IMPROVED HEVC LOSSLESS COMPRESSION USING TWO-STAGE CODING WITH SUB-FRAME LEVEL OPTIMAL QUANTIZATION VALUES

Xun Cai, Student Member, IEEE*

Research Laboratory of Electronics
Department of EECS
Massachusetts Institute of Technology
Cambridge, MA, 02139
cx2001@mit.edu

Qunshan Gu

Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
qunshan@google.com

ABSTRACT

Lossless video coding is used when perfect preservation of video data is required. In HEVC, lossless coding is accomplished by bypassing the transform and quantization stages. Prediction residuals are coded with the entropy coder in the spatial domain .

In this paper, two-stage coding with sub-frame adaptive quantization is proposed. The DCT is firstly applied to the prediction residuals and the DCT coefficients are quantized. The quantized DCT coefficients and the quantization error are coded. Adaptive quantization parameters are used for each Coding Unit. Simulation results show that the proposed method significantly outperforms the HEVC lossless coding.

Index Terms— Lossless coding, HEVC, two-stage coding, adaptive quantization

1. INTRODUCTION

Lossless video coding is used when perfect preservation of video data is required. It is useful in many video compression applications. For example, lossless coding can be used in archiving important videos. In video editing, lossless coding can prevent accumulation of quantization error in repeated encoding and decoding operations. In addition to only maintaining visual perception, lossless coding is useful in preserving numerical data, such as in medical applications and preserving watermark information in videos.

In many video and image compression systems, different approaches are used in lossless coding and lossy coding. In lossy coding, it is important to represent a video with a small number of bits, while maintaining an acceptable visual quality. It is well known that typical video data can be approximated fairly accurately in the Discrete Cosine Transform (DCT) domain [1], and the DCT is extensively used in lossy video compression. In lossless coding, on the other hand, it is important to preserve the numerical video data with fewer bits. In this case, the DCT cannot be applied in a straightforward manner. The DCT coefficients are float-point numbers and have to be quantized. In [2], it is verified that the quantization error is introduced even when the smallest quantization parameter is used. The quantization error is against the perfect preservation requirement.

To replace the DCT for lossless coding, many approaches have been proposed. One approach is to use integer transforms. For example, in JPEG 2000 image compression standard, 5/3 reversible wavelet transform [3] is used. In VP9 video compression standard [4], Walsh-Hadamard transform is used in lossless coding. In addition, many other types of integer transforms are proposed [5][6].

Another approach is based on spatial-domain coding. The simplest spatial-domain approach, for example, is to code the prediction residuals without transform and quantization. This method is adopted in H.264/AVC [7][8] and HEVC [9] video compression systems, by bypassing the transform and quantization steps [2]. The entropy coder is directly applied to the intra/inter prediction residuals. Another more sophisticated spatial-domain approach is based on the DPCM. In these DPCM-based methods, a pixel is predicted from its surrounding causal pixels. For example, in the latest H.264/AVC reference software, vertical and horizontal DPCM is used for intra prediction residuals. In the latest HEVC reference software with Range Extensions, directional DPCM is used in encoding motion-compensated residuals. Other DPCM-based systems can be found in [10][11][2].

The third approach is based on the two-stage coding. In the two-stage coding, video signals are separated into two layers and encoded. One example implemented on H.264/AVC can be found in [12] where lossy coding is first applied, followed by lossless coding of the coding error. In addition, we refer to [13] for other two-stage coding and image processing applications.

In this paper, we explore two-stage coding in HEVC lossless coding. One fundamental issue in two-stage coding is how to separate video signals into two parts that lead to an efficient representation of video signals. In this paper, sub-frame adaptive quantization is proposed. Simulation results show that this method significantly improves the HEVC lossless coding. We note that there are two major differences between the proposed method and the one reported in [12]. First, the work in [12] uses frame-level separation while we use block-level separation. Second, block-level adaptive quantization is used in the proposed method.

This paper is organized in the following way. In Section 2, lossless coding in HEVC is discussed. In Section 3, sub-frame adaptive quantization two-stage lossless coding is proposed. In Section 4, the proposed method is implemented on a modified HEVC reference model. Simulation results show that the proposed method significantly outperforms HEVC lossless coding. In Section 5, we summarize the paper and discuss future work.

*Xun Cai was a summer intern at Google Inc. when this work was done.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

2. OVERVIEW OF HEVC LOSSLESS CODING

In HEVC coding standard, lossless coding is accomplished by coding the intra/inter prediction residuals directly in the spatial domain. In lossless coding, the HEVC codec follows the Coding Unit(CU)/Prediction Unit(PU) split and intra/inter prediction routines, and prediction residuals are obtained. To encode these residuals, transform and quantization steps are bypassed. The residual signals are coded directly using the Context Adaptive Binary Arithmetic Coder (CABAC). The block diagram shown in Figure 1 illustrates this procedure. One advantage of HEVC lossless coding strategy is computational efficiency, when the computationally intensive DCT is bypassed. Another advantage is its consistency with HEVC lossy coding structure.

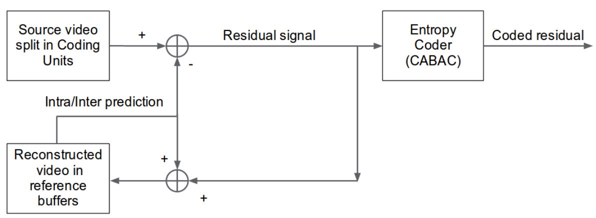

Fig. 1. HEVC lossless coding

However, the coding performance is compromised in this implementation. The HEVC entropy coder is designed for encoding blocks of quantized DCT coefficients, which are assumed to be sparse and uncorrelated. Specifically, a block of quantized DCT coefficients contain many zero coefficients, which can be efficiently encoded using coefficient scanning and significant coefficients map [9]. For a typical block, only a few large significant non-zero coefficients require a large number of bits. For a typical block of residuals in the spatial domain, however, it is usually not sparse and correlation among residual intensities is strong. In addition, the statistics of the quantized DCT coefficients is significantly different from that of residuals. As a result, the HEVC CABAC may not be efficient in encoding residual intensities in the spatial domain.

To improve the HEVC lossless coding, one solution is to design the entropy coder specifically for encoding spatial domain intensities. This requires a careful study of residual statistics in the spatial domain, and the entropy coder may become very complicated. Another solution is to use better representations of residual signals in favor of the entropy coder. We observe that the HEVC CABAC is state-of-art in coding blocks of quantized DCT coefficients. In addition, the HEVC CABAC is efficient in encoding uncorrelated coefficients with small amplitudes. This observation leads to the method discussed in the following sections.

3. TWO-STAGE LOSSLESS CODING WITH SUB-FRAME ADAPTIVE QUANTIZATION

3.1. Two-stage lossless coding

The proposed approach separates a block of residual signal into two parts through quantization. The first part is quantized DCT coefficients and the second part is the quantization error. To be specific, the DCT coefficients of prediction residuals are first obtained. These DCT coefficients are quantized as the DCT block. The quantized coefficients are used to reconstruct a lossy decoded block and the it is subtracted from the residual block. This quantization error is encoded as the spatial block. By choosing an appropriate quantization parameter for each block, both blocks can be coded efficiently using HEVC entropy coder. The block diagram shown in Figure 2 illustrates this idea.

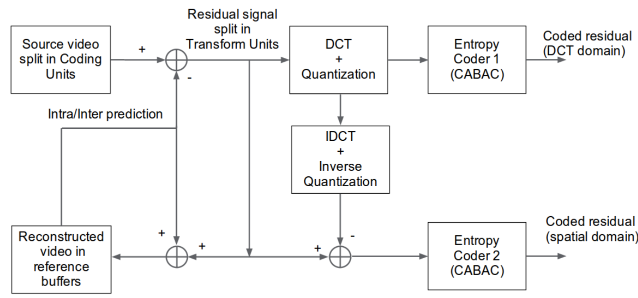

Fig. 2. Two-stage lossless coding

To understand why this approach may lead to a more efficient representation of the signal, we consider the energy distribution between the spatial domain and the DCT domain. Suppose a large quantization parameter is used. Since typical residual signal in the DCT domain only consists of a small number of non-zero large coefficients, a larger quantization parameter will discard more non-zero small DCT coefficients in the DCT domain. In this case, the energy compaction of the DCT can be better utilized and the DCT block can be efficiently coded. On the other hand, suppose a small quantization parameter is used. This results in a smaller quantization error. In addition, one can show that with a small enough quantization parameter, the spatial block can be nearly uncorrelated. These facts indicate that the spatial block can be coded efficiently when a small quantization parameter is used. In practice, the number of bits spent on both blocks may be smaller, if the energy is distributed to both domains in a correct way, relative to coding in a single domain. The correct energy distribution between two blocks is achieved by selecting a proper quantization parameter for the block to be coded. The effect of the quantization parameter on the coding efficiency is shown in more detail in Section 4.

3.2. Sub-frame adaptive quantization selection

As discussed above, the best coding performance can be achieved if a correct quantization parameter is chosen. In practice, the best quantization parameter may depend on the characteristics of the DCT coefficients. For example, when the block is smooth, we may choose a small quantization parameter so that most energy is distributed in the DCT domain. In contrast, when the block is noisy, we may choose a large quantization and most energy is distributed in the spatial domain. As a result, it is necessary to use adaptive quantization parameters for blocks with different characteristics. In this paper, this is accomplished by allowing a different quantization parameter for each CU.

Table 1. Coded block flag for two-stage coding

DCT block	Spatial block	Coded block flag
Zero	Zero	0
Non-zero	Non-zero	11
Zero	Non-zero	100
Non-zero	Zero	101

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

4.1. Implementations

Our proposed method is implemented on the HEVC reference software 6.3. To be specific, a quantization parameter is chosen for each CU. Prediction residuals are obtained using HEVC intra/inter prediction. The residual in this CU is split into a quad-tree of Transform Units (TU), where the DCT is applied. The DCT coefficients of a TU are quantized using the chosen quantization parameter, and the quantized coefficients are coded. Encoding of quantization error in the same TU follows. Both the DCT and the spatial blocks are coded with the CABAC. At the decoder side, the quantization error is added to the lossy reconstruction from the quantized DCT coefficients to form the lossless reconstruction.

The coded block flags for both parts are modified accordingly. The coded block flag of each TU is set to zero only when both DCT and spatial blocks are zero. In the case where either block is non-zero, the following strategy applies. A one-bit flag is used to specify whether both blocks are non-zero. If only one-block is non-zero, one additional one-bit flag is used to specify whether block is non-zero. This is illustrated in Table 1.

In adaptive quantization, a base quantization parameter for each frame is signaled in the frame header. The delta quantization parameter is signaled at the CU level and coded using the HEVC variable quantization parameter coding. The quantization parameter for each CU can be selected in many ways. In our simulations, the delta quantization parameters range from -7 to +7 and the quantization parameter is chosen by exhaustive search. By doing this, we are able to see the highest coding gain possible from the adaptive quantization. This exhaustive search, however, will increase the encoder complexity. We note that in practice, it can be replaced with other methods (e.g. binary search) to accelerate the encoding process. Finally, the default lossless coding mode (Q=0) is included as complement to the two-stage coding. The performance of the modified system is compared against the latest HEVC 12.0 lossless coding.

4.2. Experimental results: two-stage lossless coding on a single TU

We first consider a block (TU) of residual signal obtained from a video sequence. This residual block is separated into two parts (DCT block and spatial block) using a quantization parameter of 15. The numerical intensities of three blocks are shown in Figure 3.

From this block, we observe that the original residual block has a relatively large energy and the intensities are highly correlated. On the other hand, we observe that two blocks after separation have desirable properties favored by the entropy coder. To be specific, the DCT block is sparse and the energy is concentrated in a very small number of non-zero coefficients. The spatial block has much smaller intensities and higher sparsity compared to the original residual block. As a result, both blocks may be coded efficiently by the entropy coder. The total number of bits spent on both blocks may be smaller than on a single residual block. This observation is true for many blocks in practical video sequences, if a correct quantization parameter is chosen. As a result, the proposed method may lead to a significant bit saving compared to the HEVC lossless coding. This is verified in the following subsections.

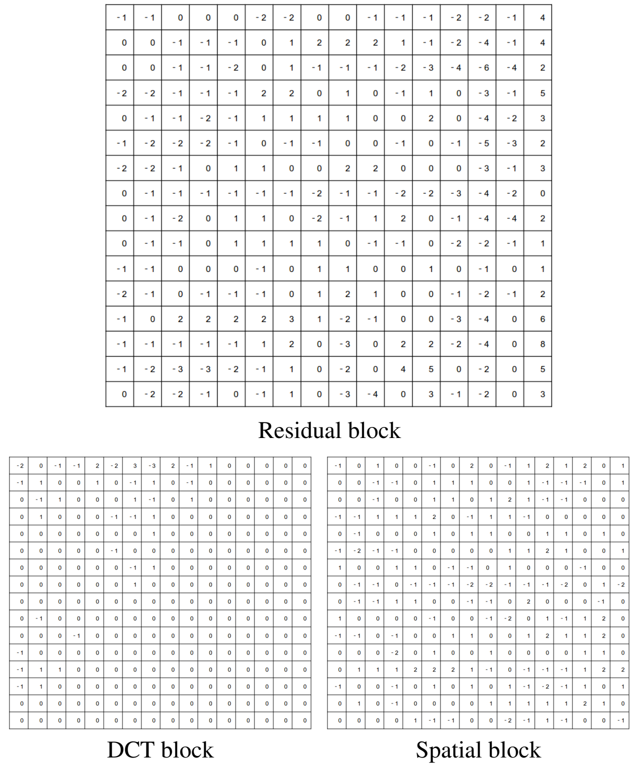

Fig. 3. Two-stage lossless coding of a single TU

4.3. Experimental results: single frame with a fixed Q

In this subsection, we encode the first frame in the "crew_1280_720" as an I frame, with a frame level Q. We plot the number of bits used to losslessly code the frame as a function of Q, using both the original HEVC and two-stage lossless coding. The result is shown in Figure 4.

As we can see in the results, two-stage lossless coding has a large coding gain compared to the original HEVC codec. The coding gain can be as high as 11.4% when the optimal quantization parameter (12) is used . From the result, the impact of the different quantization parameters on the energy distribution between two blocks can be observed. The number of bits first decreases and then increases as Q increases. This indicates that the energy is redistributed between the DCT domain and the spatial domain when Q changes. When Q is too small, most energy is distributed the DCT domain. In this case, the DCT blocks are not likely to be sparse and too many bits are wasted on the DCT block. When Q becomes too large, most energy is distributed in the spatial domain. The spatial intensities may be strongly correlated and large-valued. This prevents an efficient coding of the spatial blocks. The largest coding gain is achieved only when the tradeoff in both domains is correctly balanced.

Table 2. Profile Parameters

GOP structure	intra only	intra and inter
Profile	Intra main	Low delay main
# of frames per sequence	10	50

4.4. Experimental results: single frame with a variable Q

In this subsection, we allow sub-frame adaptive quantization. We plot the number of bits as a function of frame-level base Q and allow delta Q to vary from -7 to +7. Other experimental settings are the same as the previous subsection. The result is shown in Figure 4. As expected, when Q is adapted to the local characteristics of residual signals, additional coding gain can be obtained. In this case, the highest coding gain reaches as high as 12.4%, an additional 1.0% coding gain compared to using a frame-level Q.

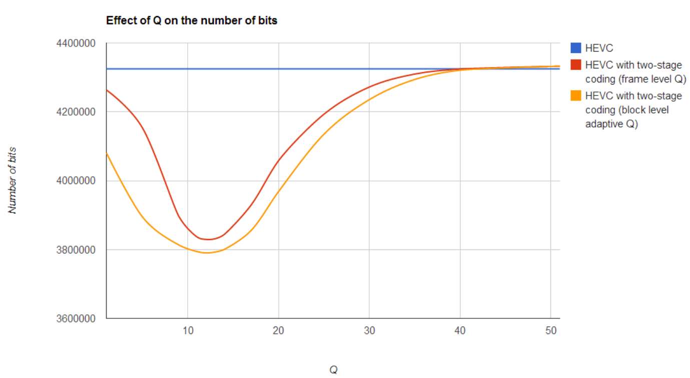

Fig. 4. Single I frame two-stage coded with variable Q

4.5. Experimental results: sequences with variable Q

To see the overall coding gain of our proposed method, we encoded a set of sequences in different resolutions using the two-stage coding with variable Q. The experimental parameters are listed in Table 2 and 3. The results are compared with the HEVC lossless mode, shown in Table 4.

From the results, two-stage lossless coding with sub-frame adaptive quantization outperforms the HEVC lossless coding. The coding gain for I frames is higher than for P/B frames. One explanation is that the correlation in intra prediction residuals is higher than in inter prediction residuals. As a result, intra blocks tend to benefit more than inter blocks from the two-stage coding. In addition, the coding gain for high resolution videos is higher than that for low resolution videos, due to a similar reason.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed to use sub-frame adaptive quantization in two-stage lossless coding. Simulation results show that the proposed method is an efficient representation of the residual signals in lossless coding. The proposed method can be implemented on other video coding systems in addition to HEVC. In addition, the idea of two-stage coding can be extended beyond quantizing the DCT coefficients. Other approaches that effectively separate the signal into two parts can be used. For example, we may detect whether impulsive values exist in a block and separate them from the rest smooth part. We also observe that the proposed method uses fewer bits than DCT-based lossy coding at a very high quality. As a result, the two-stage coding can be applied to high quality lossy coding applications with proper modifications.

Table 3. Sequence Parameters

Sequence	Resolution	Frame rate	Frame level Q
crew	1280x720	60	12
night	1280x720	60	14
mobcal	1280x720	60	18
cyclists	1280x720	60	12
gipsMotion	1280x720	60	13
BasketballDrill	832x480	50	20
BQMall	832x480	60	22
Flowervase	832x480	30	14
Mobisode2	832x480	30	13
PartyScene	832x480	50	24
RaceHorses	832x480	30	16

Table 4. Performance of Two-stage Coding

Setting	Average coding gain
720 Intra	12.14%
720 Intra+Inter	5.69%
480 Intra	4.48%
480 Intra+Inter	1.64%

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

6. REFERENCES

[1] Nasir Ahmed, T Natarajan, and Kamisetty R Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. 100, no. 1, pp. 90–93, 1974.

[2] Minhua Zhou, Wen Gao, Minqiang Jiang, and Haoping Yu, "HEVC lossless coding and improvements," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1839–1843, 2012.

[3] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi, "The JPEG 2000 still image compression standard," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 36–58, 2001.

[4] Jim Bankoski, Ronald S Bultje, Adrian Grange, Qunshan Gu, Jingning Han, John Koleszar, Debargha Mukherjee, Paul Wilkins, and Yaowu Xu, "Towards a next generation open-source video codec," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013.

[5] Wilfried Philips, "The lossless DCT for combined lossy/lossless image coding," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. IEEE, 1998, pp. 871–875.

[6] Ying-Jui Chen, Soontorn Oraintara, and Truong Nguyen, "Video compression using integer DCT," in *Image Processing, 2000. Proceedings. 2000 International Conference on*. IEEE, 2000, vol. 2, pp. 844–845.

[7] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[8] Gary J Sullivan, Pankaj N Topiwala, and Ajay Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Optical Science and Technology, the SPIE 49th Annual Meeting*. International Society for Optics and Photonics, 2004, pp. 454–474.

[9] Gary J Sullivan, Jens Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[10] Yung-Lyul Lee, Ki-Hun Han, and Gary J Sullivan, "Improved lossless intra coding for H.264/MPEG-4 AVC," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2610–2615, 2006.

[11] Yih Han Tan, Chuohao Yeo, and Zhengguo Li, "Residual DPCM for lossless coding in HEVC," .

[12] Jun-Ren Ding, Jiun-Yu Chen, Fu-Chun Yang, and Jar-Ferr Yang, "Two-layer and adaptive entropy coding algorithms for H.264-based lossless image coding," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1369–1372, 2008.

[13] Jae S Lim, "Two-dimensional signal and image processing," *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.*, vol. 1, 1990.