SSIM-BASED ADAPTIVE QUANTIZATION IN HEVC

Chuohao Yeo, Hui Li Tan, Yih Han Tan

Signal Processing Department
Institute for Infocomm Research
1 Fusionopolis Way, Singapore 138632

ABSTRACT

HEVC is an emerging video coding standard that can achieve significant compression gains compared to H.264/AVC due to the inclusion of numerous new coding tools. In particular, it allows for a flexible quadtree based block partitioning of each coding tree unit (CTU) and an ability to switch quantization parameters (QP) on a sub-CTU level. In this paper, we present an approach for selecting quantization parameters for each block of pixels on the basis of optimizing the SSIM of the entire picture. Our simulation results show that when SSIM is the quality metric, the proposed approach is able to give average BD-Rate gains of 5.5% to 7.4% compared to using a constant QP per picture while having a negligible increase in encoding runtime. In addition, our proposed method also significantly outperforms the MPEG-2 TM5 adaptive quantization algorithm implemented in the HEVC reference software.

***Index Terms*—** HEVC, SSIM, Rate distortion optimization, Adaptive quantization

1. INTRODUCTION

Since April 2010, ISO/IEC MPEG and ITU-T VCEG, under the Joint Collaborative Team on Video Coding (JCT-VC), have been working on the development of the High Efficiency Video Coding (HEVC) standard; the first version of this standard is expected to be finalized in January 2013 [1]. Extensive evaluations have already shown that HEVC demonstrates significant compression gains compared to H.264/AVC as well as previous coding standards [2]. HEVC is able to achieve these gains by including numerous new coding tools, such as a coding tree unit (CTU) structure which can be recursively divided into coding units (CU) in a quadtree fashion, variable-sized prediction units (PU), recursive quadtree transforms, motion signaling using multiple candidate motion vector prediction and motion merging, higher precision and longer tap interpolation filters for motion compensation, increased number of intra prediction modes and sample adaptive offset [1].

As in H.264/AVC, HEVC allows for the quantization parameter (QP), which controls how the decoder dequantizes the received coefficient levels, to change within a slice; this is a useful feature for rate control or perceptually motivated adaptive quantization. Unlike H.264/AVC, which only allows for QP to change at most once per macroblock (MB), HEVC allows for QP to change at a sub-CTU, or quantization group (QG), level [3]. Furthermore, the granularity at which QP modification occur can be signaled by the encoder. In the current HEVC reference software[1], HM8.0, a rate control algorithm and an adaptive quantization algorithm based on MPEG-2 TM5 [4] are implemented in the encoder which demonstrate how this QP modification can be effected.

In this work, we take a further look into how perceptually motivated adaptive quantization could be done in the emerging HEVC standard. Based on our previous work on structural similarity index (SSIM) based rate distortion optimization (RDO) [5], we view the adaptive quantization problem as one of optimizing the SSIM of the reconstructed picture. By doing so, we can derive the necessary updates of the QP and Lagrange multiplier that is used for RDO. Furthermore, we have to adapt the method to HEVC as it has a quadtree based block structure and allows for QP to change at a sub-CTU level. Our simulation results show that when SSIM is the quality metric, our proposed approach is able to give average BD-Rate gains of 5.5% to 7.4% compared to using a constant QP per picture at no significant increase in encoding runtime. In addition, our proposed method also significantly outperforms the MPEG-2 TM5 adaptive quantization algorithm as implemented in the HEVC reference software.

2. BACKGROUND AND RELATED WORK

2.1. HEVC

As excellent overview articles of HEVC are available, e.g., [1], we will only briefly describe the features of HEVC that are relevant to this work.

In HEVC, each picture is partitioned into non-overlapping CTUs that can range in size from 16x16 pixels to 64x64 pixels in the Main Profile. Each CTU can be recursively subdivided in a quadtree fashion into CUs, down to a minimum size of 8x8 pixels in the Main Profile; the actual minimum CU size can be signaled in a higher level sequence parameter set. Each leaf CU, which is a CU that is not subdivided,

[1]Available at `https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/`

can be further partitioned into PUs; each PU carries information about how that block of pixels is to be predicted. At the same time, the residual samples after prediction of each leaf CU can be recursively subdivided into square transform units (TU), and each TU undergoes a separate dequantization and inverse transform during the decoding process.

QP modification is allowed once per QG. Fig. 1 shows an example of the interaction between the QGs, CTU and CUs. A QG is a square area within a CTU, and the minimum QG size is signaled in a higher level picture parameter set. The QG of a leaf CU that is larger than or equal to the minimum QG size is the CU itself; otherwise, the QG of that CU is the smallest QG that the CU is located within. The QP difference between the desired QP and the predicted QP is signaled only for the first TU with non-zero residual coefficient levels within each QG. The predicted QP is computed as $(\text{QP}_A + \text{QP}_B + 1) >> 1$, with QP_A and QP_B being computed as follows [3]. Let X denote the TU at the top-left of the current QG, and let A and B denote the TUs to the left and above of X respectively. QP_A is set to be the QP used in A if X is not at the left boundary of the current CTU, otherwise it is set to be the previous coded QP. Similarly, QP_B is set to be the QP used in B if X is not at the top boundary of the current CTU, otherwise it is set to be the previous coded QP.

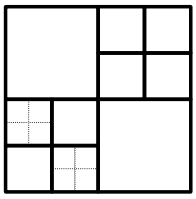

Fig. 1. Illustration of QGs in HEVC. A 64x64 CTU is shown partitioned into CUs ranging in size from 32x32 to 8x8, while QGs corresponding to a minimum QG size of 16x16 are shown with bold lines.

2.2. Adaptive quantization in MPEG-2 TM5

In MPEG-2 Test Model 5 (TM5), an adaptive quantization method describes how to scale the quantization step size according to the spatial activity in the MB relative to its average over the previous coded frame [4]. As implemented in HM8.0, this adaptive quantization is adapted as follows. For the k-th QG of size $2N \times 2N$ within the picture, the variances, $\{\sigma^2_{k,i}\}^3_{i=0}$, of each of the 4 $N \times N$ sub-block within it is computed. The spatial activity of the k-th QG is then computed as $A_k = 1 + \min(\sigma^2_{k,0}, \sigma^2_{k,1}, \sigma^2_{k,2}, \sigma^2_{k,3})$. The average spatial activity of the entire picture is computed as $\bar{A} = \frac{1}{K}\sum_{j=0}^{K} A_j$, where K is the total number of QGs in the picture. The computed QP offset for the k-th QG is then given as:

$$\Delta\text{QP}_k = 6\log_2\left(\frac{SA_k + \bar{A}}{A_k + S\bar{A}}\right)$$

where $S = 2^{\frac{\Delta\text{QP}_{max}}{6}}$ and ΔQP_{max} is the maximum allowable absolute difference from the slice QP.

2.3. SSIM in rate distortion optimization

The structural similarity index (SSIM) has been proposed as a image quality metric that is more correlated with human perceptual quality than mean square error (MSE) [6]. The SSIM between two image regions x and y is defined as [6]:

$$\text{SSIM} = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right)\left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right),$$

where μ_x and μ_y are the means of x and y respectively, σ_x^2 and σ_y^2 are the variances of x and y respectively, σ_{xy} is the cross-covariance between x and y, and c_1 and c_2 are two constants used for numerical stability.

There has been a number of works that use a derivative of SSIM as the distortion metric when performing RDO in the context of H.264/AVC. Some works replace SSE with (1-SSIM) in RDO, with various methods of computing an appropriate Lagrange multiplier, e.g., [7, 8]. In our previous work [5], we replace SSE by (1/SSIM), and showed that it is approximately equivalent to using a spatially adaptive Lagrange multiplier with the existing SSE-based RDO, where the scaling can be computed analytically. Recently, there has also been a reported work on performing SSIM-inspired RDO in HEVC based on divisive normalization [9].

3. PROPOSED APPROACH

We first review key aspects of our previous work [5], and then describe how it can be applied to HEVC.

3.1. SSIM-based RDO

We denote the original image region by x and the reconstructed image region by y. By using an additive distortion model for y, i.e., $y = x + e$, and assuming the additive noise term, e, has zero mean and is uncorrelated with x, we can compute the following approximation of SSIM:

$$\text{SSIM} \approx \frac{2\sigma_x^2 + c_2}{2\sigma_x^2 + \text{MSE} + c_2}$$

We then define a new SSIM-based distortion metric as:

$$\text{dSSIM} = \frac{1}{\text{SSIM}} - 1 \approx \frac{\text{MSE}}{2\sigma_x^2 + c_2}$$

We use this as the distortion metric and optimize the following Lagrangian cost during RDO:

$$\begin{aligned} J &= N \cdot \text{dSSIM} + \lambda R \\ &\approx N\left(\frac{\text{MSE}}{2\sigma_x^2 + c_2}\right) + \lambda R \\ &= \frac{1}{2\sigma_x^2 + c_2}\left(\text{SSE} + \left(2\sigma_x^2 + c_2\right)\lambda R\right). \end{aligned}$$

where N is the number of pixels in the MB. We can also equivalently optimize $J = \text{SSE} + \left(2\sigma_x^2 + c_2\right)\lambda R$ with an appropriately computed λ.

By applying the condition that the overall rate of coding the frame is kept the same after using dSSIM as the distortion metric in RDO, we find that the following should be used as the Lagrange multiplier for the i-th MB [5]:

$$\lambda_i = \frac{2\sigma_{x_i}^2 + c_2}{\exp\left(\frac{1}{M}\sum_{j=1}^{M}\log\left(2\sigma_{x_j}^2 + c_2\right)\right)}\lambda_{\text{SSE}}. \qquad (1)$$

where $\sigma_{x_i}^2$ is the local source variance for the i-th MB, and λ_{SSE} is the Lagrange multiplier that is used in JM for SSE based RDO. We also found that this choice of Lagrange multiplier could also be used to determine the implicit QP offset by making use of the relation $\lambda_{\text{SSE}} = \beta \cdot 2^{(\text{QP}-12)/3}$ [10], i.e.,

$$\Delta\text{QP}_i = 3\left(s_i - \frac{1}{M}\sum_{j-1}^{M} s_j\right), \qquad (2)$$

where $s_i = \log_2\left(2\sigma_{x_i}^2 + c_2\right)$.

3.2. SSIM-based adaptive quantization in HEVC

Since RDO is also used in HM, we can use (1) and (2) to compute the Lagrange multiplier scaling and QP offset to use for each QG so as to achieve the optimal SSIM. However, there are some issues to be resolved. As noted earlier, in HEVC, unlike H.264/AVC, QP can change multiple times in each CTU (up to once every QG). Also, mode decision is done in a recursive quadtree manner for each CTU down to the leaf CU. This means that the QP decision is now intertwined with the CTU mode decision process, whereas in H.264/AVC, it was possible to test combinations of QP and MB mode decisions in a relatively straightforward manner. Furthermore, in H.264/AVC, we could use a single Lagrange multiplier for the entire MB in its mode decision process, while in HEVC, there could potentially be a variety of Lagrange multipliers used within each CTU in its mode decision process.

To illustrate this, we consider the example in Fig. 2. Suppose that X is a $2N \times 2N$ CU, and it can either be coded as a $2N \times 2N$ CU, X_0, or be coded as 4 $N \times N$ CUs, $\{X_{0,i}\}_{i=0}^{3}$. Also the QG can have a minimum size of $N \times N$; i.e., a different QP can be used for each $N \times N$ CU. Due to the recursive quadtree nature of CUs in HEVC, mode decision for each CU is also performed in a recursive quadtree fashion in HM8.0; mode decision is first carried out for X_0 and separately for each $X_{0,i}$, before comparing the RD costs of X coded as X_0 and X coded as $\{X_{0,i}\}_{i=0}^{3}$ to make the final decision for X. The complication is that if the Lagrange multiplier scaling in (1) is used, the Lagrange multiplier used for X_0's mode decision and each of $X_{0,i}$'s mode decision could be different.

We propose to resolve this as follows. First, the Lagrange multiplier and QP to be used in each CU's mode decision is computed as in (1) and (2) respectively, where the variance used is that of the QG in which the CU resides. This choice is made since the QP is a free parameter within each QG, so this would allow each QG to pick the QP that is perceptually optimal. Second, and consistent with the above choice, when deciding if a CU should be split into sub-CUs or not, the Lagrange multiplier used is that which is computed for the CU.

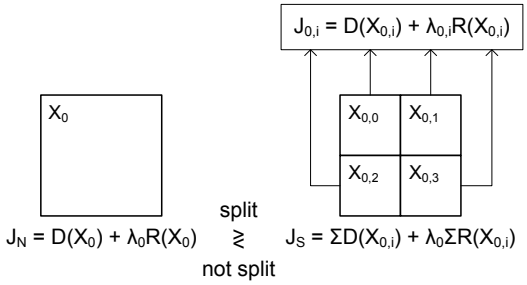

Fig. 2. Illustration of proposed method.

Going back to our above example in Fig. 2, we would compute λ_0 and QP_0 for X_0 with a QG size of $2N \times 2N$ and use it for X_0's mode decision. Similarly, we would compute $\lambda_{0,i}$ and $\text{QP}_{0,i}$ for each $X_{0,i}$ with a QG size of $N \times N$ and use it for $X_{0,i}$'s mode decision. Finally, we would use λ_0 in computing the RD cost of X coded as X_0 and X coded as $\{X_{0,i}\}_{i=0}^{3}$ when making the final decision for X.

4. EXPERIMENTS AND RESULTS

In our experiments, we encoded the HEVC test sequences using the HEVC Main Profile random-access configuration described in the HM8.0 common test conditions [11]. This uses a hierarchical B picture structure with a GOP size of 8 frames, and an Intra frame inserted approximately every second. The anchor used is HM8.0 without any modifications. In addition to testing our proposed approach implemented in HM8.0 as described in Section 3, we also ran experiments using the MPEG-2 TM5 adaptive quantization algorithm that is implemented in HM8.0 (option "-aq 1").

For both tested methods, we considered different minimum QG sizes, ranging from 64x64 (option "-dqd 0"), which is the CTU size in common test conditions, down to 8x8 (option "-dqd 3"), which is the minimum CU size in common test conditions. We also limited the maximum QP offset to 3 (option "-aqr 3"). Four QP points, $\{22, 27, 32, 37\}$, were used to encode each sequence.

Fig. 3 illustrates the various SSIM-Rate performance of the proposed method for different QG size granularity for the sequence BQSquare, while Table 1 shows the BD-Rate numbers [12], which uses SSIM as the quality metric, of the proposed approach and the MPEG-2 TM5 adaptive quantization algorithm in HM8.0 compared against the HM8.0 anchor. Note that a negative BD-Rate number means that the method being compared uses less rate than the HM8.0 anchor for the same SSIM score. The results suggest that our proposed approach is able to achieve an average compression improve-

Table 1. BD-Rate (%) using SSIM as quality metric vs HM8.0 anchor

Sequence	Proposed Approach				MPEG-2 TM5 Step 3			
	BD-Rate (-dqd 0)	BD-Rate (-dqd 1)	BD-Rate (-dqd 2)	BD-Rate (-dqd 3)	BD-Rate (-dqd 0)	BD-Rate (-dqd 1)	BD-Rate (-dqd 2)	BD-Rate (-dqd 3)
Traffic	-9.6%	-11.0%	-10.6%	-9.6%	-0.7%	0.1%	1.2%	2.3%
PropleOnStreet	-5.8%	-7.5%	-7.9%	-7.0%	2.0%	3.2%	4.7%	5.9%
Nebuta	-2.0%	-0.0%	0.1%	0.2%	3.3%	3.7%	3.9%	3.9%
SteamLocomotive	-0.5%	-0.0%	0.7%	0.7%	-0.2%	0.2%	0.4%	0.7%
Kimono	0.8%	1.5%	2.0%	2.3%	0.5%	1.2%	1.7%	1.8%
ParkScene	-5.0%	-6.5%	-6.2%	-5.4%	1.8%	2.1%	2.8%	3.5%
Cactus	-0.6%	-0.9%	-0.5%	0.0%	0.4%	1.0%	1.9%	2.7%
BasketballDrive	-9.7%	-9.8%	-9.0%	-8.3%	2.4%	3.3%	4.4%	5.3%
BQTerrace	-3.5%	-6.2%	-7.3%	-7.2%	0.4%	1.1%	2.0%	2.7%
BasketballDrill	-20.0%	-19.9%	-19.1%	-18.1%	1.0%	1.9%	2.7%	3.7%
BQMall	-2.9%	-5.0%	-4.9%	-4.1%	0.6%	1.2%	2.4%	3.6%
PartyScene	-2.3%	-5.5%	-6.8%	-5.3%	0.9%	1.2%	2.0%	3.1%
RaceHorsesC	-2.0%	-4.0%	-3.1%	-2.2%	1.6%	2.2%	2.8%	3.8%
BasketballPass	-11.7%	-15.2%	-15.8%	-14.7%	1.9%	3.1%	4.8%	5.9%
BQSquare	-10.7%	-17.9%	-21.7%	-21.9%	-0.1%	0.7%	1.7%	3.0%
BlowingBubbles	0.2%	-2.8%	-5.7%	-4.0%	-0.3%	0.0%	0.9%	2.0%
RaceHorsesD	-6.7%	-9.1%	-9.8%	-8.4%	1.6%	2.4%	3.4%	4.5%
BasketballDrillText	-17.0%	-18.4%	-17.2%	-16.5%	1.3%	1.9%	3.3%	4.4%
ChinaSpeed	-1.4%	-2.4%	-3.1%	-3.2%	2.9%	3.5%	5.1%	6.0%
SlideEditing	1.0%	-2.4%	-0.9%	-1.8%	0.3%	0.4%	2.1%	3.2%
SlideShow	-5.5%	-7.5%	-7.9%	-7.8%	-0.7%	-0.5%	1.0%	2.1%
Average	-5.5%	-7.2%	-7.4%	-6.8%	1.0%	1.6%	2.6%	3.5%

ments of 5.5% to 7.4%, depending on the minimum QG size that is used. We find that that across each sequence and also in the average performance, there is diminishing returns associated with increasing QG size granularity, and could even lead to a decrease in compression gain. This is not surprising; while a finer QG size granularity enables finer control of QP, it also leads to higher QP signaling overhead. We also observe that the MPEG-2 TM5 adaptive quantization algorithm implemented in HM8.0 does not lead to any SSIM improvements (as shown here) or PSNR improvements (PSNR results not shown due to space constraints).

We also measured the encoding runtime of the proposed method and that of the anchor for each of the test sequences and QP points, and the geometric mean of the runtimes was computed. We find that the proposed method introduces virtually no encoding complexity overhead, as it uses 98% to 100% of the anchor encoding runtime on average, depending on the minimum QG size used. In certain cases, the encoding runtime is even smaller than that of the anchor, as the proposed approach leads to a video being coded at a lower rate.

5. CONCLUSIONS

We have proposed an SSIM-based adaptive quantization algorithm for HEVC that is based on optimizing a SSIM-derived distortion metric for each block of pixels. The proposed approach involves computing a local scaling of the Lagrange multiplier to be used in RDO for each block and a QP offset based on that computed scaling. No training or multi-pass encoding is required. Experimental results of the proposed method implemented on HM8.0 demonstrates that it can achieve average compression improvements of 5.5% to 7.4% across the HEVC test sequences while maintaining the same SSIM score. One promising direction of future work is to investigate how best to choose an appropriate minimum QG size within this adaptive quantization framework.

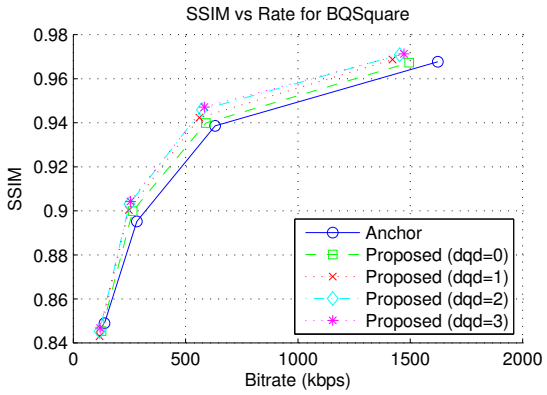

Fig. 3. SSIM vs Rate plot of "BQSquare" for anchor and proposed method.

6. REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*.

[2] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*.

[3] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "WD9: Working Draft 9 of High-Efficiency Video Coding," in *JCTVC-K1003*, Shanghai, China, Oct 2012.

[4] ISO/IEC/JTC1/SC29/WG11, *MPEG-2 Test Model 5*, chapter 10. "Rate control and quantization control", Mar. 1993.

[5] C. Yeo, H. L. Tan, and Y. H. Tan, "On rate distortion optimization using SSIM," *IEEE Transactions on Circuits and Systems for Video Technology*.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[7] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[8] S. Wang, A. Rehman, W. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*.

[9] A. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," in *IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2012, pp. 497 –502.

[10] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[11] F. Bossen, "Common test conditions and software reference configurations," in *JCTVC-J1100*, Stockholm, Sweden, Jul 2012.

[12] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," in *ITU-T SC16/Q6, VCEG-M33*, Austin, USA, Apr. 2001.