ADAPTIVE QUANTIZATION BY SOFT THRESHOLDING IN HEVC

Lee Prangnell [1], Victor Sanchez [1], and Rahul Vanam [2]
[1] Department of Computer Science, University of Warwick, Coventry, England, UK
[2] InterDigital Communications, Inc., San Diego, California, USA

Introduction

Main Objective

- Quantize transform coefficients individually to improve signal reconstruction quality.

Approach

- *Soft Thresholding*: our technique employs a continuous monotonically decreasing function to define the importance of a transform coefficient in terms of signal reconstruction.

Contributions

- Suitable for a variety of video applications that support the HEVC codec, including consumer electronics technology.
- Improved signal reconstruction quality resulting in improved coding efficiency performance.
- Reduction of non-zero quantized transform coefficients resulting in faster entropy coding and decoding times.
- An emphasis on applying lower quantization for low frequency transform coefficients and coarser quantization for high frequency transform coefficients, resulting in improved visual quality.

URQ

URQ, which is typically used in combination with Rate Distortion Optimized Quantization (RDOQ), is a block level quantization approach that equally quantizes all transform coefficients in a Transform Block (TB) according to a QP value [1]. URQ does not take into account the importance of individual transform coefficients during the quantization process, which represents an important shortcoming of the technique.

Following intra or inter prediction, for each TB ranging from 4 × 4 to 32 × 32 samples, a finite precision approximation of the DCT is applied to the residual signal to compute the transform coefficients. After linear transformation by DCT, a TB comprises low frequency components consisting of a DC coefficient and the AC coefficients close in proximity in addition to medium and high frequency AC coefficients.

Encoder Side

A transform coefficient $C(x,y)$, located at coordinates (x,y) within an $N \times N$ TB, is quantized to a transform coefficient level value l, as given by (1):

$$l(x,y) = \frac{C(x,y) \cdot Q + offset}{2^{21 + \frac{QP}{6} - \log_2 N}} \quad (1)$$

where Q is the multiplication factor associated with the QP value and *offset* is a constant value that specifies the error caused by rounding and the level of deadzone.

Decoder Side

At the decoder side, a transform coefficient is recovered by inverse quantization, as given by (2):

$$C'(x,y) = \frac{l(x,y) \cdot IQ \cdot 2^{\frac{QP}{6}}}{2^{\log_2 N - 1}} \quad (2)$$

where $C'(x,y)$ is the recovered coefficient located at coordinates (x,y) within an $N \times N$ TB and IQ is the scaling factor used for inverse quantization. Table 1 tabulates Q and IQ values for six QP values in URQ.

Table 1. Q and IQ values for six QP values in URQ.

QP	0	1	2	3	4	5
Q	26214	23302	20560	18396	16384	14564
IQ	40	45	51	57	64	72

Related Work

Alternative HEVC adaptive quantization methods have been previously proposed to improve upon URQ. These methods are as follows:

- Structural Similarity Index Metric (SSIM).
- Intensity Dependent Spatial Quantization (IDSQ).
- Adaptive Quantization for Screen Content Videos (AQSCV).
- N-Level Quantizer (NLQ).

In contrast to URQ, SSIM [2], IDSQ [3], and AQSCV [4], our technique modifies quantization parameter (QP) values at the transform coefficient level instead of at the block level. NLQ is a hard thresholding technique, which also works at the coefficient level. However, when quantization is applied to a block using NLQ, there exists only a restricted number of different quantization levels that can be obtained [5], which represents a shortcoming of the NLQ method.

Proposed Technique

- Based on modifying Q values by using a weight, w, and quantizing each coefficient using a specific Q value; this is quantified in (3).

$$Q' = Q \times w \quad (3)$$

- Weight w is quantified by an exponential function, which includes a distance parameter d and an energy parameter c; w is computed in (4).

$$w = e^{-\left(\frac{d}{c}\right)^2} \in [0,1] \quad (4)$$

- Weight w is then defined as a continuous, monotonically decreasing function that determines the importance of a transform coefficient in terms of reconstructing the signal (see Fig. 1).

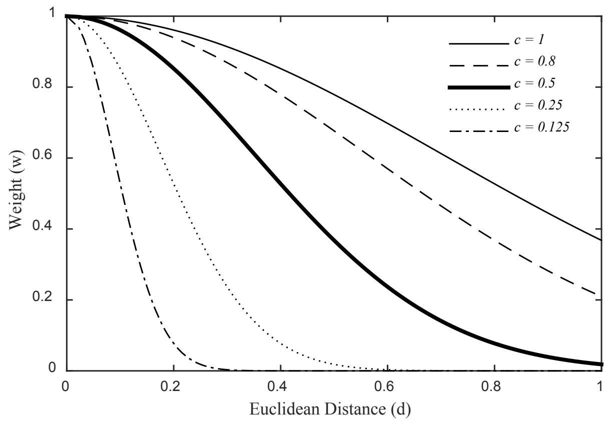

Fig. 1. Weight w for various values of the energy parameter c and Euclidean distance d.

- The inverse quantization process is also modified accordingly, which equates to the following: $IQ' \times Q' = 2^{20}$ for both luma and chroma components.

Distance Parameter

- The distance parameter d is quantified in the spatial domain using the normalized Euclidean distance formula (see Fig. 2), which is quantified in (5).

$$d = \sqrt{\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{(x_1 - x_{\max})^2 + (y_1 - y_{\max})^2}} \in [0,1] \quad (5)$$

DC *0* **0**	**AC** *1.0000* **0.2357**	**AC** *2.0000* **0.4714**	**AC** *3.0000* **0.7071**
AC *1.0000* **0.2357**	**AC** *1.4142* **0.3333**	**AC** *2.2361* **0.5271**	**AC** *3.1623* **0.7454**
AC *2.0000* **0.4714**	**AC** *2.2361* **0.5271**	**AC** *2.8284* **0.6667**	**AC** *3.6056* **0.8499**
AC *3.0000* **0.7071**	**AC** *3.1623* **0.7454**	**AC** *3.6056* **0.8499**	**AC** *4.2426* **1**

Fig. 2. Transform coefficients in a 4 × 4 TB. The positions of the low frequency components are displayed in darker shades. Numerical values represent the Euclidean distance of each coefficient from the DC coefficient before (italics) and after (bold) normalization.

Energy Parameter

- The energy parameter c in (3) controls the decay of the exponential function for the purpose of gradually decreasing Q values. This parameter is computed in (6).

$$c = \frac{E}{E_{\max}} \in [0,1] \quad (6)$$

- E is an estimate of the total energy of an $N \times N$ TB. Let us denote the n^{th} recovered transform coefficient within an $N \times N$ TB as C'_n, with $n = 1$ and $n = N \times N$ denoting the DC coefficient and the coefficient located at coordinates ($x=N$, $y=N$), respectively, following a zig-zag order.
- The value of E for the n^{th} coefficient in an $N \times N$ TB is calculated in (7).

$$E_n = \sum_{m=1}^{n-1} (C'_m)^2 \quad (7)$$

- $E_{\max}$ is estimated in the frequency domain as the energy of an $N \times N$ TB in which all residual values are $r = 2^b - 1$, where b represents the bit depth of the data.

Research Methodology

Common Test Conditions recommended by JCT-VC. The official test sequences used have resolutions of 2560 × 1600, 1920 × 1080, 832 × 480, 416 × 240 and 1280 × 720, which represent classes A, B, C, D and E, respectively. Configurations: All Intra, Low Delay B, Low Delay P and Random Access using the Main Profile (MP), the High Efficiency (HE) profile and the QP values: 22, 27, 32 and 37.

Results & Discussion

Experimental Results (Encoding)

- In comparison with the HM 16 reference software, the most noteworthy luma and chroma BD-Rate gains attained by our technique are as follows: 5.5% (Y), 13.2% (Cb) and 11.7% (Cr) for HD sequences in Class E using the Low Delay B configuration and Main profile, which equates to greatly improved visual quality.
- In comparison with NLQ, the most significant average luma and chroma BD-Rate improvements achieved by our method are as follows: 2.7% (Y), 8.4% (Cb) and 8.4% (Cr) using the Low Delay P configuration and Main profile.
- Performs particularly well on the inter predicted residual transform coefficients in Class E sequences.
- Figure 3 shows an RD-Plot for the KristenAndSara sequence in Class E, which shows BD-Rate gains of 6.4% for Y-PSNR.

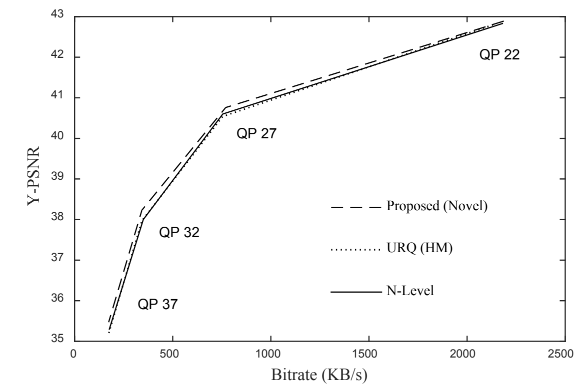

Fig. 3. RD-Plot showing the Y-PSNR improvements of our technique compared with URQ and NLQ for the sequence KristenAndSara in Class E using the Low Delay B Main configuration.

Experimental Results (Complexity)

- Our approach yields maximum speed improvements of 3.6% and 11.5% for encoding and decoding, respectively (see Fig. 4).

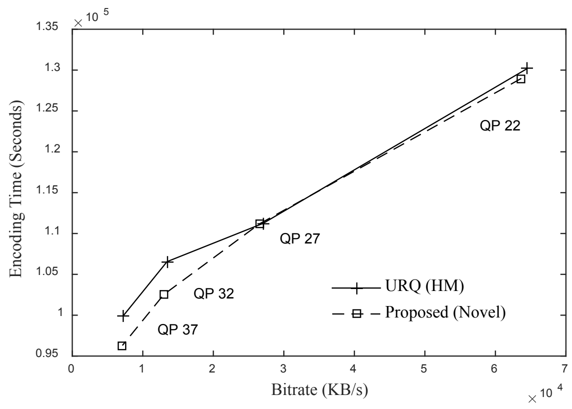

Fig. 4. Plot showing the improved encoding time performance of our technique compared with URQ.

Discussion

- Our technique yields improved coding efficiency performance because it adaptively utilizes high QP values for low energy transform coefficients in TBs.
- Our method reduces the number of non-zero quantized transform coefficients. This results in faster entropy coding and decoding times.

Conclusion

We have proposed an adaptive, transform coefficient level method of quantization for HEVC that takes into consideration the importance of transform coefficients in terms of reconstructing the signal. We compared our technique with URQ and NLQ. BD-Rate reductions of up to 5.5% (Y), 13.2% (Cb), and 11.7% (Cr), and decreases in encoding and decoding times, were attained.

[1] G. Sullivan, J-R. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, 2012.
[2] Y. Chuohao, T. Hui Li and T. Yih Han, "SSIM-Based Adaptive Quantization in HEVC," *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 1690-1694.
[3] M. Naccari and M. Mrak, "Intensity Dependent Spatial Quantization with Application in HEVC," *IEEE Int. Conf. Multimedia and Expo*, San Jose, CA, 2013, pp. 1-6.
[4] J. Nam, D. Sim and I.V. Bajic, "HEVC-based Adaptive Quantization for Screen Content Videos," *IEEE Int. Symp. Broadband Multimedia Systems*, Seoul, 2012, pp. 1-4.
[5] R. Gweon and Y. Lee, "N-Level Quantization in HEVC," *IEEE Int. Symp. Broadband Multimedia Systems*, Seoul, 2012, pp. 1-5.