Effects of Variable Block Size Motion Estimation in Transform Domain Wyner-Ziv Coding

Dohyeong Kim, Bonghyuck Ko, Hiuk Jae Shim, Byeungwoo Jeon

School of Information and Communication Engineering
Sungkyunkwan University
Suwon, Korea
hangdo2001@skku.edu, o3eda@skku.edu, waitnul@ece.skku.ac.kr, bjeon@skku.edu

ABSTRACT

In the Wyner-Ziv coding, compression performance highly depends on the quality of the side information since better quality of side information brings less channel noise and less parity bit. However, as decoder generates side information without any knowledge of the current Wyner-Ziv frame, it doesn't have optimal criterion to decide which block is more advantageous to generate better side information. Hence, in general, fixed block size motion estimation (ME) is performed in generating side information. By the fixed block size ME, the best coding performance cannot be attained since some blocks are better to be motion estimated in different block sizes. Therefore if there is a way to find appropriate ME block of each block, the quality of the side information might be improved. In this paper, we investigate the effects of variable block sizes of ME in generating side information.

Keywords: Wyner-Ziv coding, variable block, motion estimation, side information

1. INTRODUCTION

Video coding has traditionally focused on reducing spatial and temporal redundancy in the video sequence. In order to do so, the popular conventional video coding standard such as MPEG-1/2/4 and H.26x has used block-based intra and inter prediction. In this structure, computational complexity of the encoder is much higher than that of the decoder [1]. That is because the ME is the most complex process of the overall process and the encoder tries to find out the best coding mode among the possible coding modes, while decoder just performs decoding process according to the knowledge sent from the encoder. This structure is not suitable for the applications where the encoder has limited resources such as power.

It has been known that Distributed Video Coding [2, 3] can provide solutions to this problem. In the Distribute Video Coding, there can be typically many encoders, and there exists only a few decoders. In this structure, the decoder performs joint decoding with the bitstreams sent from the independent encoders. This means that the decoder is responsible to explore the source statistics and all the process at the encoder can be shifted to the decoder. As a matter of fact, the Slepian-Wolf [2] and the Wyner-Ziv theorems [3] state that it is possible to encode correlated sources independently while still achieving optimum compression performance as long as decoding is performed jointly. Various video source coding techniques based on this idea have emerged lately and, as a whole, are known as distributed video coding (DVC). As resource-constrained video encoding has become an important issue, the needs for extremely low complexity video encoding are on the increase. DVC fits well in many new applications where the encoder uses limited resources. For example, wireless video surveillance, mobile video can be considered as this situation.

The Wyner-Ziv coding is a special case of Distributed Video Coding [4], [5]. In the Wyner-Ziv coding, decoder reconstructs video sequence by correcting noise in side information using channel code such as turbo code or LDPC code. The performance of the Wyner-Ziv coding is highly dependent on the quality of side information. Since the noise contained in the side information is corrected using the parity bits sent from the encoder. As a result, the better the quality of the side information is, the less the bits are requested by the decoder.

In spite of the importance of side information, the study on the variable block size ME has not been carried out sufficiently. Therefore, there exists no optimal criterion to decide the best block size in generating side information. In the Wyner-Ziv coding, as the decoder performs ME without any knowledge about the original Wyner-Ziv frame, the quality of side information can not guarantee overall coding performance, and can be easily affected by the motion in the video sequence. That is, if the original video sequence contains non-linear motion or occlusions between frames, it is hard to generate side information of high quality. In this case, variable block size ME might be an effective sophistigation to improve the quality of the side information by minimizing the MAD value between blocks in frame interpolation process.

In this paper, we investigate the effects of variable block sizes of ME in generating side information. The remainder of this paper is organized as follows. Section 2 presents the overall coding process of the Wyner-Ziv coding. Section 3 describes in detail the idea of variable block size-based side information generation. Section 4 presents experiments and results. Conclusion and comments on future work are presented in Section 5.

This work was supported by the Korea Science and Engineering Foundation (KOSER) NRL Program grant funded by the Korea government (MEST) (ROA-2006-000-10826-0(2008)).

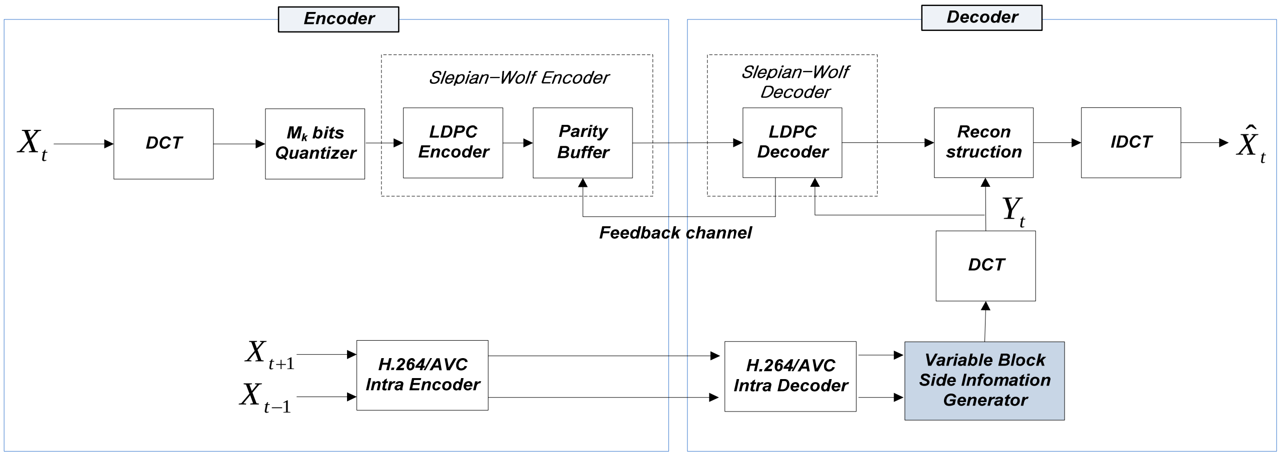

Fig. 1: Overall Codec Structure

2. OVERALL CODEC STRUCTURE

Fig. 1 illustrates the overall architecture of the Transform Domain Wyner-Ziv coding (TDWZ). The variable block size-based side information generator module is also depicted. Detailed coding process is as follows.

2.1 Encoding Process

First of all, as previously proposed [5], [6], we split video sequence into odd frames as key frames and even frames as Wyner-Ziv frames. The key frames are coded with H.264/AVC intra coder. After splitting the video sequence into two categories, 4x4 integer DCT transform is applied to the Wyner-Ziv frame, and the grouped DCT coefficients of the same frequency of a Wyner-Ziv frame form whole 16 DCT bands. Each band is quantized by M-bits uniform scalar quantizer. This quantizer reserves the upper M bits of the frame and gets rid of the lower bits. The bitstreams are reordered into bitplanes, and then sent to LDPC encoder. The LDPC encoder generates the parity bits in order from the most significant bitplanes. Then, the parity bits are transmitted to decoder when the parity request from decoder is acknowledged through feedback channel.

2.2 Decoding Process

The decoding process is as follows. First of all, the decoder reconstructs the key frames as received from H.264/AVC intra encoder. By utilizing the reconstructed key frames, the decoder generates 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 LDPC decoder [7] and the reconstruction module. The LDPC decoder computes soft-input using side information and transform coefficients modeled as Laplacian, and then soft-output is produced by the LDPC decoding algorithm. Based on soft-output value, soft-input is updated and LDPC decoding is performed iteratively. After iteration, bit plane is reconstructed and bit error rate (BER) is calculated. If calculated BER is higher than a pre-defined threshold value, the decoder requires additional parity bits to correct the noise in the side information. On the contrary, if calculated BER is lower than the threshold value, decoding of the bit plane in progress is terminated and the next bit plane is decoded. Finally, after LDPC decoding, in reconstruction module, the lower bits, removed in the quantization process, are reconstructed using both the quantized bit stream and side information.

3. SIDE INFORMATION GENERATION

In general, the Wyner-Ziv decoder performs three key functions: side information generation, channel code decoding, and reconstruction. Among them, the variable block sizes targets for improving the quality of the side information by using optimal block sizes in ME.

3.1 Motion Estimation with Fixed Block Size

Usually, the Wyner-Ziv decoder generates side information using frame interpolation assuming linear motion between key frames. Thus, 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. That is because decoder generates side information without any knowledge of the current Wyner-Ziv frame. Therefore, as decoder can not decide the best block size in ME process, fixed block size ME such as 8x8 is performed to generate side information, in general. Fig. 2 illustrates the side information of the Foreman sequence generated by fixed 8x8 and 4x4 block ME respectively. In Fig 2, the side information has much noise in the back ground region generated by 4x4 block ME. However, for the same frame, the side information generated by 8x8 block doesn’t appear so much noisy in the back ground region. Because in case of smaller block size-based motion estimation, the number of reliable pixel is not enough to be distinguished from adjacent block on account of spatial correlation. Hence, adjacent wrong block can be selected as best matching block in terms of MAD. Relatively, in case of larger block based motion estimation, the number of pixel is more abundant in comparison to

4x4 block fixed ME 8x8 block fixed ME

Fig. 2: Side Information for 37th frame (Foreman)

4x4 block fixed ME 8x8 block fixed ME

Fig. 3: Side Information for 111th frame (Foreman)

smaller block based ME. However, this assumption is not always true. Fig. 3 illustrates another generated side information of the Foreman sequence. The left is the side information generated by 4x4 block ME and the right is the side information generated by 8x8 block ME. As shown in the figure, although the subjective quality looks similar, the PSNR value of side information generated by 4x4 block ME is higher than that by 8x8 block ME. From these observations, if there exists a fine way to accommodate variable block size ME, the quality of side information can be improved.

3.2 Motion Estimation with Variable Block Size

Fig. 4 illustrates one possible procedure of variable block ME. Detailed algorithm is as follows.

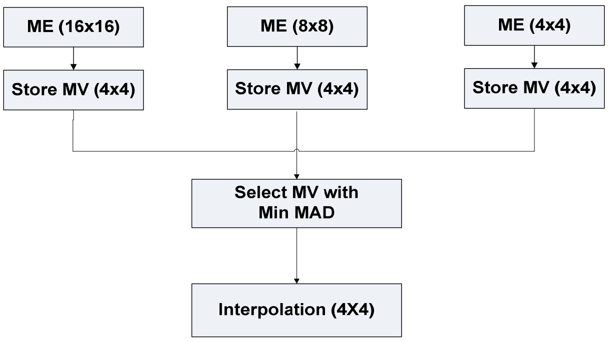

Fig. 4: Possible Procedure for Variable Block Size ME

3.2.1 Motion Estimation

First of all, 16x16 block ME is performed in a macroblock between key frames. The acquired motion vector by the 16x16 sized ME is stored in every 4x4 block for compatibility with the smaller sized ME in later stage. Hence, each 4x4 block shares the same motion vector value acquired by 16x16 ME. For the same macroblock, 8x8 block ME is performed four times, as there are four constituting 8x8 blocks. In this case, the motion vectors are

16	16	8	8
16	16	8	8
8	8	4	4
8	8	4	4

Fig. 4: An example of macroblock pattern

also stored in 4x4 block size. Therefore, 4x4 block shares the same motion vector for each 8x8 block. Lastly, 4x4 block ME is performed sixteen times for the macroblock, as there are sixteen different 4x4 blocks in the macroblock. In this case, each 4x4 block stores different motion vectors. In this way, side information for a key frame is generated for all the macroblock.

3.2.2 Motion Vector Selection and Interpolation

After motion estimation process, for each 4x4 block, side information generator compares MAD value in order to select motion vector with the minimum MAD. Fig. 4 illustrates an example of macroblock pattern which indicates the motion vector type. The largest block in Fig. 4 indicates a macroblock and sixteen small blocks mean each 4x4 block that contains motion vector. The number written in each block indicates the motion vector type which was acquired and stored in the process of variable motion estimation. Finally, motion compensated frame interpolation is performed for each 4x4 block size.

4. EXPERIMENTS AND RESULTS

In order to evaluate the performance of the variable block size-based ME in Wyner-Ziv video decoding, following three configurations are considered:

a) 8X8 Fixed Block ME: TDWZ (previously proposed in [5] and several refinement methods for side information proposed in [8]) with 8x8 fixed block ME.
b) Variable Block ME: 16x16, 8x8, 4x4 variable block ME as in Section 3.2.

Table 1: Quantization Matrix of each Sequence.

Sequence		Foreman	Coastguard
Quantization Matrix number for WZ frame	2	39	37
	4	34	34
	6	32	30
	8	25	26

The test conditions are as follows. For the test sequences, we use whole 15Hz Foreman and Coastguard sequences in QCIF format. In the experiment, different quantization [9] is applied to obtain four rate-distortion points as depicted in Table 1. For the fixed block sized ME, block size is set

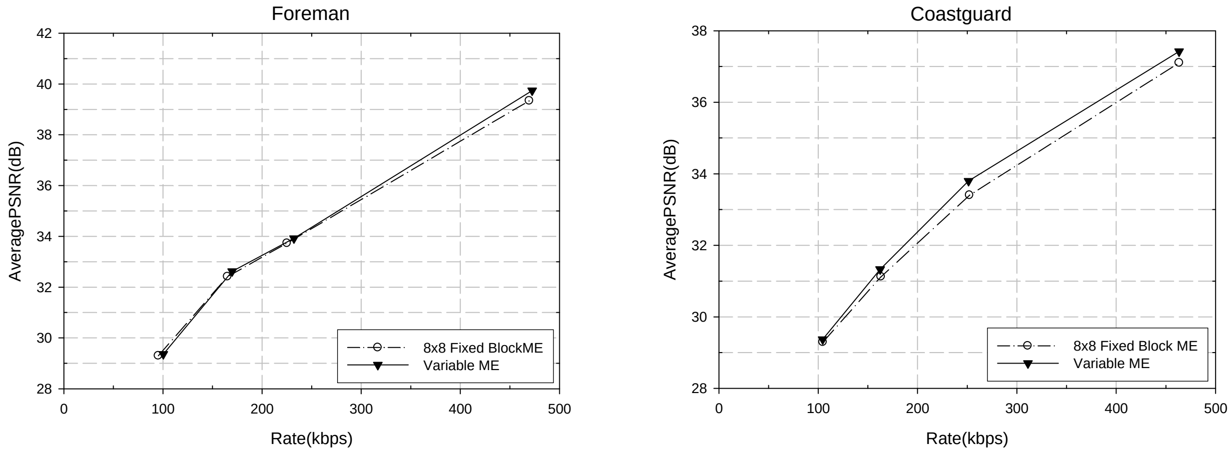

Fig. 5: R-D Performance Comparison.

as 8x8, and search range is set as ± 32, and refinement range is set as ± 4. For the variable block sized ME, block size is set as 16x16, 8x8, 4x4, and search range is set as ± 32, ± 32, ± 4, respectively and refinement range is set as ± 4. Fig. 5 shows R-D performance comparison for the Forman and Coastguard sequences. In the R-D performance, the Foreman sequence does not show noticeable coding gain. The coastguard sequence shows maximum PSNR gain up to 0.28dB. Note that the Foreman sequence has smaller gain than the Coastguard sequence. That is because, side information refinement module such as side matching module [10] is optimized to 8x8 block in the codec structure, relatively, 4x4 block based motion estimation shows less quality especially in the region with non-linear or occlusion. As the Coastguard sequence has lots of linear region for the Foreman sequence, our experiments show better coding performance with the Coastguard sequence. Above all, as the motion estimation performs without any knowledge of the current Wyner-Ziv frame, the acquired motion vector between key frames in terms of minimum MAD value cannot always guarantee the real motion vector of the video sequence. That is why the method in Section 3.2 does not show outstanding coding gain.

5. CONCLUSIONS

In this paper, we investigated the effect of variable block size motion estimation in order to find a way to improve the quality of side information. The presented method enables to make a little more accurate side information. The method could be used as a criterion for the channel decoder to decide more accurate block in performing motion estimation process to generate side information. As a future work, we are going to develop improved variable block motion estimation based on successive decoding.

6. REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, pp. 560-576, July 2003.

[2] J. D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources", IEEE Trans. on Information Theory, vol. IT-19, pp. 471 - 480, July 1973.

[3] A. Wyner and J. Ziv, “The rate-distortion fuction for source coding with side information at the decoder,” IEEE Trans. Information Theory, vol. 22, pp. 1-10, July 1976.

[4] A. Aaron, S. Rane, R. Zhang, and B. Girod, “Wyner-Ziv coding for video: Applications to compression and error resilience,” Proc. Of IEEE Data Compression Conference, DCC-2003, Snowbird, UT, March 2003.

[5] A. Aaron, S. Rane, E. Setton and B. Girod, “Transform-domain Wyner-Ziv codec for video,” Proc. Visual Communications and Image Processing, VCIP-2004, San Jose, CA, January 2004.

[6] A. Aaron, R. Zhang and B. Girod, “Wyner-Ziv Coding for Motion Video”, Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, November 2002.

[7] D. Varodayan, A. Aaron and B. Girod, “Rate-adaptive distributed source coding using Low-Density Parity-Check codes,” Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, November 2005.

[8] B. Ko, H. J. Shim, and B.Jeon, “Wyner-Ziv coding with spatio-temporal refinement based on successive turbo decoding,” IEEE International Conference on Multimedia and Expo, Hannover, Germany, June 2008.

[9] C. Brites, J. Ascenso, F. Pereira, “Improving Transform domain Wyner-Ziv Video coding performance,” IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 2006.

[10] B. Ko, H. J. Shim, and B. Jeon, “Wyner-Ziv Video Coding with Side Matching for Improved Side Information,” IEEE Pacific Rim Symposium on Image Video and Technology, Santiago, Chile, pp. 816-825, December 2007.