the encoder with its own powerful error correction capability. In turbo coding, virtual channel noise can be taken care of in two different ways - one is a symbol-level approach, and the other is bit-level approach. Although some degree of correlation still exists between bits constituting a symbol, correlation inherently does belong to symbol level. In this respect, the independence assumption of bit planes taken by previously proposed bit-plane-based turbo coding has a problem in calculating channel likelihood [5].

On the contrary, the Hyper-trellis turbo coder treats several combined bit transition corresponding to one symbol length as one symbol transition, thus calculation of the metric data for a state transition in trellis is obtained for the symbol length corresponding to the combined transition, not a bit level transition. Of course, channel likelihood is calculated on symbol-level. Therefore, the hyper-trellis-based turbo-coder does not have a problem of approximation unlike the bit-plane-based counterpart.

4 Experimental Results

In order to evaluate the coding performance of the proposed scheme, following two configurations are considered:

- i) Conventional: Hyper-trellis-based PDWZ with frame interpolation[5]
- ii) Proposed: Hyper-trellis-based PDWZ with frame interpolation using the proposed side matching

The test conditions are as follow. For the test sequences, we use Foreman (the one without Siemens logo), Mobile, and Stefan sequences in QCIF format (176 x 144 pixels). The Wyner-Ziv frame rate is 15 Hz like IST-PDWZ [4] - that is, all the even frames are coded as the Wyner-Ziv frames. In the experiment, no coding is carried out for the key frames and they are assumed to be perfectly reconstructed by decoder without any coding error. It is assumed so because the coding of the key frames is a separate issue from the Wyner-Ziv frames. In the evaluation of PSNR and bit rate, only those of Wyner-Ziv frames are considered since we are interested in evaluating coding performance of Wyner-Ziv frames only.

In the experiment, different quantization levels are applied, $2^M \in \{2, 4, 8, 16\}$, to obtain four rate-distortion points. For the Turbo coding, we use a generating function, $[\frac{1+D+D^3+D^4}{1+D^3+D^4}]$. Puncturing period is 32, and pseudo random interleaver is used. For the frame interpolation, block size is 8×8 , search range is ± 16 , and refinement range is ± 3 . For the side matching, decision threshold $\mathcal{E} = 8$, and search modes are forward, backward, and bi-direction as described in Section 3.2.

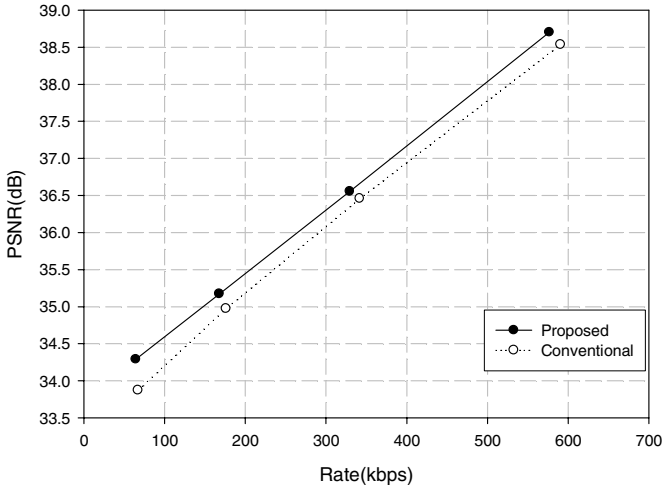


Fig. 7. RD performance of the proposed scheme with Foreman (frames 200~300)

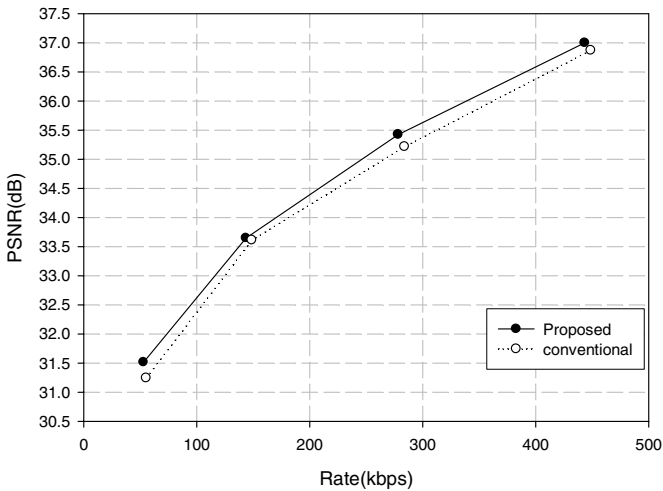


Fig. 8. RD performance of the proposed scheme with Mobile (frames 200~300)

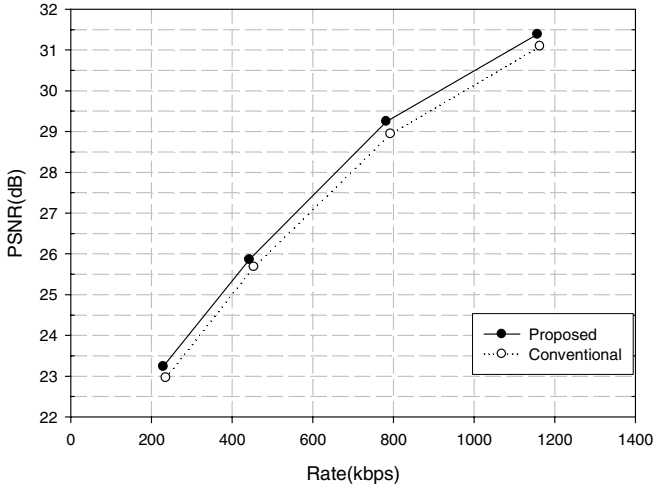


Fig. 9. RD performance of the proposed scheme with Stefan (frames 200~300)

In case of the Foreman sequence, PSNR is increased by up to maximum 0.42dB, minimum 0.09dB by the proposed scheme. In the minimum case, bit-rate is reduced by about 12Kbps. Since the LSB parts of a pixel in Wyner-Ziv frame are not transmitted from the encoder, the decoder inherits the LSB parts of the generated side information. Therefore, in LSB parts, reduction of burst noise aids to recover quantization error by the reconstruction module, consequently higher PSNR is obtained. On the other hand, in MSB parts, reduced burst noise makes turbo decoding easier with smaller number of parity bit requests, consequently lower bit-rate is obtained. Therefore better side information in decoding process is shown to increase PSNR and to decrease bit-rate at the same time in our experiments. Similar results are also observed with the Mobile and the Stefan sequences.

Table 1 shows average performance improvements which lead us to conclude that our proposed scheme is better than the conventional scheme[5], and as we mentioned above, performance improvements occur in both PSNR and bit-rate.

Table 1. Average performance improvements of proposed scheme over the conventional one [5]

| Sequence | BDPSNR(dB) | BDBR(%) |
|----------|------------|---------|
| Foreman | 0.29 | -12.66 |
| Mobile | 0.18 | -7.12 |
| Stefan | 0.33 | -6.39 |

5 Conclusion and Future Work

In this paper, we proposed using side matching in Wyner-Ziv Coding in order to increase the quality of side information. The proposed method enables us to reduce

burst type virtual channel noise, and makes more accurate side information. Therefore, the turbo decoder requests less parity bits than the conventional scheme. This means that the total transmission rate is eventually lowered in decoding process and reconstructed image has more reliable information about original data from the side-information.

The proposed scheme works well in many cases, however, we also observed, although they are rare, some cases of wrong blocks still occurring in groups after side matching in case of extremely fast non linear motions. As a future work, we will make more robust scheme which can also operate in such a very noisy circumstance having very fast non linear motion.

References

1. Slepian, D., Wolf, J.: Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory* 19, 471–480 (1973)
2. Wyner, A., Ziv, J.: The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inform. Theory* 22, 1–10 (1976)
3. Aaron, A., Zhang, R., Girod, B.: Wyner-Ziv Coding for Motion Video. In: 36th Asilomar Conference on Signals, Systems and Computers, Pacific Grove Monterey CA, pp. 240–244 (2002)
4. Ascenso, J., Brites, C., Pereira, F.: 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 (2005)
5. Avudainayagam, A., Shea, J.M., Wu, D.: Hyper-trellis decoding of pixel-domain Wyner-Ziv video coding. In: *IEEE GLOBECOM 2005 proceedings* (2005)
6. Dalai, M., Leonardi, R., Pereira, F.: Improving turbo coding integration in pixel-domain distributed video coding. In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse (2006)
7. Westerlakena, R.P., Klein Gunnewiekb, R., Lagendijka, R., Inald, L.: Turbo-Code Based Wyner-Ziv Video Compression. In: *Twenty-sixth Symposium on information Theory, Benelux* (2005)
8. Ascenso, J., Brites, C., Pereira, F.: Motion compensated refinement for low complexity pixel based distributed video coding. In: *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Como, Italia (2005)
9. Wu, S.W., Gersho, A.: Joint Estimation of Forward and Backward Motion Vectors for Interpolative Prediction of Video. *IEEE Transactions on Image Processing* 3, 684–687 (1994)