SELECTIVE HASH-BASED WYNER-ZIV VIDEO CODING

Taewon Do, Hiuk Jae Shim, Bonghyuck Ko, and Byeungwoo Jeon

School of Information and Communication Engineering
Sungkyunkwan University
Suwon, Korea
sponsor8@skku.edu, waitnual@ece.skku.ac.kr, o3eda@skku.edu, bjeon@skku.edu

ABSTRACT

Distributed video coding (DVC) is a new coding paradigm that enables to exploit the statistics among sources only in decoder and to achieve extremely low complex video encoding without any loss of coding efficiency. Wyner-Ziv coding, a particular implementation of DVC, reconstructs video by correcting noise on side information using channel code. Since a good quality of side information brings less noise to be removed by the channel code, generation of good side information is very important for the overall coding efficiency. However, if there are complex motions among frames, it is very hard to generate a good quality of side information without any information of original frame. In this paper, we propose a method to enhance the quality of the side information using small amount of additional information of original frame in the form of hash. By decoder's informing encoder where the hash has to be transmitted, side information can be improved enormously with only small amount of hash data. Therefore, the proposed method gains considerable coding efficiency. Results of our experiment have verified average PSNR gain up to 1 dB, when compared to the well-known DVC codec, known as DISCOVER codec.

Keywords: Wyner-Ziv coding, side information, hash selection

1. INTRODUCTION

The popular conventional video coding standards such as MPEG-1/2/4 and H.26x which exploit complex motion estimation at the encoder cannot meet the need for low power consumption and low computational complexity of encoder. Therefore, in many new applications where reduction of encoder complexity is an utmost important design factor such as wireless PC cameras, mobile camera phones, and sensor networks, a new video coding paradigm claiming low complex video encoding would be extremely valuable. Distributed source coding is one of the solutions for the low complexity encoding problem.

Distributed source coding is based on the Slepian-Wolf [1] and the Wyner-Ziv theorems [2] which have been proposed in 1970s'. These two theorems state that separate encoding but joint decoding of two correlated sources is just as efficient as the joint optimal encoding and decoding. Several video coding techniques based on this idea have come out recently, and they are, as a whole, known as distributed video coding (DVC). It can reduce the computational complexity burden of encoder by performing motion estimation only at decoder: actually, generation of side information at decoder is similar to motion estimation at encoder to utilize temporal correlation among frames.

The Wyner-Ziv coding is one of the representative DVC schemes [3], [4]. In the Wyner-Ziv coding, decoder reconstructs video by correcting noise on side information. Since the amount of parity bits depends on the noise in the side information, good quality of side information is necessary for better rate-distortion (R-D) performance. But in fact, it is occasionally difficult to generate side information of good quality without any original information if there are complex motions or occlusions among key frames. Therefore, in this case, helpful information on original Wyner-Ziv frame, typically called as hash is greatly desired by decoder to make a high quality of side information. In addition, it is also important, in the R-D sense, to send a just right amount of hash that is necessary to construct the side information.

In this paper, we propose a method to transmit as small hash data as possible without degrading coding efficiency. For that reason, we propose that decoder informs encoder where the hash transmission is required. This scheme fits well with DVC paradigm since the encoder does not need to do additional process to select where the hash is required. Besides, in the proposed method, encoder removes redundancy in hash data for improving coding efficiency. Generally, most information in the block is concentrated in low frequency, therefore, we propose to use the DC coefficient as hash. The hash generator transmits entropy coded hash to encoder to reduce redundancy in hash data, minimizing the transmission rate. This paper is organized as follows. Section 2 shows the overall structure of the proposed hash-based transform domain Wyner-Ziv coding. Section 3 describes in detail the proposed hash selection method and hash-based motion estimation. 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 (KOSEF) NRL Program grant funded by the Korea government (MEST) (ROA-2006-000 -10826-0(2008))

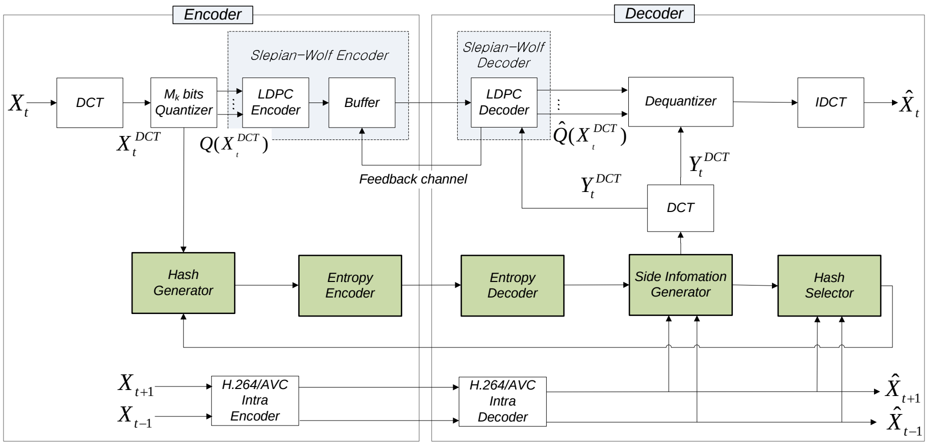

Fig. 1: Structure of the proposed hash-based Transform Domain Wyner-Ziv coding

2. PROPOSED WYNER-ZIV VIDEO CODEC

Fig. 1 illustrates the structure of the proposed selective hash-based transform domain Wyner-Ziv coding. (SH-TDWZ). The hash-based motion estimation modules such as hash generator, entropy encoder, entropy decoder, side information generator and hash selector are newly proposed on top of our previous development [8].

2.1 ENCODING PROCESS

The encoding process starts by separating the video frames into key and Wyner-Ziv frames. The key frames are encoded using the H.264/AVC intra encoder. If there is a hash transmission request from decoder, hash generator starts generating hash information of the spatially sub-sampled locations according to a checkerboard pattern and then, the hash is entropy encoded to be sent to decoder. The Wyner-Ziv frames are transformed by the 4x4 integer DCT transform and the grouped DCT coefficients of the same frequency over a WZ frame form whole 16 DCT bands. Each DCT band is quantized and reordered into bitplanes, and then the generated bitplanes are sent to LDPC encoder [5]. The LDPC encoder generates the parity bits in order from the most significant bitplanes to the least significant bitplanes. Then, the parity bits are transmitted to decoder when the parity request from decoder exists on feedback channel.

2.2 DECODING PROCESS

First of all, key frames are reconstructed as receiving bits stream from H.264/AVC intra encoder. Then, decoding a WZ frame is started. The first step in decoding WZ frame is to select where the hash bits is needed to generate accurate side information. Each frame is evaluated whether hash transmission is necessary or not by hash selector. If decoder makes the decision that hash is necessary due to complex motions or occlusion, hash generator transmits DC hash. In the other case when hash selector decides that hash is not needed, then the frame is interpolated with only the key frames assuming linear motion.

Once side information is generated, decoder estimates statistical property of the channel noise that is usually modeled as Laplacian. However, most of the channel noise on the side information is not stationary due to motions among frames. Therefore, decoder needs to estimates channel noise model [6]. It is important to make the estimated channel noise model similar to the real one. If the difference between them is small, the LDPC decoder can request less parity bits. It tells that there are two factors to improve R-D performance of system: one is the accuracy of channel noise model and the other is quality of the side information. The proposed method targets for improving the quality of side information, whose detail is described in next section.

3. PROPOSED HASH-BASED PROCESSING

Hash-based motion estimation [7] generates side information of better quality than the linear frame interpolation, especially, when there are complex motions among frames. However, hash is not always good in R-D sense since the consequential increment in transmission rate can annihilate both the quality gain of the reconstructed frame and the reduction in requested parity bits. Therefore, it is important for encoder to transmit hash only when the overall performance gain is expected. In this context, the proposed method reduces the hash data in two ways. First, it selects frames that definitely need hash for constructing the side information of required quality. Second, it removes statistical and spatial redundancy of hash data.

3.1 SELECTIVE HASH TRANSMISSION

Since hash transmission is not beneficial when linear frame interpolation works well, it is important for decoder to estimate whether the motion between key frames is sufficiently linear or not. To do this, once key frames are reconstructed, forward motion search is performed between the two key frames. From the estimated forward motion vector between the key frames, bi-directional motion vector of current frame is estimated assuming linear

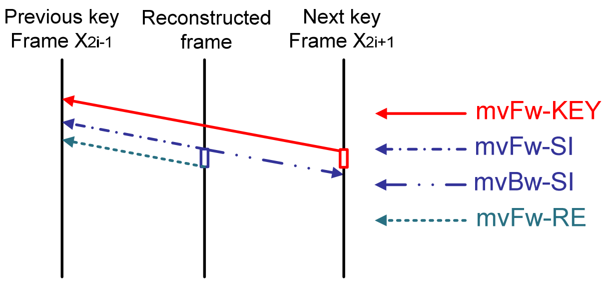

Fig. 2: Hash frame decision scheme

motion between the two key frames [8]. As a result, both key blocks are interpolated and side information is generated. If a real motion between key frames is nearly linear and estimated motion vector from linear frame interpolation is similar to the original one, motion vector estimated from the motion estimation based on side information to key frames should be similar to the one from linear frame interpolation. Otherwise, motion vector from the linear frame interpolation should be much different. Observing this, we propose that decoder decides whether to request hash bits or not from encoder using following logic.

$$If\ N\left\{ mvFw\text{-}RE = \frac{mvFw\text{-}KEY}{2} \right\} < THD, transmit\ hash$$

where $N(x)$ is the number of occurrence of event x; *mvFw-RE* is the forward motion vector from a block in the reconstructed frame to the previous key frame, and *mvFw-KEY* is the forward motion vector from a block in the next key frame to the previous key frame.

In Fig.2, *mvFw-SI* indicates the forward motion vector from a block in side information to the previous key frame, and *mvBw-SI* is the backward motion vector from a block in side information to the next key frame.

3.2 HASH-BASED MOTION ESTIMATION

To improve the coding efficiency, encoder has to remove redundancy in hash data. Usually, video data has three types of redundancy: temporal, spatial, and statistical redundancy. The proposed method removes the redundancy in hash data as following. Note that in DVC paradigm, encoder does not extract correlation among sources (that is, temporal redundancy). To simplify the redundancy reduction process at encoder, the temporal redundancy in hash data is not reduced, and only the black blocks in the checkerboard depicted in Fig. 3 are regarded as candidate blocks applied to block matching algorithm such as SAD to do motion estimation. The statistical redundancy in hash data is removed by entropy coding.

For the motion estimation, the hash which undergoes reduction of spatial redundancy is used for the candidate blocks. This hash is a DC coefficient of 4x4 integer DCT transform block. More accurate motion estimation desires more candidate blocks and the macroblock size becomes bigger. A macroblock of size 16x16 consists of 16 4x4 blocks. Only a half of 16 blocks in a macroblock is

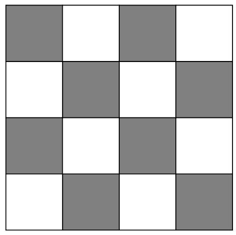

Fig. 3: 16x16 macroblock

selected as candidates for hash for motion estimation. After decoder decides that hash is necessary due to complex motion or occlusion, side information is constructed by the bidirectional hash-based interpolation. In this case, although half of DC coefficients are used for constructing the side information, the PSNR of side information is still good compared to the constructed side information without hash. This is because decoder already makes a decision that it is hard to make side information of a good quality.

4. EXPERIMENT AND RESULTS

In order to evaluate the performance of the proposed scheme, the proposed scheme is compared to that of DISCOVER codec [9]. For the test sequences, all frames of "Foreman", "Stefan", "Hall monitor" and "Football" sequences are QCIF@15Hz. In the experiment, different quantization [10] is applied to obtain eight rate-distortion points depicted in Table 1. For the frame interpolation, block size is set as 16x16, search range is set as ±32, and refinement range is set as ±4.

Fig. 4 shows R-D performances of Foreman, Stefan, Hall monitor and Football sequences. Except for the Hall monitor sequence which has nearly linear motion, the other sequences show overall PSNR gain up to 1 dB and in experiment, 30~90% of WZ frames end up with requiring hash when the side information is constructed. For the Hall monitor sequence having nearly linear motions, the proposed method does not require much hash data, since linear frame interpolation generates side information of sufficiently good quality. Therefore, there is no rate increment coming from the proposed hash transmission.

At the low bitrate case, the proposed scheme cannot guarantee improved R-D performance. It is because that increasing bits for additional hash data is relatively too many to improve PSNR in the rate.

Table 1: Quantization parameter setting for key frame

sequence		Foreman	Stefan	Hall monitor	Football
Qm	1	40	40	37	40
	2	39	39	36	39
	3	38	38	36	38
	4	34	34	33	34
	5	34	34	33	34
	6	32	32	31	32
	7	29	29	30	29
	8	25	25	25	25

* Qm : quantization matrix number for WZ frame

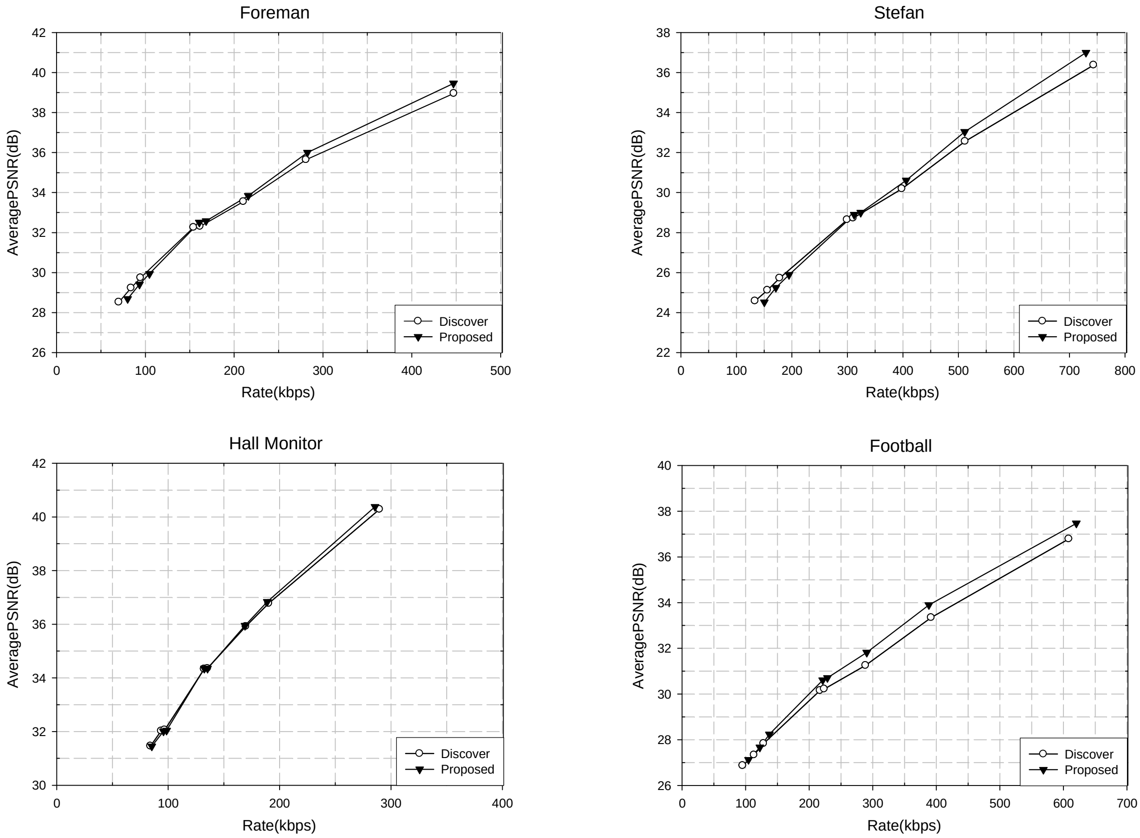

Fig. 4: R-D performance comparison (Proposed : SKKU-TDWZ[8] + Selective Hash)

5. CONCLUSIONS

In this paper, we proposed a TDWZ codec with selective hash transmission and hash-based motion estimation improving R-D performance. The hash frame selection method does not increase computational complexity in encoder because it is conducted at decoder side. Therefore, this concept fits well with DVC that tries to reduce the burden of encoder. When the frame is determined to use hash, the motion estimation is executed using the selected blocks on block checkerboard. For the sequences having non-linear motion, our proposed scheme shows the average PSNR gain up to 1dB. As a future work, we are going to study more improved hash-based motion estimation method using smaller macro block size with less hash. Decision measure is also need to be applied per block instead of frame.

6. REFERENCES

[1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. IT-19, 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 Trans. Information Theory, vol. 22, pp. 1-10, July 1976.

[3] 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, March 2003.

[4] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video," Proc. VCIP 2004, January 2004.

[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] 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, Nov. 2005.

[7] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," IEEE International Conference on Image Processing, Oct. 2006.

[8] J. Ascenso, F. Pereira, "Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding" International Conference on Image Processing (ICIP), San Antonio, USA, September, 2007.

[9] B. Ko, H. J. Shim, and B. Jeon, "Wyner-Ziv coding with spatio-temporal refinement based on successive turbo decoding," IEEE Intern. Conf. on Multimedia and Expo, Hannover, Germany, June 2008.

[10] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation", Picture Coding Symposium 2007, Lisbon, Portugal, 2007.

[11] C.Brites, J. Ascenso, F. Pereira, "Improving Transform domain Wyner-Ziv video coding performance," IEEE Intern. Conf. on Acoustics, Speech and Signal Processing, Toulouse, May 2006.