An Adaptive Perceptual Quantization Algorithm Based on Block-Level JND for Video Coding

Guoqing Xiang[1,2], Xiaodong Xie[2,*], Huizhu Jia[2], Xiaofeng Huang[2], Jie Liu[2], Kaijin Wei[2], Yuanchao Bai[2], Pei Liao[1,2], and Wen Gao[2]

[1] SECE of Shenzhen Graduate School, Peking University, 518055, Shenzhen, China
[2] Engineering Lab for Video Technology School of EECS, Peking University, Beijing, China
{gqxiang,hzjia,xfhuang,liuzimin,xdxie}@jdl.ac.cn

Abstract. It has been widely demonstrated that integrating efficient perceptual measures into traditional video coding framework can improve subjective coding performance significantly. In this paper, we propose a novel block-level JND (just-noticeable-distortion) model, which has not only adjusted pixel-level JND thresholds with more block characteristics, but also integrated them into a block-level model. And the model has been applied for adaptive perceptual quantization for video coding. Experimental results show that our model can save bit rates up to 24.5% on average with negligible degradation of the perceptual quality.

Keywords: JND model, masking effect, adaptive perceptual quantization, visual quality, video coding performance.

1 Introduction

Traditional hybrid video coding aims to remove spatial and temporal statistical redundancies for signal compression. However, most of these methods often neglect perceptual features for better subjective video coding. Considering that human eyes are the ultimate receivers, it is worthwhile to dedicate perceptually friendly coding researches to further remove perceptual redundancies and improve subjective quality. The just noticeable distortion (JND) threshold, i.e. the distortion that observers just begin to notice, is one of the popular perceptual methods used for such applications.

There have been abundant research efforts to develop rational JND models and apply them into video coding. The existing JND models can be classified into pixel domain and transform domain, respectively. In the pixel domain, most JND models use luminance adaptation and texture masking to compute pixel-level JND [1]. X.K.Yang et al. extended the JND model with a Nonlinear Additively Masking Model (NAMM) by integrating the luminance and texture masking together [2]. Except the luminance and texture that affect the perceived distortion, some other important factors have been studied, such as Chen et al. introduced a famous FJND model [3] integrated with retina foveation model to account for the relationship between visibility and eccentricity. Most of the transform domain JND methods are modeling in DCT domain, and the

* Corresponding author.

W.T. Ooi et al. (Eds.): PCM 2014, LNCS 8879, pp. 54–63, 2014.

subband JND features higher accuracy with the consideration of channel's interactions. Among the popular DCT domain JND models, researchers focus on luminance adaptation, the spatial CSF (contrast sensitivity function) and temporal CSF effects [4] to get useful spatial-temporal JND models. However, Jia's model [4] only refers the magnitude of motion contribution to final spatial-temporal JND threshold, but the directionality of motion is neglected. Wei et al. furthered the model by introducing a gamma correction to compensate the original luminance adaptation effect and a novel temporal modulation factor to integrate temporal properties [5].

The existing JND methods have been widely used in hybrid video coding to enhance coding performance. In [7], a low-complexity perceptual rate distortion model has been introduced to replace the Lagrange RD cost model and demonstrated inter mode decision performance improvement. Chen et al. [3] optimized the quantization parameter for each MB (macro-block) based on its FJND information. The Lagrange multiplier in the rate-distortion optimization is adapted to ensure that the MB noticeable distortion is minimized. Some researchers have utilized JND models to improve the compression rates mainly by residues [6] [8] or DCT coefficients [9-10] filtering based on a hard or soft JND threshold.

From the JND models and application methods in video coding mentioned before, we can find that most the popular JND models have some characteristics as follows. First, although these models have considered adjacent characteristics of a pixel or a subband during the modeling procedure, they are just applied by separate pixel-level filtering [6, 8, 9, 10], and these methods do not consider that traditional video coding is on the basis of block units, which prefers more on smooth compression. And pixel-level filtering may introduce much artifact distortion fluctuation in a block. Second, the computational complexity is too large for most JND models, such as using the canny operator in [5]. At last, none of them pay attention to the fact that the quantization parameter decided by traditional mode decision has not considered perceptual properties, namely, the *QP* can be perceptually adjusted to remove more visual redundancies or enhance perceptual quality. In order to resolve these problems, an adaptive perceptually quantization method for video coding based on a block-level JND model is proposed in this paper. It not only integrates the pixel-level JND into the block-level JND, but also uses perceptual quantization to improve compression efficiency.

The reminder of this paper is organized as follows. In Section 2, the main structure of the proposed block-level JND model is introduced. Based on our block-level JND model, an adaptive perceptual quantization method is integrated into video coding framework in Section 3. In Section 4, the experimental results are shown and discussed. The Section 5 draws the conclusions of our work.

2 Proposed Block-Level JND Model

The proposed block-level JND model is computed by two steps, computing pixel-level JND and then integrating it into block-level JND. The pixel-level JND in [5] can be expressed for every 4×4 block in an image as the integration of a spatial JND value JND_s and a temporal modulation factor F_T,

$$JND_T(n,i,j) = JND_S(n,i,j)\cdot F_T(n,i,j) \tag{1}$$

where n is the index of a 4×4 block in the image, i and j are the DCT coefficients indices. Then the block-level JND can be computed as,

$$JND_{blcok}(k) = \alpha \cdot lnD_{block}(k) \tag{2}$$

where k denotes the kth 8×8 block in a macro-block, JND_{block} is the JND threshold of a 8×8 block, α is an empirical control parameter. D_{block} means the HVS perceptual sensitivity of a block computed by the integration of its pixel-level JND_T and the block energy, which will be detailed later.

2.1 The Pixel-Level JND Threshold

The spatial JND threshold is the product of luminance adaptation factor, the contrast masking factor and the frequency property of DCT sub-band, it can be calculated as,

$$JND_S(n,i,j) = T_{basic}(n,i,j) \cdot F_{lum}(n) \cdot F_{contrast}(n,i,j) \tag{3}$$

where T_{basic} is the base JND threshold generated by the CSF effect, F_{lum} is the luminance adaptation and $F_{contrast}$ denotes the contrast masking effect.

First, the basic threshold is considered. The HVS is sensitive to spatial frequencies, and the spatial frequency of the (i, j)th subband in the nth DCT block is related to block dimension N, which can be computed as in [5],

$$\omega_{i,j} = \frac{1}{2N}\sqrt{\left(i/\theta_x\right)^2 + \left(j/\theta_y\right)^2} \tag{4}$$

where θ_x , θ_y denotes the horizontal and vertical visual angle respectively, and they are the same as,

$$\theta_x = \theta_y = 2 \cdot \arctan\left(1/\left(2 \cdot R_{vd} \cdot P_{ch}\right)\right) \tag{5}$$

where R_{vd} stands for the ratio of viewing distance [5] to picture height P_{ch}. Then the basic threshold for DCT subband can be calculated as,

$$T_1(n,i,j) = s \cdot \frac{1}{\phi_i \phi_j} \cdot \frac{\exp\left(c\omega_{ij}\right)/\left(a + b\omega_{ij}\right)}{r + (1-r)\cdot \cos^2 \varphi_{ij}} \tag{6}$$

where s accounts for the spatial summation effect with an empirical value 0.25, parameter r is set to 0.6, and the normalization factors ϕ_i or ϕ_j are expressed as

$$\phi_m = \begin{cases} \sqrt{1/N}, m = 0 \\ \sqrt{2/N}, m > 0 \end{cases} \tag{7}$$

and more parameters can be found in [5]. And because of the difference between 4×4 block and 8×8 block, for ϕ_m in (7), the $\phi_i\phi_j$ will make the basic perceptual distortion fail by half, so we set the basic threshold of a 4×4 block as follows to ensure perceptual distortion consistency with [5]

$$T_2 = 2 \cdot T_1 \tag{8}$$

In order to take more block-level luminance characteristics into account [12], we adjust each basic threshold as

$$T_{basic} = T_2 \cdot \left(\frac{C(0,0,n)_{4\times4}}{C(0,0)_{8\times8}} \right)^{\tau} \tag{9}$$

where $C(0,0,n)_{4\times4}$ and $C(0,0)_{8\times8}$ is the nth 4×4 block DC coefficient in a 8×8 block and the 8×8 block DC coefficient respectively, the parameter τ here equals to 0.649.

The luminance adaptation for HVS mainly depends on average luminance intensity value $\bar{I}$ of each 4×4 block, and it is described as in [5]

$$F_{lum}(n) = \begin{cases} \left(60-\bar{I}\right)/150+1, & \bar{I} \le 60 \\ 1, & 60 < \bar{I} < 170 \\ \left(\bar{I}-170\right)/425+1, & \text{others} \end{cases} \tag{10}$$

It is easy to know that in smooth and edge areas the distortion can be more easily recognized than texture areas with high energy, so we should compute the contrast masking factor according to different type of blocks, namely Plane, Edge and Texture. However, the Canny operator may be too complicated in [5] for block classification, here we replaced it with a famous and useful block classification method in DCT domain [13] used for JPEG encoder. According to [13], the 4×4 block can be divided into four indicative areas as show in Fig.1, where *DC*, *L*, *M*, and *H* denotes the absolute sums of DCT coefficients in different areas respectively. By easily computing the (*L*+*M*)/H, *L/M*, *M*+*H* and comparing them with experimental threshold μ_1, μ_2, α_1, α_2, β_1, β_2, γ etc., we can decide the block type as Plane, Edge and Texture quickly. Generally, the larger values of *L/M* and (*L*+*M*)/*H* for a block means higher possibility to be Edge area, and usually smaller *M*+*H* indicates Plane block. More details and the comparison procedure can be found in [13].

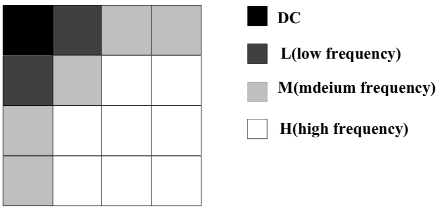


Fig. 1. Classification indicators for a 4×4 DCT block

Therefore through employing the elevation factor in [5],

$$\psi(n,i,j) = \begin{cases} 1, & \text{for Plane and Edge block} \\ 2.25 & \text{for } \left(i^2+j^2\right) \le 4 \text{ in Texture block} \\ 1.25 & \text{for } \left(i^2+j^2\right) > 4 \text{ in Texture block} \end{cases} \tag{11}$$

we can compute $F_{contrast}$ as follows,

$$F_{contrast}=\begin{cases}\psi, & \text{for } \left(\mathrm{i}^2+\mathrm{j}^2\right)\le 4 \text{ in Plane and Edge block}\\ \psi\cdot\min\left(4,\max\left(1,\left(\dfrac{C(n,i,j)}{T_{basic}(n,i,j)\cdot F_{1um}(n)}\right)\right)^{0.36}\right), & \text{others}\end{cases}\tag{12}$$

where $C(n,i,j)$ denotes the (i,j)th DCT coefficient in the nth block.

In order to consider the temporal effect, we need to evaluate the temporal modulation factor F_T. Many works have demonstrated that there are no independent characteristics between the spatial and temporal frequencies, so we should take spatial factor into account for the temporal modulation factor. In [5], F_T is derived as

$$F_T=\begin{cases}1 & f_s<5cpd\ \&\ f_t<10Hz\\ 1.07^{(f_t-10)} & f_s<5cpd\ \&\ f_t\ge 10Hz\\ 1.07^{f_t} & f_s\ge 5cpd\end{cases}\tag{13}$$

where f_s accounts for spatial frequency discussed above and f_t denotes the temporal frequency, which have relationships with motion vectors acquired by motion estimation, current frame rate, and eyes move velocity, etc. Detailed calculation procedure has been described in [5].

Finally, the pixel-level JND threshold in classic 4×4 DCT domain can be obtained as (1), which is the basis of block-level JND model.

2.2 The Block-Level JND Threshold

The block-level JND threshold is proposed based on the following facts. Firstly, observers are more easily attracted to a block or an area than a pixel in an image. Among most natural scenes, the distortion beyond a block-level perceptual distortion threshold in a block will be more easily noticed than a pixel-level difference out of its threshold. Secondly, observers can be more easily attracted to high frequency content and is more sensitive to the distortion of low frequency areas, such as edges and noise in plane area, respectively. And human eyes have less interest in the medium frequency areas which contain much information and energy, and become less sensitive to their distortion. Therefore, it is rational to take subband pixel-level JND of a block and its energy distribution characteristics together into account to find a reasonable JND threshold for each block in a picture.

Similar to [11], a block-level JND for image is proposed, it has considered the energy distribution characteristics in a block and the difference of block types in JND modeling, and it can be expressed as,

$$D_{block}(k)=\sum_{n=0}^{M}\sum_{i=0}^{N}\sum_{j=0}^{N} JND_T(n,i,j)\cdot|C(n,i,j)|^2\tag{14}$$

where M is the number of sub-block divided in a block, here its value is 4. The larger the D_{block}, the less sensitivity to distortion of the block for HVS, i.e. the more redundancies can be removed for better compression without much visual difference. And from the expression of D_{block}, we can find that in the very low and very high frequency sub-bands, such as very simple plane area with near-zero JND_T and very complex edge areas with very low energy respectively, their values tend to be smaller than in medium frequency sub-bands with larger JND_T and energy such as modest texture areas. As a consequence, there will be less artifact distortion fluctuation in a block than pixel-level JND and we can avoid introducing too much artifact distortion in low frequency areas and protect more details in high frequency areas. At the end, the block-level JND can be computed as (2), which will be incorporated in perceptual video coding.

3 Adaptive Perceptual Quantization for Video Coding

In the traditional hybrid video coding standards, an offset of quantization parameter QP, namely ΔQP, will be used in Differential Pulse Code Modulation (DPCM) and transmitted in coded bit stream, and the QP used for residual DCT coefficient quantization or inverse quantization can be expressed as,

$$QP = QP_0 + \Delta QP \tag{15}$$

where QP_0 is original quantization parameter of current macro-block, and it will be used for uniform quantization in a macro-block. However, the QP used in the best mode coding does not explore perceptual characteristics very well. According to [14], the quantization error should be limited to

$$|e_{QP}| = |C - C_{rec}| \leq JND_{block} \tag{16}$$

where C_{rec} stands for the reconstructed DCT coefficient. Taking the maximum unnoticeable distortion into account, a perceptual quantization step should be limited to the block-level noticeable distortion, so we can get

$$QP_{step} = 2 \cdot JND_{block} \tag{17}$$

where QP_{step} is a uniform quantizer step applied to each DCT coefficient C. Then we can combine the procedure in [14] with the proposed block-level JND model in Section 2 and get the perceptual ΔQP_{JND} as,

$$\Delta QP_{JND} = \mathrm{Ceil}(K \cdot \log_2 JND_{block}) \tag{18}$$

where $Ceil(x)$ denotes the closest integer not more than the argument, K means the relationship between QP and QP_{step} and it varies from different video coding standard. At the end, the perceptually adaptive quantization parameter QP_{JND} is computed as

$$QP_{JND} = QP_0 + \Delta QP_{JND} \tag{19}$$

In order to comply with traditional macro-block video coding standard, we should average all the mentioned QP_{JND} of each 8×8 block in a macro-block for uniform quantization as,

$$QP_{JND\_MB} = \frac{1}{B}\sum_{k=0}^{B} QP_{JND}(k) \tag{20}$$

where B is the number of blocks in a coding macro-block. The QP_{JND_MB} will be used for the best mode coding to remove more perceptual unnoticeable redundancies and integrate it into video coding procedure as shown in Fig. 2. The quantization offset ΔQP_{JND_MB} values' mapping will be transmitted to final bit stream as follows

$$\Delta QP_{JND\_MB} = QP_{JND\_MB} - QP_0 \tag{21}$$

4 Experimental Results

In order to evaluate the performance of our proposed block-level JND scheme, the integration procedure is implemented on AVS Jizhun profile. The GOP length is 15 with structure IBBPBBP….The frame rate is 30 frames per second, the motion estimation is carried out at a quarter pixel resolution with search range of 16 and the RDO is enabled. The test sequences are 4:2:0 YUV format covering CIF, 720p and 1080p resolutions. We have compared the subjective quality and bit rate compression performance of video encoded by the proposed method with Yang's model [2], which is a famous and useful JND model [10]. The chosen subjective distortion measure is the Multiple Scale-Structural Similarity (MS-SSIM) [15] calculated on the luminance frames and averaged for the whole sequence. The Table 1 shows our experimental results and the Fig.3 (a) to Fig.3 (b) demonstrate subjective performance improvement directly. And we can make some discussions as follows.

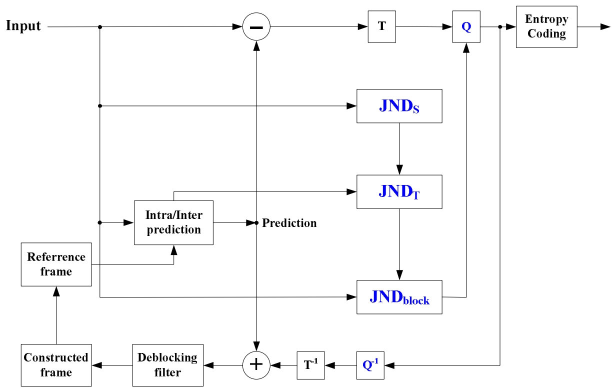
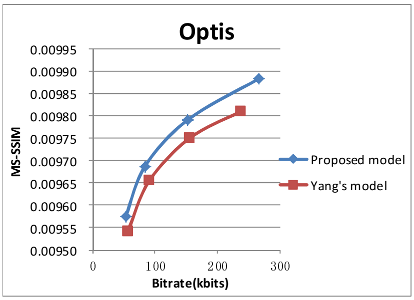


Fig. 2. Perceptual adaptive quantization video coding diagram

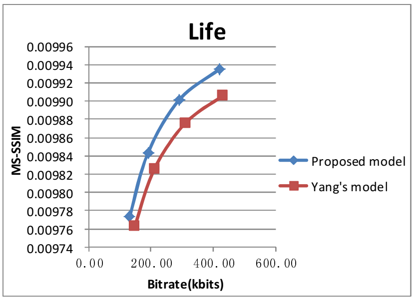
According to the Table 1, we can conclude that our block-level JND model shows less negligible subjective loss and lower MS-SSIM reduction than Yang's with the value of -0.3265% and -0.5654%, respectively. It means that the perceptually processed sequences almost have the same visual quality as original reference encoded sequences and are better than Yang's. We can see that the proposed block-level JND model yields an average 24.5% bit rate savings compared to Yang's 18.8% bit rate reduction. Meanwhile, Yang's PSNR loss is almost twice than our model, which means that because of considering block-level JND characteristics rather than directly using separate JND filtering, our model can shape perceptual noise more uniformly. The Fig.3 (a) and (b) show the integral performance by bit rate versus MS-SSIM exampled by sequence "Optics" and sequence "Life". We can easily find that our model has a better subjective performance, which is in accordance with the result of Table 1.

Table 1. Performance of proposed block-level JND model

Sequences	QP	Proposed model			Yang's model		
		ΔBitrate (%)	ΔPSNR (dB)	ΔMS-SSIM (%)	ΔBitrate (%)	ΔPSNR (dB)	ΔMS-SSIM (%)
Football (CIF)	16	-12.07%	-1.18	-0.1347%	-5.00%	-2.37	-0.2565%
	20	-13.55%	-1.32	-0.2181%	-3.97%	-1.68	-0.2490%
	24	-13.38%	-1.24	-0.3127%	-2.83%	-1.01	-0.2231%
	28	-12.59%	-1.08	-0.3835%	-1.96%	-0.58	-0.1832%
Foreman (CIF)	16	-33.44%	-2.16	-0.2154%	-17.08%	-3.20	-0.2388%
	20	-35.32%	-2.12	-0.2565%	-14.89%	-2.27	-0.2046%
	24	-32.60%	-1.72	-0.2912%	-11.70%	-1.35	-0.1624%
	28	-25.20%	-1.19	-0.2659%	-7.51%	-0.71	-0.1153%
Optis (720p)	16	-24.29%	-1.89	-0.4112%	-32.73%	-5.01	-1.1559%
	20	-36.70%	-2.34	-0.8442%	-35.60%	-3.98	-1.2634%
	24	-35.31%	-1.66	-0.7978%	-29.96%	-2.48	-1.1315%
	28	-26.24%	-1.15	-0.7721%	-20.96%	-1.57	-1.1011%
Sheriff (720p)	16	-22.89%	-1.68	-0.2139%	-33.53%	-5.99	-0.9714%
	20	-30.63%	-2.06	-0.4333%	-35.97%	-4.77	-0.9858%
	24	-33.69%	-1.85	-0.5478%	-34.16%	-3.25	-0.8771%
	28	-28.64%	-1.30	-0.4905%	-26.16%	-1.96	-0.7071%
Life (1080p)	16	-24.42%	-1.43	-0.0725%	-22.62%	-5.72	-0.3959%
	20	-27.86%	-1.54	-0.1194%	-22.86%	-4.50	-0.3994%
	24	-29.10%	-1.35	-0.1417%	-21.28%	-3.07	-0.3692%
	28	-25.95%	-1.00	-0.1609%	-16.92%	-1.93	-0.3128%
Tennis (1080p)	16	-14.94%	-1.02	-0.1425%	-16.07%	-4.22	-0.6586%
	20	-21.75%	-1.06	-0.3198%	-18.15%	-3.22	-0.7470%
	24	-15.17%	-0.58	-0.1433%	-11.44%	-2.05	-0.4653%
	28	-12.26%	-0.51	-0.1470%	-8.85%	-1.53	-0.3960%
Average		-24.50%	-1.43	-0.3265%	-18.84%	-2.85	-0.5654%



(a)



(b)

Fig. 3. Subjective performance comparison (a) sequence "Optics" (b) sequence "Life"

5 Conclusion

In this paper, we have proposed a novel block-level JND model and integrated it with video coding scheme to generate the perceptually adaptive quantization. The proposed algorithm is fully compatible with current mainstream video coding standard and also can be applied for H.264/AVC and HEVC coding framework. The experimental results show average 24.5% bit rate saving with negligible impact on the sequence perceptual quality. In the future, we plan to explore more accurate JND model and integrate its applications into hybrid video coding to further enhance the coding performance.

Acknowledgements. This work is partially supported by grants from the Chinese National Natural Science Foundation under contract No.61171139 and National High Technology Research and Development Program of China (863 Program) under contract No.2012AA011703.

References

1. Chou, C.H., Li, Y.C.: A perceptually tuned subband image coder based on the measure of Just-Noticeable-Distortion Profile. IEEE Transaction on Circuits and Systems for Video Technology 5(6), 467–476 (1995)
2. Yang, X.K., Lin, W.S., Lu, Z., Ong, E.P., Yao, S.S.: Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 609–612 (2003)
3. Zhenzhong, C., Guillemot, C.: Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model. IEEE Transactions on Circuits and Systems for Video Technology 20(6), 806–819 (2010)
4. Yuting, J., Weisi, L., Kassim, A.A.: Estimating Just-Noticeable Distortion for Video. IEEE Transactions on Circuits and Systems for Video Technology 16(7), 820–829 (2006)
5. Zhenyu, W., Ngan, K.N.: Spatio-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain. IEEE Transactions on Circuits and Systems for Video Technology 19(3), 337–346 (2009)
6. Hao, C., et al.: Temporal color Just Noticeable Distortion model and its application for video coding. In: IEEE International Conference on Multimedia and Expo (ICME), Suntec City (2010)
7. Huan, W., Xueming, Q., Guizhong, L.: Inter mode decision based on Just Noticeable Difference profile. In: 2010 17th IEEE International Conference on Image Processing (ICIP), Hong Kong (2010)
8. Chun-Man, M., King, N.N.: Enhancing compression rate by just-noticeable distortion model for H.264/AVC. In: IEEE International Symposium on Circuits and Systems, ISCAS 2009, Taipei (2009)
9. Luo, Z., et al.: H.264/Advanced Video Control Perceptual Optimization Coding Based on JND-Directed Coefficient Suppression. IEEE Transactions on Circuits and Systems for Video Technology 23(6), 935–948 (2013)
10. Qi, C., Li, S.: AVS encoding optimization with perceptual just noticeable distortion model. In: 2013 9th International Conference on Information, Communications and Signal Processing (ICICS), Tainan (2013)

11. Wilson, T.A., Rogers, S.K., Myers Jr., L.R.: Perceptual-based hyper spectral image fusion using multiresolution analysis. Optical Engineering 34(11), 3154–3164 (1995)
12. Watson, A.B.: Visually optimal DCT quantization matrices for individual images. In: Data Compression Conference, DCC 19, UT, Snowbird (1993)
13. Tong, H.H.Y., Venetsanopoulos, A.N.: A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking. In: Proceedings of the 1998 International Conference on Image Processing, ICIP 1998, Chicago, IL (1998)
14. Naccari, M., Mrak, M.: Intensity dependent spatial quantization with application in HEVC. In: 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA (2013)
15. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. IEEE Trans. on Image Proc. 19(6), 1427–1441 (2010)