

CONSTANT QUALITY CONTROL OF WYNER-ZIV FRAMES IN DCT DOMAIN DISTRIBUTED VIDEO CODING

HONGBIN LIU¹, DEBIN ZHAO¹, SIWEI MA² AND WEN GAO²

¹School of Computer Science and Technology
Harbin Institute of Technology
No. 92, West Dazhi Street, Harbin 150001, P. R. China
{hbliu; dbzhao}@jdl.ac.cn

²School of Electronics Engineering and Computer Science
Peking University
No. 5, Yiheyuan Road, Haidian District, Beijing 100871, P. R. China
{swma; wgao}@jdl.ac.cn

Received September 2010; accepted December 2010

ABSTRACT. For most DCT domain distributed video coding systems, fixed uniform quantizers are used to quantize all Wyner-Ziv (WZ) frames. However, it leads to fluctuant video quality because the WZ frame quality is also greatly influenced by the side information (SI) quality besides the quantizers. To address this issue, this letter proposes a new constant quality control strategy, which adaptively designs quantizers for each WZ frame according to its SI. We formulate the constant quality control problem into the adaptive quantization level (AQL) design problem, and propose a low complexity greedy algorithm to solve it. Experimental results demonstrate that the proposed algorithm decreases the quality fluctuation obviously and improves the PSNR (peak signal-to-noise ratio) by 0.79dB as well when compared with the state-of-the-art algorithm.

Keywords: Distributed video coding, Constant quality control, Greedy algorithm, Adaptive uniform quantizer

1. Introduction. Distributed video coding (DVC) is a new coding paradigm that provides very low complexity encoder. It is built upon the theoretic results of Slepian-Wolf [1] and Wyner-Ziv [2] theorems. A pioneer work of DVC was proposed by Aaron et al. [3,4], which is called the Stanford DVC framework. Another DVC framework known as PRISM [5] was proposed by Puri et al. The Stanford DVC framework is a very popular DVC framework, and it includes pixel domain DVC (PDVC) [3] and DCT transform domain DVC (TDVC) [4]. Since TDVC achieves significant improvement than PDVC, it becomes the most popular DVC framework and is improved by many other researchers [6,7].

In most traditional TDVC systems, fixed quantization level (FQL) uniform quantizers are employed to quantize DCT coefficients of all WZ frames. However, since the WZ frame quality is also greatly influenced by the SI quality besides the quantizers, it leads to fluctuant quality, especially on sequences containing large motions. This can hurt the video quality seriously. To address this issue, G. Wu et al. [8] proposed to reduce the SI quality fluctuation. However, it is difficult to control the SI quality quantitatively. The state-of-the-art algorithm was proposed by S. Sofke et al. [9], in which adaptive quantization levels (QLs) are designed for each WZ frame according to the SI. It computes average distortion of two adjacent key frames for each DCT band, and selects QL for each band of WZ frame independently according to its corresponding average distortion. Though [9] achieves smoother WZ frame quality than [4], there are two limitations. First, fluctuant key frame quality leads to fluctuant WZ frame quality since the latter is dependent on the former. Second, each band is controlled to respective target distortion independently, which brings accumulated control error.

To overcome the above limitations, this letter proposes an improved constant quality control strategy, in which quality control of WZ frame is independent of key frame quality and distortions of all DCT bands are controlled together. In the proposed algorithm, we model the constant quality control problem of WZ frame into an AQL design problem, and propose a greedy algorithm to solve it.

The rest of this paper is organized as follows. In Section 2, constant quality control of WZ frame is formulated into AQL design problem, and the greedy algorithm is applied to solve it. In Section 3, AQL design algorithm is implemented in the TDVC codec. Section 4 presents the experimental results. Finally, Section 5 concludes this letter.

2. Problem Formulation of Constant Quality Control of WZ Frame. In this section, constant quality control of WZ frame is formulated into AQL design problem and the greedy algorithm is applied to solve it.

In TDVC, when encoding a DCT band, the consuming rate depends on the quantizer, the SI and the employed channel code; while the distortion depends on the quantizer and the SI. Consequently, the rate and distortion functions of the i th band can be written as $R_i(Q_i, SI_i, C_i)$ and $D_i(Q_i, SI_i)$, where Q_i , SI_i and C_i are the quantizer, SI and channel code of the i th band respectively. After defining the rate and distortion functions, the constant quality control problem can be formulated as:

$$\begin{aligned} & \min \sum_i R_i(Q_i, SI_i, C_i) \\ & \text{subject to } \sum_i D_i(Q_i, SI_i) = D_{tar} \quad i = 1 \dots k, \end{aligned} \quad (1)$$

where D_{tar} is the target distortion of WZ frame and k is the total bands number.

Problem in (1) is difficult to solve since there are so many variables. Fortunately, it can be simplified if some prerequisites are determined. First, the quantizer is fixed to uniform quantizer as most TDVC researchers do. Because uniform quantizer is only determined by its quantization step (QS) if the deadzone width is decided, Q_i in (1) can be replaced by its QS or its QL. Second, the SI generation algorithm is preselected. This makes SI constant since SI depends only on already decoded adjacent key frames which are determinate, given the SI generation algorithm. Finally, turbo code is chosen as the channel code and is kept unchanged. From the above analysis, we can see that SI_i and C_i are constant and Q_i is defined by its QL (which is denoted as QL_i). Therefore, $R_i(\cdot)$ and $D_i(\cdot)$ depend only on QL_i and (1) can be simplified to AQL design problem as:

$$\begin{aligned} & \min \sum_i R_i(QL_i) \\ & \text{subject to } \sum_i D_i(QL_i) = D_{tar} \quad i = 1 \dots k. \end{aligned} \quad (2)$$

To solve the AQL design problem, greedy algorithm is proposed as an alternative of dynamic programming algorithm [10], since computational complexity of the latter can grow exponentially with the band number, which is too complex for the TDVC encoder.

The input of the greedy algorithm is the distortion rate (DR) points of all DCT bands and the target distortion, while the output is QLs of all bands. The solution procedure is as follows. First, connect adjacent DR points to form the DR curve (X-axis is distortion; Y-axis is rate) for each band. For each DR curve, we start from the point with the largest QL and thus the maximum rate. Then, we repeatedly decrease the total rate by decreasing QL of certain band, as long as the total distortion is no larger than the target distortion. In each repeat stage, first, for each curve, slope of the line connecting current point and the next point (with smaller QL than current point) is calculated. Then, the curve with the largest minus slope (meaning of minus slope is the (rate decrease)/(distortion increase))

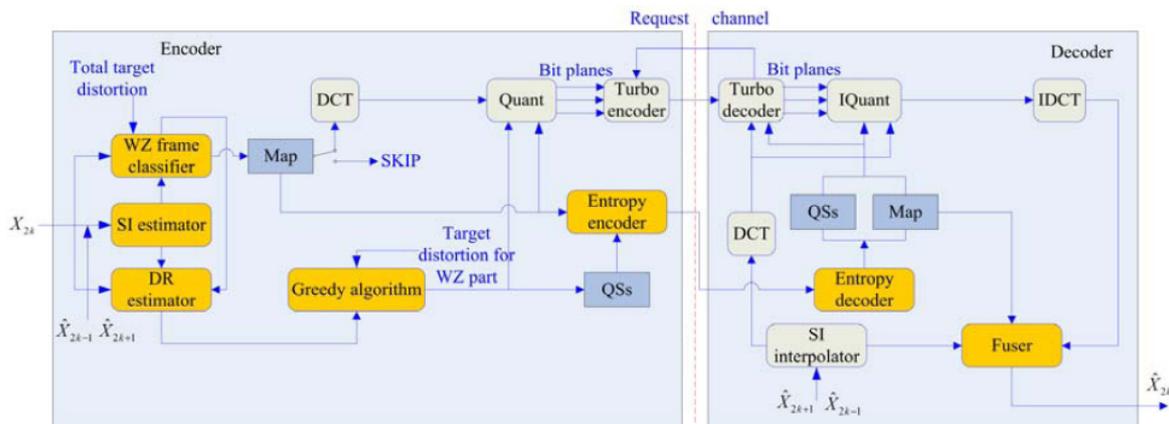


FIGURE 1. Proposed constant quality control framework of WZ frame

ratio) is firstly chosen as the candidate curve and it steps forward to the next point. If the resulted total distortion is still no larger than the target distortion, the next stage starts. Otherwise, the candidate curve steps back, and the curve with the second largest minus slope is chosen as the candidate curve. This process continues until a certain curve steps forward or all curves are tested. The stage is repeated until no further QL decrease is allowed.

In the proposed greedy algorithm, if there are i DR points in one curve, we only need to deal with at most $i-1$ slopes for it. Consequently, the computational complexity increases linearly with total DR points, which is much simpler than the dynamic programming algorithm.

3. Proposed Constant Quality Control Framework of WZ Frame. In this section, the above AQL design algorithm is implemented into practical TDVC codec. Figure 1 illustrates the proposed constant quality control framework of WZ frame. As stated in Section 2, DR points of all bands are required in the AQL design algorithm. However, the real DR points are unavailable since the real decoding process is not performed at the encoder. Therefore, we have to estimate the DR point which requires estimating the SI, the rate and the distortion of each band. Estimation methods are not described here since they are not the focus of this letter; interesting readers can refer to [11,12] for details. We will mainly discuss the encoding and decoding procedures of WZ frames in the following.

At the encoder, first, the estimated SI is generated. Then, the WZ frame is classified (by the WZ frame classifier in Figure 1 into two block sets according to the estimated SI: skipped part and WZ part. The skipped part is not encoded and the WZ part is WZ encoded. The skipped part is introduced because that SI quality of some blocks is already higher than a threshold (which relates to the target distortion) and thus these blocks need not to be encoded. To distinguish these two parts, a binary classification map is generated according to:

$$map_i = \begin{cases} 1 & MSE_i < \alpha \cdot MSE_{tar} \\ 0 & \text{else} \end{cases} \quad i = 1, 2, \dots, n. \quad (3)$$

Here, MSE_i denotes MSE (mean squared error) between the original value and the estimated SI of the i th block. MSE_{tar} is the average target MSE of one block, which is derived from the target distortion. n is the total block number in a WZ frame. map_i indicates the encoding mode of the i th block, and $map_i = 1$ and 0 denote the skipped mode and WZ mode respectively. α is a scale factor controlling number of the skipped blocks, which is set to 1.5. The block size in classification is set to 16×16 to balance the classification accuracy and the map rate.

Afterwards, DR points of all allowable QLs (the largest allowable QL is set to 256 for all bands because the larger QL is seldom used) of each DCT band is estimated for the WZ part. Meanwhile, the target distortion of the WZ part is also calculated:

$$tar_dis_{WZ} = tar_dis - dis_{skip}. \quad (4)$$

Here, tar_dis_{WZ} and tar_dis are the target distortion of the WZ part and the WZ frame respectively. dis_{skip} is the estimated distortion of the skipped part, i.e., distortion of the estimated SI for the skipped part. Then, the proposed AQL design algorithm is used to determine QLs of all DCT bands for the WZ part.

Finally, the encoding procedure starts, in which the classification map, the QLs and the WZ part are encoded. A run-length entropy coder is used to encode the classification map. The QL in this letter is confined to numbers that are power of 2 as [4] did, thus only $\log_2(QL)$ needs to be transmitted. Since QLs consumes only few bits, they are binary encoded directly. The WZ part is encoded by traditional WZ frame encoder but using the adaptively devised QLs.

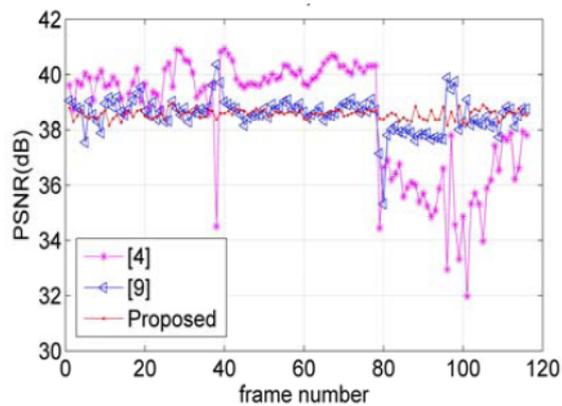
At the decoder, the SI is first generated and then the classification map and QLs are decoded. Afterwards, the WZ part is decoded and reconstructed by traditional WZ frame decoder, according to the classification map and QLs. Finally, the fuser in Figure 1 combines the WZ part and SI of the skipped part to form a complete WZ frame.

4. Experimental Results. In this section, we evaluate the proposed algorithm objectively in terms of PSNR variance and rate distortion (RD) performance.

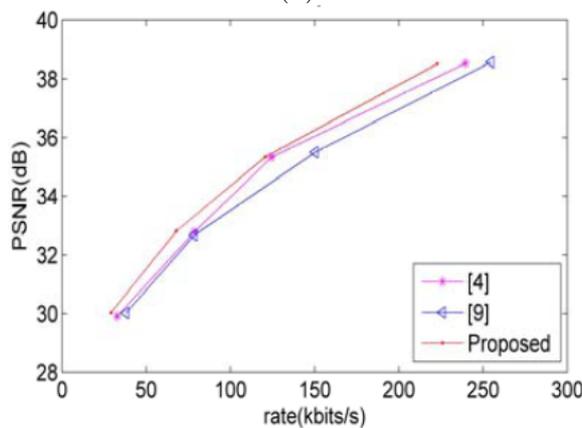
FQL based TDVC [4] and WZ frame quality control algorithm in [9] are selected as the benchmarks. In the experiment, three QCIF (176×144) 30Hz sequences are used. Instead of testing each sequence separately, we concatenate the first 150, 165 and 150 frames of *Foreman*, *Hall Monitor* and *Coastguard* respectively to compose a concatenated sequence as [9] did. The concatenated sequence includes both low and high motions, and it also contains scene changes. Therefore, it can test robustness of the proposed algorithm very well.

GOP (group of pictures) size is set to 2. In each GOP, the first frame is encoded as key frame and the second frame is encoded as WZ frame. The encoding rate is 15Hz. For encoding key frames, H.264/AVC reference software JM 13.2 [13] with main profile is used, and key frame quality control algorithm in [9] is also employed. Key frames and WZ frames are encoded into four quality levels. In each quality level, same target quality is used for both key frames and WZ frames for [9] and the proposed algorithm; while for [4], QLs of WZ frames are adjusted manually to achieve similar average quality to key frames. When comparing RD performance, the tool proposed in VCEG-AE07 [14] is used.

Since key frames are the same for all methods, we only compare PSNR variance and RD performance of WZ frames. Table 1 compares the PSNR variance and average PSNR of different methods. We can see that these three methods achieve similar average PSNR; however, the proposed algorithm achieves much smaller PSNR variances, i.e., much smoother quality than other two methods. For giving an intuitive comparison, Figure 2(a) presents PSNR of all WZ frames at quality level 1. It can be observed that the gap between the highest and lowest PSNR is substantially decreased by the proposed algorithm. Besides, the lowest PSNR is also significantly improved. Since the video quality is usually influenced by the lowest quality, the proposed algorithm can bring obvious improvement. Advantages of the proposed algorithm mainly attribute to two reasons: first, distortions of all DCT bands in the WZ frame are controlled together; therefore, the accumulated control error in [9] is avoided. Second, quality control of the WZ frame is independent of key frame quality, thus fluctuant key frame quality will not bring additional quality fluctuation to WZ frame. RD performance is also compared in Figure 2(b).



(a)



(b)

FIGURE 2. PSNR and RD performance comparison of different methods, (a) PSNR of all WZ frames at quality level 1, (b) RD curves

TABLE 1. PSNR variance and average PSNR of WZ frames for different methods

Quality Level		1	2	3	4
Variance	[4]	4.4751	4.2198	3.6455	3.0800
	[9]	0.3597	0.3946	0.2214	0.2243
	Proposed	0.0158	0.0332	0.0264	0.0351
PSNR(dB)	[4]	38.53	35.33	32.78	29.91
	[9]	38.57	35.50	32.67	30.02
	Proposed	38.51	35.34	32.81	30.03
	Target	38.50	35.50	32.80	30.00

We can see that the proposed algorithm is a little better than [4]; however, it improves the PSNR by 0.79dB (decreases the bit rate by 17%) when compared with [9].

5. Conclusions. In this paper, we have proposed an improved WZ frame quality control algorithm for TDVC to overcome limitations of existing quality control algorithms. The constant quality control problem of WZ frame is modeled into AQL design problem and the low complexity greedy algorithm is applied to solve it. Experimental results have clearly demonstrated advantages of the proposed algorithm.

Acknowledgment. This work is partially supported by National Science Foundation (60736043) and National Basic Research Program of China (973 Program, 2009CB320905). The authors would like to thank Dr. Bo Wu for providing FQL based TDVC software.

REFERENCES

- [1] D. Slepian and J. Wolf, Noiseless coding of correlated information sources, *IEEE Trans. on Information Theory*, vol.22, pp.471-480, 1973.
- [2] A. D. Wyner and J. Ziv, The rate-distortion function for source coding with side information at the decoder, *IEEE Trans. on Information Theory*, vol.22, pp.1-10, 1976.
- [3] A. Aaron, R. Zhang and B. Girod, Wyner-Ziv coding of motion video, *Proc. of Asilomar Conference on Signals and Systems*, Pacific Grove, CA, 2002.
- [4] A. Aaron, S. Rane, E. Setton and B. Girod, Transform-domain Wyner-Ziv codec for video, *Proc. of SPIE Visual Communications and Image Processing*, San Jose, CA, 2004.
- [5] R. Puri and K. Ramchandran, PRISM: A new robust video coding architecture based on distributed compression principles, *The 40th Allerton Conference on Communication, Control and Computing*, Allerton, IL, 2002.
- [6] T. C.-W. Lei and S.-J. Chern, New distributed video coding paradigm with partial boundary matching algorithm, spatio-temporal texture synthesis and image inpainting, *International Journal of Innovative Computing, Information and Control*, vol.6, no.12, pp.5405-5419, 2010.
- [7] R. Martins, C. Brites, J. Ascenso and F. Pereira, Refining side information for improved transform domain Wyner-Ziv video coding, *IEEE Trans. on Circuits and Systems for Video Technology*, vol.19, pp.1327-1341, 2009.
- [8] G. Wu, L. Sun and F. Huang, Consistent-quality distributed video coding framework, *International Conference on Image Processing*, Singapore, 2004.
- [9] S. Sofke, F. Pereira and E. Muller, Dynamic quality control for transform domain Wyner-Ziv video coding, *Journal on Image and Video Processing*, vol.2009, no.5, pp.1-15, 2009.
- [10] A. Ortega and K. Ramchandran, Rate-distortion methods for image and video compression, *IEEE Signal Processing Magazine*, vol.15, no.6, pp.23-50, 1998.
- [11] D. Kubasov, J. Nayak and C. Guillemot, Optimal reconstruction in Wyner-Ziv video coding with multiple side information, *Workshop on Multimedia Signal Processing*, Crete, Greece, 2007.
- [12] C. Brites and F. Pereira, Encoder rate control for transform domain Wyner-Ziv video coding, *International Conference on Image Processing*, San Antonio, TX, 2007.
- [13] http://iphome.hhi.de/suehring/tml/download/old_jm/.
- [14] S. Pateux and J. Jung, An excel add-in for computing Bjøntegaard metric and its evolution, *ITU-T SG16 Q.6 Document, VCEG-AE07*, Marrakech, 2007.