Journal of Visual Communication and Image Representation 50 (2018) 280–289

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

A perceptually temporal adaptive quantization algorithm for HEVC☆

Guoqing Xiang[a], Huizhu Jia[a,e,*], Mingyuan Yang[b], Xinfeng Zhang[c], Xiaofeng Huang[d], Jie Liu[a], Xiaodong Xie[a]

[a] National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, China
[b] Beijing BOYA-HUALU Technology Inc., Beijing 100080, China
[c] Ming-Hsieh Department of Electrical Engineering University of Southern California, Los Angeles, CA 90089-2564, USA
[d] Nvidia Company, Pudong New Area, Shanghai, China
[e] Cooperative Medianet Innovation Center and Beida (Binhai) Information Research, Tianjin, China

ARTICLE INFO

Keywords:
Inter-frame dependency
AQ
QP offset
JND
SSIM
Subjective quality

ABSTRACT

Adaptive quantization (AQ) proves to be an effective coding tool to improve the performance of video coding. This paper presents a perceptually temporal AQ method to improve the subjective coding performance for High Efficiency Video Coding (HEVC). We first put forward a perceptual quality oriented motion estimation algorithm, which is conducted with a spatial-temporal just noticeable distortion (JND) model. Then one perceptual feature in temporal domain is proposed to develop our AQ method, which can generate different quantization parameter (QP) offsets for each coding unit (CU). The proposed method fully utilizes the temporal and perceptual characteristics of each CU, which can produce more visual-friendly QP offsets distribution. Experiments are conducted on HM16.0 (HEVC reference software), and with SSIM (Structure Similarity Index Metric) as the distortion metric, more than 8.08% and 7.95% rate savings can be obtained for Low-Delay-P (LDP) and Low-Delay-B (LDB) configurations on average, respectively. The subjective quality evaluation demonstrates that the proposed AQ method can achieve comparable visual quality as the HM16.0 while the proposed method can yield remarkable bitrate reductions.

1. Introduction

Adaptive quantization (AQ) is an important tool in video coding for the performance improvement. It adjusts quantization parameter (QP) for distinctive picture areas according to different spatial, temporal or perceptual characteristics. Typically, there are two kinds of AQ methods, one is the constructive part of the rate control, for which the QP of each coding unit (CU) is decided by rate control algorithms. The other one is achieved by preprocessing the video to get better subjective quality. In this study, we focus on the latter topic which is specified as the preprocessing based AQ method.

There are several types of AQ schemes based on preprocessing. One scheme is designed with spatial or temporal features empirically. For example., the x265 [1], which is an optimized implementation of HEVC standard, has integrated two spatial AQ methods (denoted as AQ-1 Method and AQ-2 Method) and one temporal method (CUTree Method) to improve the visual quality of video coding. However, the spatial methods only considering the spatial effect and the CUTree Method cannot work well when the reference frames are obviously different from current coding frames. In the HM16.0, another spatial AQ method is adopted [6], which is originated from Test Model 5 (TM5 Model) of MPEG-2 [7]. It scales quantization step according to the spatial activity of one CU relative to the frame-level average of the spatial activity. It should be noted that this AQ method utilizes the minimum variance of four sub-units in each CU to represent the whole CU's spatial complexity, which results in underestimated variance for large size CU and limits the performance improvement [8,19].

Besides utilizing the spatial and temporal pixel characteristics, a block level AQ (BLAQ) scheme has been incorporated in HM16.0 [9], which is based on the property that each block has its own quantization parameter. However, the BLAQ method finds a proper QP for each block by enabling the Rate Distortion Optimization (RDO) procedure for each candidate QP, which increases the computational complexity significantly. In [10] Lee et al. utilized the importance of transformed coefficients in terms of signal reconstruction and proposed a transform coefficient-level quantization technique based on a soft threshold approach. The coefficient-level AQ scheme required additional inverse QP adjustment operation in decoder. What's more, the soft threshold only

☆ This paper has been recommended for acceptance by Zicheng Liu.
* Corresponding author at: National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, China.
E-mail addresses: gqxiang@jdl.ac.cn (G. Xiang), hzjia@jdl.ac.cn (H. Jia), mingyuan.yang@boyahualu.com (M. Yang), xinfengz@usc.edu (X. Zhang), hehuang@nvidia.com (X. Huang), liuzimin@jdl.ac.cn (J. Liu), xdxie@jdl.ac.cn (X. Xie).

https://doi.org/10.1016/j.jvcir.2017.11.011
Received 12 February 2017; Received in revised form 20 September 2017; Accepted 16 November 2017
Available online 20 November 2017

relates to the distance between the positions of the current coefficient and DC coefficient in the transform block, and different quantization on the coefficient level within one CU may bring visible distortion fluctuation.

Another main QP offset determination scheme is based on the Lagrange multiplier since there is a direct relationship between QP and Lagrange multiplier. The relationship can be formulated as Eq. (1) in HM16.0 [11], where *c* is a constant parameter related to configurations. One can first propose an adaptive Lagrange multiplier model and then the updated QP can be calculated.

$$\lambda = c * 2^{(QP-12)/3}. \tag{1}$$

In [8], Yeo et al. presented a method utilizing pixel variance to scale the Lagrange multiplier in RDO for each CU in HEVC. With the updated Lagrange multiplier, they can get QP offset for each CU according to Eq. (1). The QP modification granularity can be signaled by the encoder. However, if it is used for all depths of CUs, the method requires transmitting a lot of bits for all the QP offsets, which limits the coding performance improvement [8]. Furthermore, this method only utilized the spatial variance to design the SSIM-based RDO AQ method, which is limited for sequences occupied with a lot of homogeneous areas [19].

Temporal inter-dependency is a significant characteristic for videos, and can be utilized for optimizing the Lagrange multiplier in RDO algorithm too. In [12], Li et al. investigated the RDO with inter-frame dependency and proposed a RDO scheme by adapting λ with *SSD* (Sum of Square Difference). However, it is designed for improving PSNR (Peak Signal to Noise Ratio) and only the λ is updated with QP being invariable for CU.

As a conclusion, on one hand, although there are several AQ methods for HEVC, most of the existing methods are spatial ones, the temporal inter frame dependency has not be exploited well. On the other hand, the existing few temporal AQ methods for HEVC, such as CUTree Method in x265, only takes the CU-level reference importance to decide the QP offset. However, considering the distortion and bitrate of each CU in one coding frame should be balanced with the overall bitrate by the rate distortion constraint within one frame [1], only taking the CUs' influence in the temporal chain will not fulfill the requirement well. What's more, most of these existing AQ methods have not taken the human visual characteristics into consideration, which cannot improve the perceptual video compression performance sufficiently.

In this paper, we propose a perceptual AQ method aimed to improve the subjective coding performance. Firstly, a perceptual motion estimation method is presented with an improved just noticeable distortion (JND) model. The JND model is improved with the latest spatial one in [14] by taking temporal masking effect into consideration. Secondly, a perceptually temporal AQ method is proposed, which utilizes a perceptually temporal feature to produce more visual-friendly QP offsets distribution for CUs. Finally, evaluation results of the proposed AQ method are shown based on SSIM metric and subjective quality test.

The rest of this paper is organized as follows: The related works are reviewed in Section 2. In Section 3, a perceptual motion estimation method is presented and the proposed AQ method is introduced in detail. Extensive experiments are shown in Section 4. Section 5 presents our conclusion.

2. Related works

In this section, the background on AQ for HEVC are introduced at first and then in order to comprehend the nature of AQ methods, related typical AQ methods are presented including AQ methods in x265, TM5 Method and the SSIM-based RDO Method.

2.1. Quantization group in HEVC

For the HEVC, QP modification is allowed within each QG (Quantization Group) as shown in Fig. 1. A QG is a square area within a CTU (Coding Tree Unit), and the minimum QG size is signaled in a high-level picture parameter set. The QP difference between the desired QP and the predicted QP is signaled. If each depth CU has its own QP offset, more bits will be required to transmit in final bitstream. In Fig. 1, the maximum QG size is 32 × 32 and 16 × 16 is the minimum QG size as shown with bold lines, which means the QPs for the 3 CUs with size 32 × 32 or the 3 CUs with size 16 × 16 can be different from each other, while the 4 CUs with size 8 × 8 with dashed lines can only have the same QP.

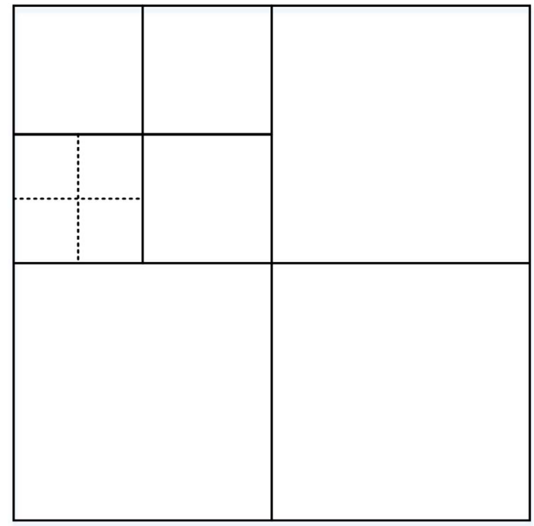

Fig. 1. The QGs in HEVC. The maximum QG size is 32 × 32 and 16 × 16 is the minimum QG size as shown with bold lines.

2.2. AQ methods in x265

In x265 [1], two spatial AQ methods and one temporal method have been integrated to improve the visual quality of compressed videos. For AQ-1 Method, only the pixel variance of each CU is calculated to get its QP offset directly by Eq. (2),

$$\Delta Qp_{d,k} = s_1 * \log_2(\sigma_{d,k}^2), \tag{2}$$

where $\sigma_{d,k}^2$ is the variance of the kth CU in depth d, s_1 is a control parameter.

Different from AQ-1 Method, AQ-2 Method utilizes one power function to get one temporary CU-level QP offset $\Delta QP_{d,k}^{\star}$ at first as Eq. (3)

$$\Delta QP_{d,k}^{\star} = \sqrt[8]{\sigma_{d,k}^2 + 1}. \tag{3}$$

Then the average factor $\overline{S}_d$ and its square value $\overline{S^2}_d$ for the frame are calculated by averaging all the CU temporary QP offsets $\Delta QP_{d,k}^{\star}$ and the square of them in depth d. The frame-level average QP offset $\Delta Qp_{d,f}$ is determined as

$$\Delta QP_{d,f} = \overline{S}_d - \frac{1}{2} * (\overline{S^2}_d - 14)/(\overline{S}_d), \tag{4}$$

Finally, the QP offset of each CU is decided by the difference between the CU-level and frame-level QP offset as Eq. (5).

$$\Delta Qp_{d,k} = s_2 * (\Delta Qp_{d,k}^{\star} - \Delta Qp_{d,f}), \tag{5}$$

where s_2 is a control parameter.

Besides the two spatial methods, for the temporal AQ method in x265, namely the CUTree Method, it decides the QP offset adaptively with the amount of propagation distortion using a relative propagation cost model as Eq. (6),