HVS-BASED PERCEPTUAL QUANTIZATION MATRICES FOR HDR HEVC VIDEO CODING FOR MOBILE DEVICES

Dan Grois and Alex Giladi

Comcast, Philadelphia PA, USA
Correspondance Email: Dan_Grois@comcast.com

ABSTRACT

This work presents development and implementation of novel perceptual quantization matrices for coding High Dynamic Range (HDR) mobile device-based video content. The proposed perceptual quantization matrices are based on Human Visual System (HVS) and utilized for reducing video transmission bit-rate and for optimizing perceived visual quality of video content to be displayed on mobile devices, such as tablets and smartphones.

According to the proposed video coding scheme, perceptual quantization matrices are first calculated based on Human Visual System (HVS) characteristics and on predefined viewing conditions, and then utilized during the encoding loop for removing non-perceptible visual information, while making an especial emphasis on the Ultra High Definition (UltraHD) resolution and H.265/MPEG-HEVC video coding standard.

Based on extensive experimental results, visual quality of the HDR UltraHD video content is significantly improved, for substantially the same bit-rate, in terms of the popular objective quality metric SSIMPlus. On the other hand, *the video transmission bit-rate is significantly reduced by up to about 25%*, while keeping visual quality of the video content, to be displayed on a mobile device screen, substantially at the same level.

INTRODUCTION

There is currently a strong demand for high resolution video content, particularly for the high-definition (HD) and UltraHD video content for a variety of mobile devices, such as tablets, smartphones and even smartwatches. According to the recent Cisco® report (1), the IP video traffic is expected to be 82% of all Internet traffic by 2022, and there is a continuous need to decrease video transmission bit-rate, especially for delivery over wireless or cellular networks without reducing visual presentation quality.

In addition, the HDR UltraHD video content is recently attracting a lot of attention due the relatively high luminance levels and fine shadow details, which extend much beyond conventional Standard Dynamic Range (SDR) content. The HDR technology makes it possible to present highly bright signals along with very dark signals on the same video frame, thereby providing a high contrast ratio within the same image. In addition, the HDR video content is usually combined with a Wide Color Gamut (WCG), such as BT.2020 (2), (3), thereby enabling to present video with a significantly extended color spectrum.

Particularly, HDR has gained its popularity after the development and approval of the High Efficiency Video Coding (HEVC) standard (4), i.e. H.265/MPEG-HEVC (2)-(9), in 2013. As known, HEVC was especially designed for coding of HD and UltraHD video content with a much larger coding gain (10)-(13) compared to its predecessor H.264/MPEG-AVC (14), thereby reducing both spatial and temporal video content redundancies in a much more efficient way, which in turn significantly assisted in compression of the HDR UltraHD video content. However, coding of the HDR video content still remains challenging due to users' demands for high visual quality, which in turn requires allocating more bits and increasing a video coding depth (e.g., from 8 bits to 10 bits). In addition, the transmission bandwidth is normally limited due to a typical limitation of the existing network infrastructure, especially in case of the transmission over wireless/cellular networks. As a result, in order to stay within the transmission bandwidth limits, the high-resolution HDR video content is often compressed with visually perceived coding artifacts. Moreover, encoding of the HDR content normally consumes significant computational resources due to a requirement to preserve fine details within the HDR video. Therefore, there is a strong demand to improve perceived visual quality of the compressed HDR video substantially without increasing its bit-rate (15)-(18).

One the most popular approaches for improving video quality is related to considering spatial frequency sensitivity of Human Visual System (HVS). As known, the HVS system is a part of the central nervous system, which enables processing of visual details and generating non-image photo response functions by obtaining and processing visible information (19)-(21). Thus, for example, during the coding process, the values of the Discrete Cosine Transform (DCT) frequency coefficients can be attenuated by applying quantization matrices: i.e. lower spatial frequencies are usually quantized with smaller quantization parameters (QPs), while higher spatial frequencies - with larger QPs. However, the improvements in visual quality of the related state-of-the-art approaches (22),(23) are relatively small, and more efficient solutions are desirable. In addition, most of the state-of-the-art pre-processing methods are designed for the relatively low-resolution SDR video content, and as a result, these methods found to be mostly inefficient for HDR UltraHD. In turn, this is also true for the coding schemes that aim to remove fine details below a predefined visibility threshold, which is referred as Just Noticeable Difference (JND) (24)-(29). As a result, the state-of-the-art JND-based schemes do not provide sufficient video quality improvement for the HDR UltraHD video content as well.

This work is a continuation work of (30) with a special emphasis on mobile devices, such as tablets and smartphones, which have relatively small screens. In this work, authors present a novel coding scheme, in which highly efficient perceptual quantization matrices for coding the HDR UltraHD *mobile device video content* are developed, thereby allowing to significantly *reduce cellular/wireless network transmission bit-rate* by achieving a much better bit allocation balance among brighter and darker video scenes to be presented on relatively small mobile device screens.

Specifically, the bit-rate is reduced by up to about 25% based on SSIMPlus objective quality metrics (31)(32) in terms of BD-BR rate (33), while keeping visual quality substantially at the same level. This paper is organized as follows. First, the authors provide the background and detailed overview of perceptual quantization matrices. Then, the proposed perceptual quantization matrices for coding of the HDR UltraHD video content are presented, the design and development of which was motivated and inspired by the Contrast Sensitivity Function

(CSF) of a human visual system as defined by Barten in (19)-(21). After that, the test methodology and evaluation setup are discussed, followed by extensive experimental results and conclusions.

PERCEPTUAL QUANTIZATION MATRICES: BACKGROUND AND OVERVIEW

The human visual system (HVS) is considered to be a very complex system, while a level of contrast that is required to generate a response perceived by HVS is known as a contrast threshold (34) of a sinusoidal luminance pattern. In turn, an inverse of this threshold is called "contrast sensitivity" (35), and it varies as a function of a spatial frequency (34). The relationship between the spatial frequency and contrast sensitivity is known as a contrast sensitivity function (CSF) that differs for achromatic and chromatic scenes (34). In addition, the HVS is more sensitive to low spatial frequencies than to high spatial frequencies (36)-(39), and by assuming that HVS is isotropic, it can be modeled as a nonlinear point transformation that is followed by a Modulation Transfer Function (MTF):

$$H(f) = a(b + c \cdot f) \cdot \exp\left(-c(f)^{d}\right) \qquad [1]$$

where *a, b, c,* and *d* are constants and *f* is a radial frequency in cycles per degree. This HVS-based model was first proposed by Mannos & Sakrison in 1974 (40), and then, in 1987, modified by Daly (36), thereby setting *a, b, c,* and *d* constants to the following values: *a=2.2, b=0.192, c=0.114* and *d=1.1* and as a result, obtaining the following MTF:

$$H(x,y) = \begin{cases} 2.2 \cdot \left(0.192 + 0.114 \cdot \tilde{f}(x,y)\right) \cdot exp\left(-(0.114 \cdot \tilde{f}(x,y))^{11}\right), & if\ \tilde{f}(x,y) > f_{peak} \\ 1.0\,, & otherwise \end{cases} \qquad [2]$$

where f_{peak} is the exponential radial peak frequency, and $\tilde{f}(x,y)$ is a radial frequency in cycles per degree.

Later, this approach was practically used in developing a HVS-based quantization table for the JPEG still image compression standard (41),(42). Authors of (42) derived this table by incorporating a HVS model developed by Daly (36)-(39) with an uniform quantizer, and further claiming that by replacing the JPEG quantization table with their HVS-based quantization table, obvious perceptual quality improvements are achieved. More specifically, the authors of (42) applied a 1st order low-contrast MTF of the HVS model proposed by Daly for generating a HVS-based quantization table for the baseline JPEG image compression standard (42), as follows below.

First, for obtaining a desired HVS-based quantization table, it is required to express radial frequencies as discrete horizontal and vertical frequencies in a DCT domain (42). So, let's denote horizontal discrete frequencies in the above-mentioned DCT domain as follows (42):

$$f(x) = \frac{x-1}{pitch \times 2N},\ for\ \ x = 1, 2, 3, \dots, N \qquad [3]$$

where *N* is a number of horizontal frequencies, and $pitch$ is the display dot pitch. Similarly, the vertical discrete frequencies are defined by:

$$f(y) = \frac{y-1}{pitch \times 2N},\ for\ \ x = 1, 2, 3, \dots, N \qquad [4]$$

where N is a number of horizontal frequencies that is the same as the number of horizontal frequencies.

After converting $f(x)$ and $f(y)$ to corresponding radial frequencies, and scaling them to the viewing distance in millimeters (mm), this results in the following equation 42:

$$f(x,y) = \frac{\pi \times \sqrt{f(x)^2 + f(y)^2}}{180 sin^{-1}(\frac{1}{\sqrt{1+D^2}})} \qquad [5]$$

In order to consider the viewing angle θ and to consider the visual MTF fluctuations as a function of it, the frequencies $f(x,y)$ are normalized by an angular-dependent function $S\big(\theta(x,y)\big)$ as follows (42):

$$\tilde{f}(x,y) = \frac{f(x,y)}{S\big(\theta(x,y)\big)} \qquad [6]$$

where $S\big(\theta(x,y)\big)$ is defined by (36)(42)(43):

$$S\big(\theta(x,y)\big) = \frac{1-w}{2} cos(4\theta(x,y)) + \frac{1+w}{2} \qquad [7]$$

and

$$\theta(x,y) = arctan\left(\frac{f(y)}{f(x)}\right) \qquad [8]$$

It is further assumed that the display dot pitch is 0.25 millimeters (mm), i.e. about 100 dots per inch, the aspect ratio is 1:1, the display size of ~128x128 millimeters (mm) is required to display 512x512 image, and the corresponding viewing distance is set to four times image height (42)(44), which is 512mm in this case. Also, as it can be seen from equation [7], when the value of *w* decreases, then the angular-dependent function $S\big(\theta(x,y)\big)$ decreases around the angle of 45 degrees, which is turn result in the decrease of $H(x,y)$ values (42)(40). This is actually well-known as an oblique effect of the human visual system (HVS), since it is much less sensitive to objects/details located under the 45 degrees angle (45).

The HEVC video coding standard (5) allows usage of perceptually-tuned frequency-dependent quantization matrices, instead of applying a constant quantization parameter (QP) on each coding block. These matrices better suit the HVS characteristics by allowing to quantize higher frequencies in a stronger manner, while their sizes vary from 4x4 to 32x32. However, the specification of the HEVC standard (5) only defines default quantization matrices for 4x4 and 8x8 transform blocks, and the rest of the matrices, i.e. for transform block sizes of 16×16 and 32×32, are obtained by upsampling the original 8×8 perceptual quantization matrix respectively. More specifically, the original 8×8 matrix is replicated: each block in the 8×8 matrix is replicated to the 2×2 area of the 16×16 transform block and to the 4×4 area of the 32×32 transform block.

Therefore, depending on the transform block type (i.e. used for *Intra* or *Inter*-picture prediction) and transform block size (i.e. 4×4, 8×8, 16×16 or 32×32), the HEVC standard employs twenty quantization matrices: 8 matrices for *Y* (*Luma*) component and 6 matrices for each of *Cb* and *Cr* (*Chroma*) components.

In addition, HEVC allows the use of other quantization matrix values (i.e. *customized* quantization matrix values). For that, the above-mentioned customized quantization matrix values can be transmitted within the HEVC bitstream Sequence Parameter Set (SPS) or

Picture Parameter Set (PPS), while coding these customized values by using so called Differential Pulse Code Modulation or in short DPCM (see (5) for more details). Similarly, the 16×16 and 32×32 quantization matrices are obtained by upsampling corresponding 4×4 and 8×8 quantization matrices.

In spite of the fact that the HEVC default perceptual quantization matrices are based on HVS, they were initially developed and tested on low-resolution JPEG images (42), such as 512×512 pixels. Therefore, they almost didn’t provide any benefits for UltraHD video content, that has the 3840x2160 resolution in terms of luma samples, which is the most popular resolution nowadays. As a result, this is currently also a reason for the relatively low popularity of these default perceptual quantization matrices, which most often are not used at all, especially for the mobile device content coding.

In the following section, the design and development of novel perceptual quantitation matrices for coding of the HDR UltraHD mobile video content is presented, further being inspired by investigating the contrast sensitivity function (CSF) of a human visual system. The novel perceptual quantization matrices significantly improve perceived video quality without a need for pre-processing and without an increase in coding computational complexity.

PROPOSED NOVEL PERCEPTUAL QUANTIZATION MATRICES: DESIGN GUIDELINES AND DEVELOPMENT ASPECTS

As discussed in the previous section, the contrast sensitivity of the human eyes, and more generally – of the human visual system as the whole, is one of the main factors how humans perceive achromatic or chromatic images. Therefore, when developing HVS-based models, it is especially important to determine the CSF as accurately as possible. For that, it is important to consider substantially all known HVS characteristics that have any impact on the CSF. With this regard, in addition to the HVS-based models developed by Mannos & Sakrison (40) and Daly (36)-(38) at the end of the 20th century, Barten in his paper from 2004 (21) proposes a more accurate HVS-based physical model/formula for the contrast sensitivity of the human eye. Particularly, in his work (21), Barten considers a plurality of HVS parameters, such as photon noise, neural noise, external noise, lateral inhibition, eye pupil diameter, eye pupil size, angular size of the object, luminance conditions, etc.

As a result, Barten’s HVS-based CSF model shown in (21) and presented below, is considered to be the most accurate for representing the HVS contrast sensitivity, and considered to be the best CSF model to date.

$$CSF = \frac{1}{m_t} = \frac{M_{opt}(u)}{k \cdot \sqrt{\frac{2}{T} \cdot \left(\frac{1}{X_0^2} + \frac{1}{X_{max}^2} + \frac{u^2}{N_{max}^2} \right) \cdot \left(\frac{1}{\eta p E} + \frac{\Phi_0}{1 - e^{-(u/u_0)^2}} \right)}} \qquad [9]$$

where X_0 is an angular object size and the optical Modulation Transfer Function (MTF) of the human eye is defined as $M_{opt}(u) = e^{-2\pi^2\sigma^2u^2}$. Also, the dependency on the pupil size is defined by $\sigma = \sqrt{\sigma_0^2 + (C_{ab} \cdot d)^2}$, where d is the pupil diameter, C_{ab} is a constant describing the increase of σ, and σ_0 is a constant as well.

In turn, the pupil diameter varies according to the average luminance of the observed area and can be approximated (21)(46) as $d = 5 - 3 \cdot tanh\left\{0.4 \cdot log\left(\frac{L \cdot X_0^2}{40^2}\right)\right\}$, while the retinal illuminance is defined as $E = \frac{\pi d^2}{4} L \cdot \left\{1 - \left(\frac{d}{9.7}\right)^2 + \left(\frac{d}{12.4}\right)^4\right\}$.

Further, for a typical scenario, the constants can be defined as follows (21): the signal to noise ratio k is set to 3.0; σ_0 is equal to 0.5 arcmin; T is set to 0.1 seconds and C_{ab} to 0.08 arcmin/mm; the quantum efficiency of the eye η is 0.03; the spectral density of the neural noise Φ_0 is set to $3 \cdot 10^{-8}$ sec deg^2; the maximum angular size of the integrated area of the noise X_{max} is equal to 12 degrees; the maximum number of cycles over which the eye can integrate the information N_{max} is set to 15 cycles; the spatial frequency above which the lateral inhibition ceases u_0 is set to 7 cycles/degree, and the photon conversion factor p is approximately equal to $1.2 \cdot 10^6$ photons/sec/deg^2/Td (21).

However, as it is seen, the Barten's HVS-based CSF model of equation [10] is very complex and its usage for an accurate determining of efficient perceptual quantization matrices to be employed during the video coding loop, either in HEVC or in other emerging video coding standards, is found to be very challenging.

Therefore, at the 1st step, the authors of this work designed the above-mentioned perceptual quantization matrices, to be employed during the video coding loop, by fitting the Daly's HVS-based model of equation [2] into the Barten's HVS-based CSF model equation [9], and as a result of this fitting, generating a corresponding multiplier for each coefficient of the Frequency Weighting Matrix (FWM), which is defined as H(x,y) in equation [2]. It should be noted that in this work, a special emphasis was made on developing perceptual quantization matrices which are optimized for coding of video content to be presented on *relatively small HDR UltraHD mobile device displays*, which typically vary between 5 to 13 inch in diameter.

In turn, the FWM is adjusted to the Barten's HVS-based CSF model, thereby resulting in a *CSF-tuned human visual FWM*. Further, it should be noted that deriving the corresponding CSF-tuned human visual FWM directly from the above-mentioned very complex Barten model of equation [9] didn't lead to a desired result, and therefore the authors have chosen first to fit the simpler Daly's model of equation [2] into that of Barten, and then to derive the CSF-tuned human visual FWM by respectively adjusting coefficients of FWM. By such a way, the CSF-tuned human visual FWM have been determined in a more practical and in a more precise manner, further supported by the detailed experimental results in the due course of this paper.

Then, at the 2nd step, the CSF-tuned human visual FWM coefficients (which are normalized into a range between *0* and *1*) are empirically optimized for *relatively small HDR UltraHD mobile device displays* by gradually attenuating high frequencies in a much stronger manner than low frequencies, and further giving priority to luminance (*Luma*) over chrominance (*Chroma*) due to the fact that human eye is more sensitive to *Luma* changes than that of *Chroma* (20). Finally, the corresponding *Luma* and *Chroma* perceptual quantization matrices coefficients for both intra and inter-picture prediction coding are derived respectively from the above-mentioned empirically optimized coefficients of the CSF-tuned human visual FWM.

Thus, upon determining FWM, the HVS-based luminance quantization table (QT) of *Figure 1(b)* is derived from the following equation:

$$QT_{Luma} = round\left(\frac{QP}{H(x,y)}\right) = round\left(\frac{16}{H(x,y)}\right) \qquad [10]$$

where each coefficient is rounded to an integer value according to (44) with the quantization parameter (QP) of *16* (that is set to *16* due to providing the best results in terms of visual quality). It should be noted that the perceptual quantization table is more often called as a perceptual “quantization matrix” (QM), or a perceptual “scaling list”, and therefore in this work, these terms are used interchangeably.

Figure 1 below presents an optimized 8x8 CSF-tuned human visual Frequency Weighting Matrix (FWM) for Intra-picture prediction along with the corresponding *Luma* perceptual quantization matrix. It should be noted that the *Luma* perceptual quantization matrix for intra-picture prediction of *Figure 1(b)* is obtained by applying equation [10] on the CSF-tuned human visual FWM of *Figure 1(a)*, while the QP value is set to “*16*” for providing best results in terms of visual quality.

1.0000	**0.5333**	**0.4103**	**0.3721**	**0.3077**	**0.2807**	**0.2500**	**0.1758**
0.5333	**0.4103**	**0.3721**	**0.3077**	**0.2424**	**0.1860**	**0.1702**	**0.1429**
0.4103	**0.3721**	**0.3077**	**0.2162**	**0.1702**	**0.1667**	**0.1368**	**0.1203**
0.3721	**0.3077**	**0.2162**	**0.1569**	**0.1495**	**0.1280**	**0.1127**	**0.0884**
0.3077	**0.2424**	**0.1702**	**0.1495**	**0.1280**	**0.1067**	**0.0847**	**0.0615**
0.2807	**0.1860**	**0.1667**	**0.1280**	**0.1067**	**0.0714**	**0.0497**	**0.0473**
0.2500	**0.1702**	**0.1368**	**0.1127**	**0.0847**	**0.0497**	**0.0426**	**0.0383**
0.1758	**0.1429**	**0.1067**	**0.0884**	**0.0615**	**0.0473**	**0.0383**	**0.0325**

(a)

16	**30**	**39**	**43**	**52**	**57**	**64**	**91**
30	**39**	**43**	**52**	**66**	**86**	**94**	**112**
39	**43**	**52**	**74**	**94**	**96**	**117**	**133**
43	**52**	**74**	**102**	**107**	**125**	**142**	**181**
52	**66**	**94**	**107**	**125**	**150**	**189**	**260**
57	**86**	**96**	**125**	**150**	**224**	**322**	**338**
64	**94**	**117**	**142**	**189**	**322**	**376**	**418**
91	**112**	**150**	**181**	**260**	**338**	**418**	**492**

(b)

Figure 1 - (a) The 8x8 *Luma* CSF-tuned human visual Frequency Weighting Matrix for *Intra*-picture prediction**;** (b) The corresponding 8x8 *Luma* perceptual quantization matrix for *Intra*-picture prediction.

Similarly, the *Chroma* CSF-tuned human visual Frequency Weighting Matrix for Intra-picture prediction is presented in *Figure 2(a)* below, along with the corresponding *Chroma* perceptual quantization matrix in *Figure 2(b)*:

1.0000	**0.5000**	**0.3721**	**0.3077**	**0.2319**	**0.2025**	**0.1758**	**0.1067**
0.5000	**0.3721**	**0.3077**	**0.2319**	**0.1616**	**0.1127**	**0.1013**	**0.0796**
0.3721	**0.3077**	**0.2319**	**0.1429**	**0.1013**	**0.0958**	**0.0724**	**0.0615**
0.3077	**0.2319**	**0.1429**	**0.0914**	**0.0833**	**0.0669**	**0.0567**	**0.0478**
0.2319	**0.1616**	**0.1013**	**0.0833**	**0.0669**	**0.0523**	**0.0448**	**0.0422**
0.2025	**0.1127**	**0.0958**	**0.0669**	**0.0523**	**0.0444**	**0.0395**	**0.0355**
0.1758	**0.1013**	**0.0724**	**0.0567**	**0.0448**	**0.0395**	**0.0349**	**0.0304**
0.1067	**0.0796**	**0.0523**	**0.0478**	**0.0422**	**0.0355**	**0.0304**	**0.0281**

(a)

16	**32**	**43**	**52**	**69**	**79**	**91**	**150**
32	**43**	**52**	**69**	**99**	**142**	**158**	**201**
43	**52**	**69**	**112**	**158**	**167**	**221**	**260**
52	**69**	**112**	**175**	**192**	**239**	**282**	**335**
69	**99**	**158**	**192**	**239**	**306**	**357**	**379**
79	**142**	**167**	**239**	**306**	**360**	**405**	**451**
91	**158**	**221**	**282**	**357**	**405**	**458**	**526**
150	**201**	**306**	**335**	**379**	**451**	**526**	**570**

(b)

Figure 2 - (a) The 8x8 *Chroma* CSF-tuned human visual Frequency Weighting Matrix for *Intra*-picture prediction**;** (b) The corresponding 8x8 *Chroma* perceptual quantization matrix for *Intra*-picture prediction.

In turn, the 8x8 *Luma* and *Chroma* perceptual quantization matrices for intra-picture prediction are downsampled, resulting in the corresponding 4x4 perceptual quantization matrices:

16	**39**	**58**	**64**

16	**43**	**69**	**91**

39	52	94	117
52	94	125	189
64	117	189	376

(a)

43	69	158	221
69	158	239	357
91	221	357	458

(b)

Figure 3 - (a) The 4x4 *Luma* perceptual quantization matrix for *Intra*-picture prediction; (b) The 4x4 *Chroma* perceptual quantization matrix for *Intra*-picture prediction.

Similarly, for the Inter-picture prediction, the FWM and perceptual quantization matrix for both *Luma* and *Chroma* are presented in *Figures 4* and *5*, respectively.

1.0000	0.5294	0.4186	0.3600	0.3051	0.2727	0.2432	0.1731
0.5294	0.4186	0.3600	0.3051	0.2368	0.1818	0.1682	0.1406
0.4186	0.3600	0.3051	0.2143	0.1682	0.1607	0.1324	0.1176
0.3600	0.3051	0.2143	0.1565	0.1463	0.1250	0.1098	0.0857
0.3051	0.2368	0.1682	0.1463	0.1250	0.1047	0.0826	0.0600
0.2727	0.1818	0.1607	0.1250	0.1047	0.0700	0.0483	0.0463
0.2432	0.1682	0.1324	0.1098	0.0826	0.0483	0.0414	0.0373
0.1731	0.1406	0.1047	0.0857	0.0600	0.0463	0.0373	0.0317

(a)

18	34	43	50	59	66	74	104
34	43	50	59	76	99	107	128
43	50	59	84	107	112	136	153
50	59	84	115	123	144	164	210
59	76	107	123	144	172	218	300
66	99	112	144	172	257	373	389
74	107	136	164	218	373	435	482
104	128	172	210	300	389	482	567

(b)

Figure 4 - (a) The 8x8 *Luma* CSF-tuned human visual Frequency Weighting Matrix for *Inter*-picture prediction**;** (b) The corresponding 8x8 *Luma* perceptual quantization matrix for *Inter*-picture prediction.

1.0000	0.5000	0.3600	0.3051	0.2278	0.1978	0.1731	0.1047
0.5000	0.3600	0.3051	0.2278	0.1565	0.1098	0.0978	0.0783
0.3600	0.3051	0.2278	0.1406	0.0978	0.0938	0.0709	0.0600
0.3051	0.2278	0.1406	0.0896	0.0814	0.0655	0.0554	0.0466
0.2278	0.1565	0.0978	0.0814	0.0655	0.0508	0.0437	0.0411
0.1978	0.1098	0.0938	0.0655	0.0508	0.0434	0.0385	0.0347
0.1731	0.0978	0.0709	0.0554	0.0437	0.0385	0.0340	0.0298
0.1047	0.0783	0.0508	0.0466	0.0411	0.0347	0.0298	0.0274

(a)

18	36	50	59	79	91	104	172
36	50	59	79	115	164	184	230
50	59	79	128	184	192	254	300
59	79	128	201	221	275	325	386
79	115	184	221	275	354	412	438
91	164	192	275	354	415	468	519
104	184	254	325	412	468	529	605
172	230	354	386	438	519	605	658

(b)

Figure 5 - (a) The 8x8 *Chroma* perceptual Frequency Weighting Matrix for *Inter*-picture prediction; (b) The corresponding 8x8 *Chroma* perceptual quantization matrix for *Inter*-picture prediction.

It should be noted that the QP value for calculating the Inter-picture prediction perceptual quantization matrices by applying equation [10] is set to the value of "*18*" (instead of the value of "*16*", as used for calculating the Intra-picture prediction perceptual quantization matrices in *Figures 1* and *2*), since in case of the inter-picture prediction, the coding artifacts that may appear due to quantization are less noticeable by the human visual system.

In addition, as already noted, the rest of matrices (i.e. for transform block sizes of 16×16 and 32×32) are obtained by upsampling the original 8×8 perceptual quantization matrix respectively - the original 8×8 perceptual quantization matrix is replicated: each block in the 8×8 matrix is replicated to the 2×2 area of the 16×16 transform block and to the 4×4 area of the 32×32 transform block. In turn, the 8x8 *Luma* and *Chroma* perceptual quantization matrices for inter-picture prediction are downsampled, resulting in the following corresponding 4x4 perceptual quantization matrices:

18	43	59	74
43	59	107	136
59	107	144	218

18	50	79	104
50	79	184	254
79	184	275	412

74	**136**	**218**	**435**	**104**	**254**	**412**	**529**

(a) (b)

Figure 6 - (a) The 4x4 *Luma* perceptual quantization matrix for *Inter*-picture prediction; (b) The 4x4 *Chroma* perceptual quantization matrix for *Inter*-picture prediction.

In the following sections, the test methodology, evaluation setup and experimental results are presented and discussed in detail.

TEST METHODOLOGY AND EVALUATION SETUP

For H.265/MPEG-HEVC-based encoding, the authors employed one of the most popular HEVC-based encoders, i.e. the x265 encoder (47), the development of which was inspired by the successful development of its predecessor – the x264 video codec (48) that is based on H.264/MPEG-AVC (12),(14). In addition, x265 is considered to be the most popular practical HEVC-based implementation, further providing a flexible trade-off between coding efficiency and computational complexity (12).

As known, the x265 encoder is operated by means of a Command Line Interface (CLI) (47). In this work, for obtaining larger coding gains, the x265 encoder was set to the *2-pass mode*, further adding "*--slow-firstpass*" and "*--me=star*" commands within the x265 CLI during the *1st pass* 47). In addition, the following commands "*--bframes=5*", "*--ref=4*", "*--weightb*", "*--subme 7*", "*--qg-size=8*" were provided within the x265 CLI during the encoding process (47).

For the purpose of evaluation, the SSIMPlus (31),(32) objective video quality metric for relatively small mobile device displays was used (more specifically, for a 10.5-inch HDR display), since the evaluation by means of SSIMPlus is currently considered to be the most close to performing subjective tests, thereby covering many psycho-visual factors of a human visual system (HVS). For that, the SSIMPlus metrics makes use of "viewer intelligence", further considering a plurality of parameters, such as human visual system properties, typical viewing conditions, typical display device properties, etc. In addition, the SSIMPlus objective metrics considers temporal elements during its quality assessment and is further based on the so called "short-memory effect" of the human brain. This effect actually considers the fact that the time interval is content-adaptive, since humans tend to memorize high quality simple content much longer than low quality complex content (31),(32).

In addition, for generating test results, the Bjøntegaard-Delta bit-rate (BD-BR) measurement method was used for the R-D performance assessment in order to calculate average bit-rate differences between R-D curves for the same distortion (e.g., for the same SSIMPlus values) (33). It should be noted that *negative BD-BR values* indicate actual bit-rate savings (10)-(13), in contrast to *positive BD-BR values*, which indicate the required overhead in bit-rate to achieve the same SSIMPlus values.

EXPERIMENTAL RESULTS

For obtaining experimental results presented in this section, a special emphasis was made on video sequences having a 10-bit sample representation and HDR UltraHD spatial resolution (particularly, the 4K resolution – i.e. 2160p, or more specifically, the 3840x2160 resolution in terms of luma samples), as presented in *Table 1* below.

Tested Video Sequences	Content Type	No. of Frames	Frame Rate per Second	Resolution	Dynamic Range
“Lucy“ (provided by NBCUniversal®)	action scenes, fast motion scenes, mixed content	8425	24	3840x2160	HDR
“Everest“ (provided by NBCUniversal®)	mountains views, snow scenes, slow motion scenes	7202	23.98	3840x2160	HDR
“Warcraft” (provided by NBCUniversal®)	computer-generated content, fast motion scenes	8177	23.98	3840x2160	HDR
“Regatta” (provided by UltraHD forum®)	water scenes, fast motion scenes	5841	59.94	3840x2160	HDR

Table 1 - HDR UltraHD test video sequences.

Also, for each of these video sequences, four QP values were used: 22, 27, 32, and 37, which are QP values used for the I-frame coding, according to the Common Test Conditions (CTC) (49) with the Intra-frame period set to 1 second. Further, for obtaining SSIMPlus values, the x265 encoder was tuned for the best performance in terms of SSIM by means of the “*---tune ssim*” CLI command (47).

Table 2 below presents, in its right column, the BD-BR SSIMPlus bit-rate savings for the HEVC encoding with the proposed perceptual quantization matrices (QMs) that are optimized for *mobile devices* versus HEVC encoding with the default QMs, as defined in the HEVC specification (5)-(9). In addition, in the middle column, are presented the BD-BR SSIMPlus bit-rate savings for the HEVC encoding with the proposed perceptual QMs that are optimized for *mobile devices* versus HEVC encoding with the constant QP as defined in CTC (49) – i.e. without employing the default HEVC QMs.

Tested Video Sequences	BD-BR SSIMPlus Proposed QMs vs. no QMs	BD-BR SSIMPlus Proposed QMs vs. Default HEVC QMs
Lucy	**-16.8%**	-15.5%
Everest	**-22.2%**	-21.4%
Warcraft	**-7.8%**	-5.9%
Regatta	**-23.9%**	-22.2%

Table 2 - BD-BR SSIMPlus bit-rate savings for the HEVC encoding with the proposed perceptual QMs *versus* the HEVC encoding with default QMs, and versus the HEVC encoding with the constant QP per CTC (49).

As can be clearly seen from *Table 2*, by employing the proposed perceptual QMs, significant coding gains of up to about 25% are achieved. It should be noted that for the “Regatta” video sequence the coding gain is the most significant - the “Regatta” video content is considered to be hard to encode, since it contains many water scenes, and the proposed perceptual QMs perform much better for such content.

Below, for example, quality scores for encoding the “Regatta” video sequence with target bit rates varying between 2Mb and 5Mb are presented, thereby showing the SSIMPLus score in a range between *0* and *100*, while the larger the number - the better the video quality is (*100* is the best possible quality).

Target Bit Rate	**SSIMPlus (no QMs)**	**SSIMPlus (Default HEVC QMs)**	**SSIMPlus (Proposed QMs)**	**Minimal SSIMPlus (no QMs)**	**Minimal SSIMPlus (Default HEVC QMs)**	**Minimal SSIMPlus (Proposed QMs)**
2Mb	76.82	76.85	**78.61**	48	48	**54**
3Mb	82.06	82.10	**83.48**	58	58	**65**
4Mb	84.82	84.87	**86.10**	65	66	**72**
5Mb	87.20	87.28	**88.48**	73	73	**77**

Table 3 - SSIMPlus scores for encoding the Regatta video sequence with target bit rates of 2Mb, 3Mb, 4Mb and 5Mb, while employing the proposed perceptual QMs, HEVC default QMs and encoding without QMs (i.e. with a constant QP only).

As is clearly seen from *Table 3*, when the proposed perceptual QMs that are optimized for *mobile devices* are employed, the SSIMPlus score is significantly higher – i.e. it is up to about 2 *points* compared to encoding with a constant QP according to CTC (49), i.e. marked as “no QMs” in the above table. Similarly, the encoding with the default HEVC QMs provides a little improved visual quality compared to the above-mentioned constant QP encoding, but still much worse quality compared to the encoding with the proposed perceptual QMs. In addition, the minimal SSIMPlus score increased by a very significant number of up to *7 points* for the bit-rate of 3Mb, which is visually clearly noticeable. Conclusions are provided in the section below.

CONCLUSIONS

In this work, novel perceptual quantization matrices for HDR video coding of content to be displayed on mobile devices, such as tablets and smartphones, have been developed and discussed in detail. In addition, a special emphasis was made on the UltraHD resolution, such as the 3840x2160 (4K) resolution in terms of luma samples, and on the H.265/MPEG-HEVC video coding standard. The development of the above-mentioned novel perceptual quantization matrices has been motivated by the Daly HVS-based perceptual model, which was further fitted into the more advanced and more complex Barten model (that incorporates a variety of HVS parameters) for much more accurate generation of these matrices. As a result, visual quality of the HDR UltraHD mobile device video content encoded by employing the above-mentioned novel perceptual quantization matrices (which are especially optimized for relatively small screens) is significantly improved, for substantially the same bit-rate. Specifically, due to employing novel perceptual quantization matrices, *the video transmission bit-rate is reduced up to about 25%* in terms of SSIMPlus objective quality metric, while keeping the visual quality substantially at the same level.

REFERENCES

1. "Cisco Visual Networking Index: Forecast and Methodology, 2016–2021", Online: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html, Cisco Systems Inc., Feb. 2019.

2. "Parameter values for the HDTV standards for production and international programme exchange" BT Series, Broadcasting service (television), Recommendation ITU-R BT.709-6, 06/2015.

3. "Parameter values for ultra-high definition television systems for production and international programme exchange" BT Series, Broadcasting service (television), Recommendation ITU-R BT.2020-2, 10/2015.

4. 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*, vol. 22, no. 12, pp. 1649-1668, 2012.

5. ITU-T, Recommendation H.265 (04/13), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding.

6. ITU-T, Recommendation H.265 (10/14), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding.

7. ITU-T, Recommendation H.265 (04/15), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding.

8. ITU-T, Recommendation H.265 (12/16), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding.

9. ITU-T, Recommendation H.265 (02/18), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding.

10. D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," *Picture Coding Symposium (PCS) 2013*, pp.394-397, 8-11 Dec. 2013.

11. D. Grois, D. Marpe, T. Nguyen, and O. Hadar, "Comparative Assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders for Low-Delay Video Applications", Proc. SPIE 9217, *Applications of Digital Image Processing XXXVII,* 92170Q, Sept. 2014.

12. D. Grois, T. Nguyen, and D. Marpe, "Coding Efficiency Comparison of AV1/VP9, H.265/MPEG-HEVC, and H.264/MPEG-AVC Encoders," *Picture Coding Symposium (PCS)*, Dec. 2016.

13. D. Grois, T. Nguyen, and D. Marpe,"Performance Comparison of AV1, JEM, VP9, and HEVC Encoders", Proc. SPIE 10396, *Applications of Digital Image Processing XL*, 103960L, 7-10 Aug., 2017.

14. T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

15. M. Nezamabadi, S. Miller, S. Daly, and R. Atkins, "Color signal encoding for high dynamic range and wide color gamut based on human perception," *Proceedings of the SPIE*, 9015:90150C–90150C–12, 2014.

16. Society of Motion Pictures and Television Engineers (SMPTE), "SMPTE ST 2084:2014 – High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays," Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, Aug. 2014.

17. Society of Motion Pictures and Television Engineers (SMPTE), "Study group report: High-dynamic-range (HDR) imaging ecosystem," Technical report, Society of Motion Pictures and Television Engineers (SMPTE), Sep. 2015.

18. A. Ninan and S. Daly, "Evolution of displays through advances in next generation video," In Proceedings of the International Display Workshops (IDW), page 1312, Toki Messe Niigata Convention Center Niigata Japan, 2014.

19. P. G. J. Barten. Physical model for the contrast sensitivity of the human eye. Proceedings of the SPIE, 1666:57–72, 1992.

20. P. G. J. Barten. “Contrast sensitivity of the human eye and its effects on image quality,” volume 72, *SPIE press*, USA, Dec. 1999.

21. P. G. J. Barten, “Formula for the contrast sensitivity of the human eye,” *Proceedings of the SPIE*, 5294:231–238, 2003.

22. X. Shang, G. Wang, X. Zhao, Y. Zuo, J. Liang and I. V. Bajić, "Weighting Quantization Matrices for HEVC/H.265-Coded RGB Videos," in *IEEE Access*, vol. 7, pp. 36019-36032, 2019.

23. L. Prangnell and V. Sanchez, "Adaptive Quantization Matrices for HD and UHD Resolutions in Scalable HEVC," *2016 Data Compression Conference (DCC)*, Snowbird, UT, 2016, pp. 626-626.

24. L. Ding, G. Li, R. Wang, W. Wang, "Video Pre-processing with JND-Based Gaussian Filtering of Super-pixels", *Visual Information Processing and Communication VI*, vol. 9410, pp. 941 004, 2015.

25. D. Yuan, T. Zhao, Y. Xu, H. Xue and L. Lin, "Visual JND: A Perceptual Measurement in Video Coding," in *IEEE Access*, vol. 7, pp. 29014-29022, 2019.

26. J. Cui, R. Xiong, X. Zhang, S. Wang and S. Ma, "Perceptual Video Coding Based on Visual Saliency Modulated Just Noticeable Distortion," *2019 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2019, pp. 565-565.

27. Z. Zeng, H. Zeng, J. Chen, J. Zhu, Y. Zhang and K. Ma, "Visual Attention Guided Pixel-Wise Just Noticeable Difference Model," in *IEEE Access*, vol. 7, pp. 132111-132119, 2019.

28. P. Lee, T. Bui and S. Yao, "Improve the HEVC Algorithm Complexity Based on the Visual Perception," *2019 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2019, pp. 1-4.

29. D. Fu, Y. Wang, B. Fan and N. Ding, "HEVC/H.265 Intra Coding Based on the Human Visual System," in *IEEE Access*.

30. D. Grois, and A. Giladi, "Perceptual quantization matrices for high dynamic range H.265/MPEG-HEVC video coding," Proc. SPIE 11137, *Applications of Digital Image Processing XLII*, 111370O, 2019.

31. Eurofins®, “Which is the best objective video quality measure and why use SSIMPLUS”, Eurofins®, Online: https://cdnmedia.eurofins.com/digitaltesting/media/116613/qoe-why-ssimplus.pdf .

32. SSIMWAVE®, “SSIMPLUS Outperforms VMAF”, 2017.

33. G. Bjøntegaard, “Calculation of average PSNR differences between RD-curves”, ITU-T Q.6/SG16 VCEG 13th Meeting, Document VCEG-M33, Austin, USA, Apr. 2001.

34. G. M. Johnson and M. D. Fairchild, “On Contrast Sensitivity in an Image Difference Model”, *In Proceedings of the IS&T PICS Conference*, pp. 18-23, Portland, OR, 2002.

35. B.A. Wandell, “Foundation of Vision,” Sinear Associates, Sunderland, MA, 1995.

36. S. Daly, “Subroutine for the generation of a two dimensional human visual contrast sensitivity function,” Technical Report 233203Y, Eastman Kodak, 1987.

37. S. Daly, “The visible differences predictor: An algorithm for the assessment of image fidelity,” In A. B. Watson, editor, *Digital Images and Human Vision*, pp. 179–206, 1993.

38. S. Daly, “A visual model for optimizing the design of image processing algorithms,” *International Conference on Image Processing (ICIP)*, vol. 2, pp. 16–20, Nov.1994.

39. S. Daly, T. Kunkel, X. Sun, S. Farrell, and P. Crum, “Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance,” *Proceedings of the SPIE*, 8651:86510J–86510J–11, 2013.

40. J. L. Mannos, and D. J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," IEEE Trans. on Info. Theory, Vol. IT-20, No. 4, July 1974.

41. R. Rosenholtz and A. B. Watson, “Perceptual adaptive JPEG coding,” In Proceedings of the IEEE International Conference on Image Processing (ICIP), vol. 1, pp. 901–904, Sep. 1996.

42. L.W. Chang, C.Y. Wang and S.M. Lee, "Designing JPEG quantization tables based on human visual system," *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, Kobe, 1999, pp. 376-380 vol.2.

43. J. R. Sullivan, L. A. Ray and R. Miller, "Design of minimum visual modulation halftone patterns," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 1, pp. 33-38, Jan.-Feb. 1991.

44. K. N. Ngan, K. S. Leong and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 11, pp. 1743-1750, Nov. 1989.

45. C. Furmanski, and S. Engel, “An oblique effect in human primary visual cortex,” *Nat Neurosci* 3, 535–536 (2000).

46. Y.L.Grand, “Light Colour and Vision,” *2-nd edition*, Chapman and Hall, 1969.

47. Projects from VideoLAN, x265 software library and application, Online: https://www.videolan.org/developers/x265.html.

48. Projects from VideoLAN, x264 software library and application, Online: https://www.videolan.org/developers/x264.html

49. F. Bossen, “Common HM test conditions and software reference configurations,” document JCTVC-L1100 of JCT-VC, Geneva, CH, Jan. 2013.