THE UNIVERSITY OF WARWICK

ADAPTIVE QUANTIZATION MATRICES FOR HD AND UHD RESOLUTIONS IN SCALABLE HEVC

Lee Prangnell and Victor Sanchez

Department of Computer Science, University of Warwick, Coventry, England, UK

Department of COMPUTER SCIENCE

Introduction

- The proposed Adaptive Quantization Matrix (AQM) technique is a novel refinement to the default Human Visual System and Contrast Sensitivity Function (HVS-CSF) based Quantization Matrix (QM) technique in HEVC.
- The proposed AQM technique employs a display resolution parameter and a Euclidean distance parameter for the purpose of minimizing video compression artifacts in high resolution VDUs.
- These two parameters represent the resolution of the target VDU and the importance of transform coefficients within a Transform Block (TB) with respect to the resolution of the VDU.

HVS-CSF Default QM Technique in HEVC

- The HVS-CSF based 8×8 QM technique in [1] is employed as the default intra QM in HEVC [2].
- HEVC supports up to 32×32 TBs.
- Default 16×16 and 32×32 QMs, for the corresponding TB sizes, are obtained by upsampling and replicating the 8×8 default intra and inter QMs (see Fig. 1).

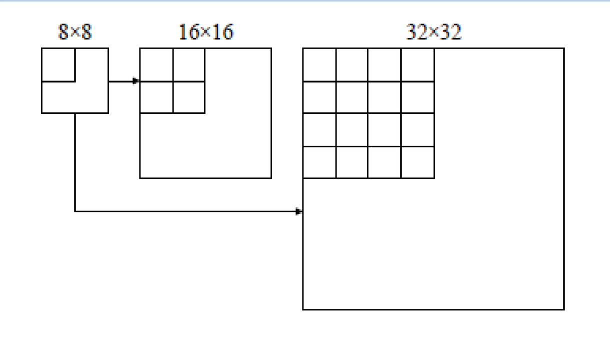

Fig. 1. The replication of 8×8 QMs to construct 16×16 and 32×32 QMs for corresponding TB sizes.

The HVS-CSF approach in [1], including the associated Modulation Transfer Function (MTF), is employed to produce a 2D Frequency Weighting Matrix (FWM), *H*(*f*), comprising floating point values, from which the default intra QM in HEVC is derived. *H*(*f*) is computed in (1):

$$H(f) = \begin{cases} a(b+cf) \times e^{-c(f)^{d}} & \text{if } f' > f_{\max}, \\ 1.0 & \text{otherwise,} \end{cases} \qquad (1)$$

where *f* is the radial frequency in cycles per degree of the visual angle subtended represented in two dimensions such that $f = f(u,v)$, where f_{max} denotes the frequency of 8 cycles per degree (i.e., the exponential peak). The MTF is computed with the constant values a=2.2, b=0.192, c=0.114 and d=1.1. The normalized radial spatial frequency, $f'(u,v)$, is derived using angular dependent function $S(\theta(u,v))$. Both $f'(u,v)$ and $S(\theta(u,v))$ are quantified in (2)-(5).

$$f'(u,v) = \frac{f(u,v)}{S(\theta(u,v))} \qquad (2)$$

$$f(u,v) = \frac{\pi}{180 \sin^{-1}\left(1/\sqrt{1+dis^{2}}\right)} \times \sqrt{f(u)^{2} + f(v)^{2}} \qquad (3)$$

$$S(\theta(u,v)) = \frac{1-s}{2}\cos(4\theta(u,v)) + \frac{1+s}{2} \qquad (4)$$

$$\theta(u,v) = \arctan\left(\frac{f(u)}{f(v)}\right) \qquad (5)$$

where *dis* represents the perceptual viewing distance of 512mm and *s* is the symmetry parameter with a value of 0.7. Parameter *s* ensures that the floating point values in *H*(*f*) are symmetric. As *s* decreases, $S(\theta(u,v))$ decreases at approximately 45 degrees; this, in turn, increases $f'(u,v)$ and decreases *H*(*f*). The discrete horizontal and vertical frequencies are computed in (6):

$$f(u) = \frac{u-1}{\Delta \times 2N}, \ \text{for } u = 1, 2..., N; \qquad f(v) = \frac{v-1}{\Delta \times 2N}, \ \text{for } v = 1, 2..., N; \qquad (6)$$

where Δ denotes the dot pitch value of 0.25mm (approximately 100 DPI) and *N* is the number of horizontal and vertical radial spatial frequencies. A static dot pitch value of 0.25mm is utilized to compute FWM *H*(*f*).

- The default QM technique in HEVC does not take into account the specific display resolution of a target VDU, nor does it take into account the importance of transform coefficients in a TB with respect to the resolution of the target VDU.
- The dot pitch can be the same value for a multitude of VDU display resolutions depending on the pixel density of the VDU. The default QM technique does not take this into account.

Related Work

- The method in [3] involves adjustments to the parameter selection of the HVS-CSF QM technique in [1]. This refinement produces a modified FWM, from which the intra and inter QMs are derived. This technique does not take into account the target VDU's display resolution.
- In [4], the authors propose a novel intra QM method that modifies the weighting values in the QM by employing a normalized exponent variable. Similar to the method in [3], this technique does not take into account the target VDU's display resolution.

Proposed AQM Technique

- We focus on integrating the proposed intra and inter AQMs into SHVC to produce a two-layered bit-stream for SNR scalability. Lower levels of quantization are applied to the EL for the purpose of reducing video compression artifacts in high resolution VDUs (see Fig. 2).

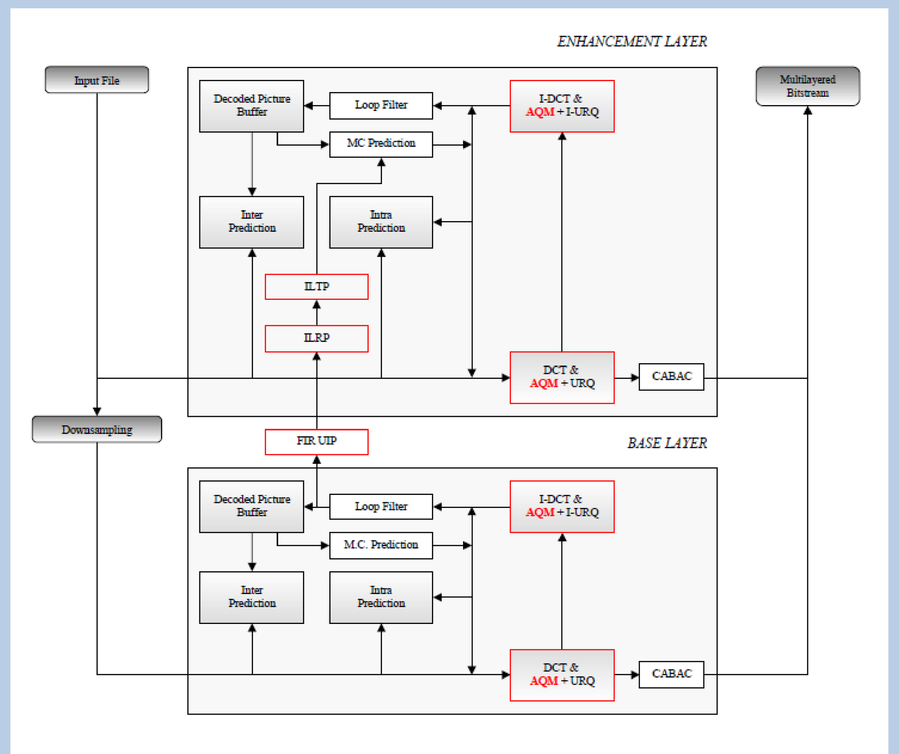

Fig. 2. A block diagram of the AQM technique in SHVC. The AQM technique is highlighted in red. This diagram shows how our method operates on an inter-layer basis in SHVC.

Our technique is based on parameter $A_{i,j}$, which is applied to each element of *H*(*f*) located at position (*i*,*j*), denoted as $H_{i,j}$. $H'_{i,j}$ is computed in (7).

$$H'_{i,j} = H_{i,j}^{A_{i,j}} \qquad (7)$$

Parameter $A_{i,j}$ is computed as a function of two parameters in (8):

$$A_{i,j}\left(d_{i,j}, w\right) = e^{-\left(\frac{d_{i,j}}{w}\right)} \in [0,1] \qquad (8)$$

where $d_{i,j}$ is the normalized Euclidean distance between the DC transform coefficient and the current coefficient located at position (*i*,*j*) in a TB, and *w* is the display resolution parameter. Euclidean distance $d_{i,j}$ is computed in (9):

$$d_{i,j} = \sqrt{\frac{(i_1 - i_2)^2 + (j_1 - j_2)^2}{(i_1 - i_{\max})^2 + (j_1 - j_{\max})^2}} \in [0,1] \qquad (9)$$

where (i_1, j_1), (i_2, j_2), (i_{max}, j_{max}) represent the position of the floating point values in *H*(*f*) associated with the DC coefficient, the current coefficient and the farthest AC coefficient, respectively. Each $A_{i,j}$ value decreases as the display resolution parameter *w* decreases. The *w* parameter is quantified in (10) and the VDU's normalized hypotenuse value *p* is computed in (11); parameter *w* rapidly decreases as *p* increases (see Fig. 3).

$$w = h_t^{-p} \in (0,1] \qquad (10) \qquad\qquad p = \frac{h_a}{h_t} \in (0,1] \qquad (11)$$

where h_t is the VDU's maximum hypotenuse value, in pixels, and where h_a is the VDU's actual hypotenuse value in the pixel domain; h_t and h_a are computed in (12) and 13), respectively:

$$h_a = \sqrt{x^2 + y^2} \qquad (12) \qquad\qquad h_t = \sqrt{x_{\max}^{2} + y_{\max}^{2}} \qquad (13)$$

where (*x*, *y*) represent the horizontal and vertical dimensions of the target VDU, respectively, and (x_{max}, y_{max}) represent, respectively, the maximum possible horizontal and vertical dimensions of the target VDU.

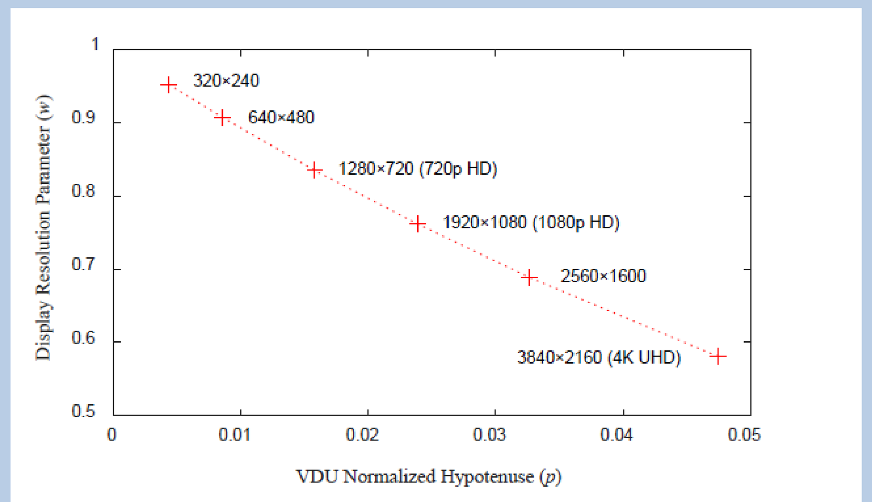

Fig. 3. Display resolution parameter *w* rapidly decreases as the VDU's normalized hypotenuse value *p*, and its display resolution, increases.

Experimental Setup

- **Reference Software:** SHVC SHM 9.0 (latest version).
- **Video Quality Metric:** BD-Rate.
- **Anchors:** Default QMs in HEVC in [1, 2] and the Sony QM technique in [3].
- **QPs:** 22, 27, 32, 37.
- **Encoding Configurations:** All Intra (Main), Low Delay (Main) and Random Access (Main).
- **Scalable Bit-Stream:** Two-layered bit-stream. The BL is aimed at HD 720p VDUs (1280×720) and the EL is aimed at 4K VDUs (3840×2160).

Results & Discussion

- In Table 1, we tabulate the average BD-Rate (for EL and BL). In Fig. 4, we show an example of the reconstruction improvements attained by our technique.

(a) (b)

Fig 4. A single frame of the HD KristenAndSara video sequence, coded using a QP = 30. Fig 4 (a) shows the improvement of the reconstruction quality of the frame using the AQMs designed for a 4K VDU with QP = 30 versus the default QMs in HEVC, as shown shown in Fig 4 (b).

Table 1. BL and EL average BD-Rate results of the proposed AQM technique compared with anchors. The results in green indicate performance improvements, the results in black indicate no improvements and the results in red indicate negative results.

Proposed AQM Technique versus Default QMs									
Class	All Intra			Low Delay B			Random Access		
	Y %	U %	V %	Y %	U %	V %	Y %	U %	V %
A (BL)	-0.6	-0.5	-0.3	-2.3	-2.6	-2.6	-3.0	-5.0	-5.4
B (BL)	-0.4	-0.1	0.0	-2.1	-2.2	-2.5	-2.9	-4.5	-5.0
C (BL)	-0.4	0.2	0.2	-3.1	-3.3	-2.9	-3.6	-5.4	-5.0
D (BL)	-0.4	0.2	0.2	-2.1	-2.3	-2.6	-2.3	-3.9	-2.9
E (BL)	-0.2	0.2	0.3	-2.9	-3.6	-4.4	-3.3	-2.8	-2.8
Average	**-0.4**	**0.0**	**0.1**	**-2.5**	**-2.8**	**-3.0**	**-3.0**	**-4.3**	**-4.2**
Class	All Intra			Low Delay B			Random Access		
	Y %	U %	V %	Y %	U %	V %	Y %	U %	V %
A (EL)	-0.8	-0.5	-0.1	-19.7	-19.1	-19.0	-37.4	-39.7	-40.1
B (EL)	-0.4	-0.1	0.0	-12.3	-14.2	-14.4	-40.4	-43.7	-44.5
C (EL)	19.0	21.1	21.5	-2.9	-2.6	-2.2	-30.0	-32.0	-31.9
D (EL)	-3.2	-3.7	-3.6	-6.3	-6.9	-7.0	-32.5	-35.2	-34.5
E (EL)	-1.6	-2.0	-2.0	-6.0	-6.7	-6.9	-29.3	-31.3	-31.1
Average	**2.60**	**2.96**	**3.16**	**-9.44**	**-9.90**	**-9.90**	**-33.92**	**-36.38**	**-36.42**

Proposed AQM Technique versus Sony QMs									
Class	All Intra			Low Delay B			Random Access		
	Y %	U %	V %	Y %	U %	V %	Y %	U %	V %
A (BL)	-2.2	0.0	-0.2	-3.2	-3.9	-4.7	-4.8	-6.9	-7.2
B (BL)	-1.1	-1.1	-1.7	-2.3	-4.7	-6.5	-3.0	-6.6	-8.2
C (BL)	-2.3	-3.9	-4.8	-1.9	-5.5	-5.3	-4.8	-11.2	-10.8
D (BL)	-2.3	-0.1	-3.2	-1.8	-2.1	-3.0	-2.7	-5.4	-5.1
E (BL)	-0.8	0.2	0.5	-3.5	-4.7	-5.7	-4.3	-3.4	-3.1
Average	**-1.7**	**-1.0**	**-1.9**	**-2.5**	**-4.2**	**-5.0**	**-3.9**	**-6.7**	**-6.9**
Class	All Intra			Low Delay B			Random Access		
	Y %	U %	V %	Y %	U %	V %	Y %	U %	V %
A (EL)	-2.3	0.0	-0.1	-52.7	-56.0	-56.6	-56.5	-58.7	-59.2
B (EL)	-1.3	-1.1	-1.6	-36.2	-42.9	-43.0	-50.6	-55.5	-56.9
C (EL)	-45.0	-47.8	-48.2	-19.9	-24.1	-23.7	-44.4	-49.1	-49.1
D (EL)	-52.6	-56.9	-57.2	-28.5	-31.6	-31.8	-39.1	-42.5	-42.3
E (EL)	-33.5	-35.4	-35.0	-6.6	-6.9	-7.1	-33.5	-35.4	-35.0
Average	**-26.94**	**-28.24**	**-28.42**	**-28.78**	**-32.30**	**-32.44**	**-44.82**	**-48.24**	**-48.50**

Most Significant BD-Rate Improvement Results Compared with Anchors (EL versus EL)

- **Sequence:** Class B HD (Default QMs in HEVC) & Class A UHD 4K Sequence (Sony QMs).
- **Encoding Configuration:** Random Access (Main).
- **BD-Rate versus Default QMs in HEVC:** -40.4% (Y), -43.7% (Cb) and -44.5% (Cr).
- **BD-Rate versus Sony QMs:** -56.5% (Y), -58.7% (Cb) and -59.2% (Cr).

The very high EL versus EL BD-Rate improvements are mainly due to the increased accuracy of inter-layer prediction for the EL. That is, our technique yields improved reconstruction of the BL, which allows for a more accurate prediction for the EL.

Encoding & Decoding Times Improvement Results Compared with Anchors

- **Encoding & Decoding Times versus Default QMs in HEVC:** -0.75% and -4.67%, respectively.
- **Encoding & Decoding Times versus Sony QMs :** -1.19% and -2.82%, respectively.

A more accurate prediction of the EL from the BL decreases the workload of the entropy coding process, which leads to improved encoding and decoding times in the proposed AQM technique.

Conclusions

A novel AQM technique for Scalable HEVC is proposed to improve video reconstruction quality, thereby reducing the visibility of compression artifacts on high resolution VDUs. Compared with anchors, the proposed method yields important coding efficiency and visual quality improvements, with a maximum luma BD-Rate improvement of 56.5% in the EL. In addition, the proposed technique attains modest encoding and decoding time improvements.

[1] C. Wang, S. Lee and L. Chang, "Designing JPEG quantization tables based on human visual system," *IEEE International Conference on Image Processing*, Kobe, Japan, 1999, vol. 2, pp. 376-380.
[2] V. Sze, M. Budagavi and G. J. Sullivan, "Quantization Matrix," in *High Efficiency Video Coding (HEVC): Algorithms and Architecture*, Springer International Publishing, 2014, pp. 158-159.
[3] M. Haque, A. Tabatabai and Y. Morigami, "HVS model based Default Quantization Matrices," *Document JCTVC-G880*, 2011, pp. 1-13.
[4] S. Jeong and B. Jeon, "Newer Quantization Matrices for HEVC," *Document JCTVC-I0126*, 2012, pp. 1-8