Adaptive Block-Size Transform Based Just-Noticeable Difference Profile for Images

Lin Ma and King N. Ngan

Department of Electronic Engineering, The Chinese University of Hong Kong,
Hong Kong SAR
{lma,knngan}@ee.cuhk.edu.hk

Abstract. In this paper, we propose a novel adaptive block-size transform (ABT) based just-noticeable difference (JND) model for images. As different transform sizes have different properties on energy compaction and detailed information preservation, traditional 8×8 discrete cosine transform (DCT) based JND model is firstly extended to 16×16 DCT based JND model, by considering spatial contrast sensitivity function (CSF), the luminance adaptation effect and the contrast masking effect based on block classification. Furthermore, in order to obtain a more accurate JND profile, a novel strategy is proposed for deciding which size of transform is employed to generate the resulting JND model. And experimental results have demonstrated that the proposed model is consistent with human visual system (HVS). Compared with other JND profiles, our proposed model could tolerate more distortions and have much better perceptual quality.

Keywords: Just-noticeable difference (JND), Adaptive block-size transform (ABT), Human visual system (HVS).

1 Introduction

JND accounts for the smallest detectable difference between a staring and secondary level of a particular sensory stimulus in psychophysics [1], which is also known as the difference limen or differential threshold. And JND model has prevailed and given a promising way to model the property of HVS accurately and efficiently in many image processing research fields, such as perceptual image compression [2]-[4], image perceptual quality evaluation [5] [6], watermarking [7], etc.

Generally, automatic JND model can be determined in the image domain, or transform domain, such as DCT and discrete wavelet transform (DWT), or even the combination of the previous two schemes. JND model generated in image domain [8] [9], which is also named as pixel-based JND, mainly focus on the background luminance adaptation and the spatial contrast masking. In [10], Yang et al. deduces the overlapping effect of luminance adaptation and spatial contrast masking to refine the JND profile in [8]. However, pixel-based JND model has not considered the sensitivity of human vision for different frequency components. Therefore it could not describe the HVS property accurately. JND model generated in the transform domain named the subband-based JND usually incorporate all the major affecting factors, namely, CSF,

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 1210–1220, 2009.

luminance adaptation, and contrast masking. In [2], the DCT JND thresholds are developed based on spatial CSF. Then the basic JND model is improved by Watson [3] in DCTune model after the implementation of contrast masking. Furthermore, Hontsch et al. [4] propose to modify the DCTune model by replacing a single pixel with a foveal region. And Zhang et al. [11] refine the JND model by formulating the luminance adaptation adjustment and contrast masking. More recently, Wei et al. [12] incorporate new formulae of luminance adaptation, contrast masking and Gamma correction to estimate the JND threshold in DCT domain. Also Zhang et al. [13] propose to estimate the JND profile by summing the effects in DCT and pixel domain together. However all the existed DCT-based JND profiles are calculated based on the 8×8 DCT, which could not explicitly exploit the properties of HVS on different transform sizes.

Recently, ABT has attracted researchers' attention for its coding efficiency in image and video compression [14] [15]. Actually, different transform sizes have different properties on energy compaction and detailed information preservation. Specifically, larger transforms could provide better energy compaction and better preservation of details, while smaller transforms could prevent more ringing artifacts during compression. Therefore, we believe that HVS performs differently with different transform sizes. By incorporating ABT together with the DCT-based JND profile, more accurate JND profile for image could be obtained.

In this paper, inspired by recent progresses on image/video compression based on ABT, a novel ABT-based JND profile for images is proposed. Firstly, traditional 8×8 DCT-based JND model is extended to 16×16 DCT-based JND model. During the extension, a psychophysical experiment is carried out to parameterize the proposed spatial CSF for further consisting with HVS. Secondly, a new strategy for adaptively adjusting the content of image is proposed to decide which transform size is employed to yield the resulting JND profile.

The rest of the paper is organized as follows. In Section 2, the proposed ABT-based JND profile for images is introduced. And experimental results are demonstrated in Section 3. Finally, Section 4 concludes the paper

2 The Proposed ABT-Based JND Model

JND profile in the DCT domain could be determined by a basic visibility threshold T_{basic} generated from the spatial CSF, the luminance adaptation α_{lum} and the contrast masking α_{cm} [11]-[13]:

$$T(m,n,i,j) = T_{basic}(i,j) \cdot \alpha_{lum}(m,n) \cdot \alpha_{cm}(m,n,i,j) , \quad (1)$$

where (m,n) denotes the position of DCT block in an image, (i,j) indicates the DCT subband, and T is the final obtained JND threshold.

2.1 DCT-Based JND Basic Threshold

Based on the band-pass property of human eyes in the spatial frequency domain, Ngan et al. [16] modeled the sensitivity characteristic of HVS as:

$$H(\varpi) = (a + b\varpi) \cdot \exp(-c\varpi), \tag{2}$$

where ϖ is the specified spatial frequency. As JND threshold is the inverse of the sensitivity modeled by (2), we can model the basic JND threshold [12] according to:

$$T_{basic}(m,n,i,j) = \frac{s}{\phi_i \phi_j} \cdot \frac{\exp(c\varpi_{i,j})/(a + b\varpi_{i,j})}{\gamma + (1-\gamma) \cdot \cos^2 \varphi_{ij}}, \tag{3}$$

where $s = 0.25$ denotes the summation effect factor, and ϕ_i, ϕ_j are the DCT normalization factors:

$$\phi_m = \begin{cases} \sqrt{1/N}, & m = 0 \\ \sqrt{2/N}, & m > 0 \end{cases}, \tag{4}$$

N stands for the dimension of the DCT block. And φ_{ij} denotes the directional angle of the corresponding DCT coefficient:

$$\varphi_{ij} = \arcsin(2\varpi_{i,0}\varpi_{0,j} / \varpi_{i,j}^2), \tag{5}$$

where $\varpi_{i,j}$ is the corresponding spatial frequency of DCT subband (i, j):

$$\varpi_{i,j} = (1/(2N))\sqrt{(i/\theta_x)^2 + (j/\theta_y)^2}, \tag{6}$$

where θ_x and θ_y are the horizontal and vertical visual angles, respectively. According to the international standard ITU-R BT.500-11, the ratio of viewing distance to picture height should be fixed which is always from 3 to 6 related to the picture height. Therefore θ_x and θ_y are identical which is calculated by:

$$\theta_x = \theta_y = 2 \cdot \arctan(1/(2 \cdot R_d \cdot P_h)), \tag{7}$$

where R_d indicates the ratio of viewing distance to image height, and P_h is the image height (in pixel).

2.2 8×8 JND Profile Extending to 16×16 JND Profile

In order to extend the traditional 8×8 JND profile to 16×16 one, the dimension of DCT block N is set as 16 to obtain the basic JND threshold T_{basic}. Furthermore, a psychophysical experiment is carried out to parameterize the three parameters a, b, and c in Equation (3).

The psychophysical experiment is designed as follows. Given a 512×512 image with pixel intensities all set to 128, noise is added to certain spatial frequency component of 16×16 DCT individually. For each test DCT subband, five amplitudes of noise

are chosen based on the pre-designed measurements by the subjective sensitivity. A group of viewers give their opinions on whether the noise could be perceived. If half of the persons vote "visible", we regard that the noise is above the JND threshold. Otherwise, the noise is below the JND threshold. Finally, we could obtain the JND threshold for the selected DCT subband. Subsequently, least mean squared error is employed to fit the obtained JND threshold to Equation (3) by:

$$(a, b, c) = \arg\min \sum_{\varpi_{ij}} [T_{\varpi_{ij}} - T_{basic}(m, n, i, j)]^2 , \tag{8}$$

where $T_{\varpi_{ij}}$ is the JND threshold obtained from the psychophysical experiment. As a result, $a = 1.83$, $b = 0.165$, and $c = 0.16$ for 16×16 JND profile, while $a = 1.33$, $b = 0.11$, and $c = 0.18$ for 8×8 JND profile [12].

As JND threshold is influenced by the intensity scale of digital image, which means that higher visibility threshold occurs in either dark or bright regions compared with medium brightness regions, luminance adaptation factor α_{lum} forms a U-shape curve. Therefore, an empirical formula based on average intensity of the DCT block is employed to represent α_{lum}:

$$\alpha_{lum} = \begin{cases} (60 - I_{ave})/150 + 1 & I_{ave} \le 60 \\ 1 & 60 < I_{ave} < 170 \\ (I_{ave} - 170)/425 + 1 & I_{ave} \ge 170 \end{cases}, \tag{9}$$

where I_{ave} is the average intensity value of the whole DCT block (16×16 or 8×8).

As to the contrast masking factor, a block based method [11]-[13] is employed to accurately describe different masking property of different blocks. Proposed block classification method is implemented in image domain. Firstly, Canny edge detector is utilized to mark the edge pixels in the image domain. Secondly, based on the amount of edge pixels in the block, the block could be categorized into three types, namely PLANE, EDGE, and TEXTURE, respectively.

For 16×16 macroblock, block categorization is defined according to:

$$Block\_categ = \begin{cases} PLANE & \sum_{Edge} < 16 \\ EDGE & 16 \le \sum_{Edge} \le 52 \\ TEXTURE & \sum_{Edge} > 52 \end{cases}, \tag{10}$$

where $\sum_{Edge}$ denotes the amount of edge pixels in a given macroblock. Based on the block classification and considering the intra-band masking effect [11] [13], the contrast masking factor α_{cm_16} for 16×16 DCT transform is defined by:

$$\alpha_{cm\_16} = \begin{cases} 1, & EDGE \quad and \quad PLANE \quad and \quad i+j<18 \\ \max(1, (\dfrac{C(m,n,i,j)}{T_{basic}(i,j)\cdot\alpha_{lum}(m,n)})^{0.36}), & EDGE \quad and \quad PLANE \quad other\ subbands \\ \min(4, 2.25\cdot\max(1, (\dfrac{C(m,n,i,j)}{T_{basic}(i,j)\cdot\alpha_{lum}(m,n)})^{0.36})), & TEXTURE \quad and \quad i+j<18 \\ 1.25\cdot\max(1, (\dfrac{C(m,n,i,j)}{T_{basic}(i,j)\cdot\alpha_{lum}(m,n)})^{0.36}), & TEXTURE \quad other\ subbands \end{cases}, \quad (11)$$

where $C(m,n,i,j)$ is the corresponding DCT subband coefficient. Detailed information about the contrast masking factor for 8×8 block could be referred to [12].

With the definition of basic JND threshold T_{basic} based upon spatial CSF, luminance adaptation α_{lum} and contrast masking α_{cm}, two JND profiles for different block-size transforms (16×16 and 8×8) are obtained. In order to explicitly exploit the property of HVS on ABT, we further propose a strategy to determine which transform block-size to be employed for yielding the resulting JND profile.

2.3 Balanced Strategy for JND Profile between Different Block-Size Transforms

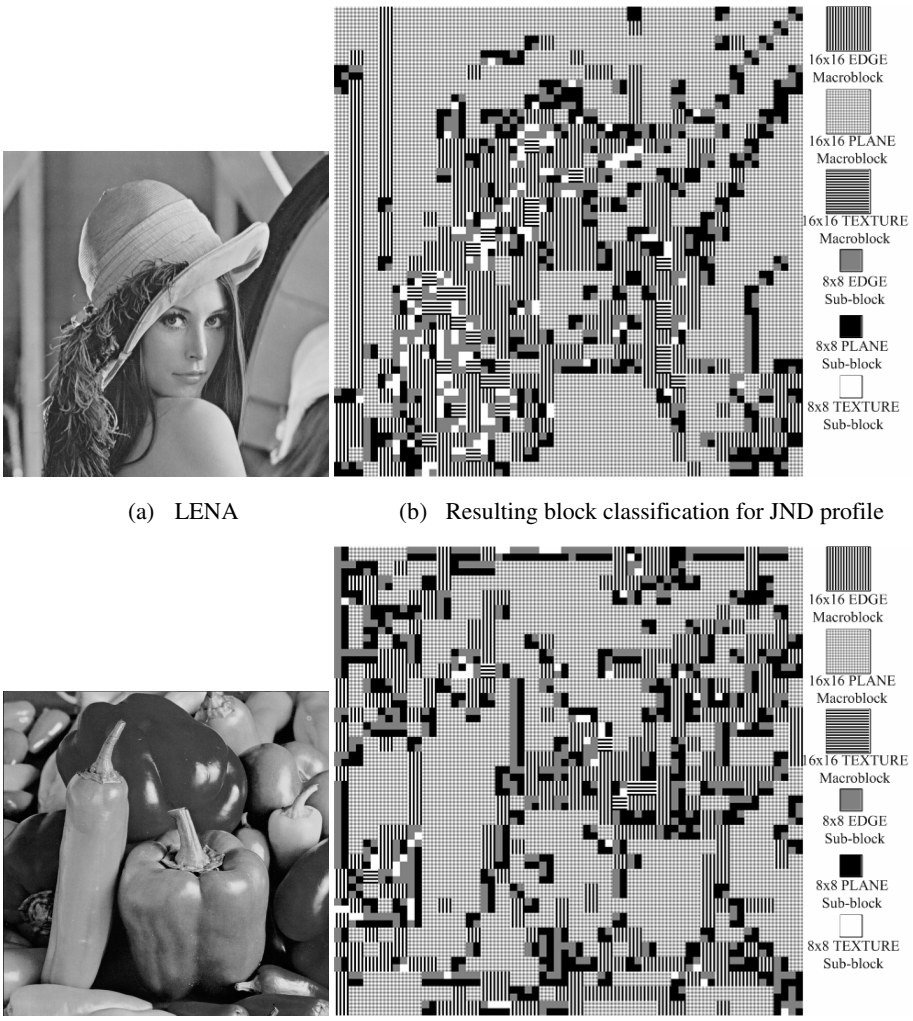
As JND profile has been extended from 8×8 to 16×16 DCT, two JND profiles based on different block-size transforms have been obtained. The 16×16 DCT would result in better energy compaction and detailed information preservation, while 8×8 DCT could represent the content and characteristic of local image blocks more efficiently. Therefore, a new strategy is proposed to adaptively adjust the image content to determine which transform block-size is utilized to generate the final JND threshold.

Table 1. Balanced Strategy between the 16×16 and 8×8 JND Profiles

16×16 macroblock type	8×8 sub-block type	JND Profile
PLANE	All 8×8 sub-blocks are PLANE	16×16 macroblock PLANE
	Otherwise	8×8 sub-block
EDGE	All 8×8 sub-blocks are EDGE	16×16 macroblock EDGE
	Otherwise	8×8 sub-block
TEXTURE	All 8×8 sub-blocks are TEXTURE	16×16 macroblock TEXTURE
	Otherwise	8×8 sub-block

As the balanced strategy is designed based on each macroblock, the image is firstly divided into 16×16 macorblocks. Since we have obtained two JND profiles based on 16×16 and 8×8 DCT, respectively, the block categorization is simply employed as the basis for balancing the two different JND profiles. Detailed information about the balanced strategy is shown in Table 1. The first column denotes the 16×16 macroblock type, while the second one indicates the 8×8 sub-block categorization. And the third column is the decision for generating the final JND profile. For example, if the 16×16 macroblock is marked as PLANE and all the 8×8 sub-blocks in the macroblock

are also marked as PLANE, 16×16 DCT-based JND model is adopted for its advantage in energy compaction and preserving detail information. Otherwise, 8×8 DCT-based JND model will be employed for each 8×8 sub-block, by considering different content information and characteristic of the sub-blocks. The results of the balanced strategy on LENA and PEPPERS are shown in Fig. 1.



(a) LENA (b) Resulting block classification for JND profile

(c) PEPPERS (d) Resulting block classification for JND profile

Fig. 1. Block classification results based on the proposed balanced strategy (Left: the original image; Right: block categorization result by considering both block type and transform block-size)

3 Experimental Results

In order to demonstrate the efficiency of our proposed JND model, distortion is inserted into each DCT coefficient of an image to evaluate the error tolerance property of HVS by:

$$\hat{M}_{typ}(m,n,i,j) = M_{typ}(m,n,i,j) + random_{(m,n,i,j)} \cdot T_{JND\_typ}(m,n,i,j), \quad (12)$$

where $\hat{M}_{typ}$ is the noise-contaminated DCT coefficient which is located on the $(i,j)th$ position of $(m,n)th$ block, $random_{(m,n,i,j)}$ takes +1 or -1 randomly to avoid introducing a fixed pattern of changes, T_{JND_typ} is the JND threshold obtained by the proposed scheme based on different transform block-sizes, *typ* denotes the final transform block-size for generating the proposed JND profile.

Table 2. PSNR Comparison between different JND Profiles

Image	Yang	Wei	Proposed JND profile
BABOON	32.53 dB	28.38 dB	27.46 dB
BARBARA	31.35 dB	29.49 dB	29.02 dB
BRIDGE	30.96 dB	29.01 dB	28.53 dB
LENA	32.72 dB	29.97 dB	29.51 dB
PEPPERS	30.78 dB	29.99 dB	29.66 dB

The proposed JND model is tested on several typical images (512×512). Also we compare our method with Yang et al.'s method [10], which evaluates the JND profile in the image domain, and Wei et al.'s JND profile [12], which is calculated in the DCT domain. Comparisons in terms of PSNR are listed in Table 2, which clearly shows that our proposed JND method yields smaller PSNR values than other JND profiles, which means that our JND profile could tolerate more distortions.

In order to provide a more convincing evaluation of the proposed JND model, subjective tests were conducted to assess the perceptual quality of the noise-contaminated images of different JND models. In the subjective test, two images were juxtaposed on the screen. One is the original image as the reference and the other is the distortion-inserted version. Ten observers (half of them are experts in image processing and the other half are not) are asked to offer their opinions on the subjective quality of the images, according to the continuous quality comparison scale shown in Table 3. In this experiment, the viewing monitor is a Viewsonic Professional series P225fb CRT display. And the viewing distance is set as 4 times the image height. Subjective test results are depicted in Table 4. The average subjective score of proposed JND model is 0.36, which means that the distortion introduced is almost invisible. And the noise-contaminated images by proposed JND model have similar visual quality with Wei et al.'s model, which is superior to Yang et al's JND profile.

Table 3. Rating Criterion for Subjective Quality Evaluation

Subjective Score	Descriptions
-3	The right on is much worse than the left one
-2	The right one is worse than the left one
-1	The right one is slightly worse than the left one
0	The right one has the same quality as the left one
1	The right one is slightly better than the left one
2	The right one is better than the left one
3	The right one is much better than the left one

Table 4. Subjective Evaluation Results (L: Noise-contaminated Image by Different JND model; R: Original Image)

Image	Yang	Wei	Proposed JND Profile
BABOON	0.5	0.2	0.2
BARBARA	1.2	0.4	0.5
BRIDGE	0.7	0.3	0.3
LENA	0.8	0.3	0.4
PEPPERS	1.0	0.4	0.4

In order to further test the efficiency of our proposed JND model, noise shaping effect is carried out for the images according to Equation (13):

$$\hat{M}_{typ}(m,n,i,j) = M_{typ}(m,n,i,j) + \tau_{typ} \cdot random_{(m,n,i,j)} \cdot T_{JND\_typ}(m,n,i,j) \,, \tag{13}$$

where τ_{typ} is an adjustable parameter to ensure the same amount of error energy (same MSE or PSNR) among the different JND models. With the constraint of same error energy, better quality of the resultant noise-contaminated image indicates a better JND profile. As our proposed JND model is based on ABT, two transform block-sizes (16×16 and 8×8) are considered. Therefore two adjustable parameters τ_{16} and τ_8 are required for the 16×16 and 8×8 transform, respectively. As 16×16 transform could better compact energy and preserve detailed information, larger distortion could be introduced with the constraint of the same error energy. Consequently, adjustable parameter τ_{16} is set larger than τ_8 for its better error masking capability.

In this paper, we take LENA image as an example. The error energy in terms of PSNR is set as 22 for comparing the visual quality after noise shaping effect. All the images shown in Fig. 2 are down-sampled with the same Gaussian kernel. From the results, we can see that the Gaussian noise contaminated image has the worst visual quality, which means that all the JND models could efficiently exploit the HVS property. And Yang et al.'s result is noisier for not considering HVS sensitivity in different spatial frequencies, while the image generated by Zhang et al.'s model appears noisy in the smooth region for its inaccurate modeling of contrast masking. Furthermore,

Fig. 2. Noise-contaminated images via different JND Models (PSNR = 22 dB). (a) Original Lena; (b) Gaussian-noise contaminated image; (c) Yang et al's model [10]; (d) Zhang et al's model [11]; (e) Wei et al's model [12]; (f) The proposed JND model.

compared with Wei et al.'s method, the proposed model could produce higher quality, especially around the brims and the surface of LENA's hat, for its accurate modeling of image content based on different transform block-sizes.

4 Conclusions

In this paper, a novel ABT-based JND profile is proposed for images by considering different energy compaction and detailed information preservation property of different transforms to exploit the HVS properties. A new strategy is proposed for each macroblock to decide what transform block-size is to be employed to represent the image content. Based on the proposed model, our JND profile could tolerate more distortions with the same visual quality compared with other JND models, which means that our model is more effective in exploiting the HVS properties.

Acknowledgement

This work was partially supported by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003). The authors are grateful to Zhenyu Wei for providing the JND code and valuable discussions.

References

1. Weber's Law of Just Noticeable Differences, http://www.usd.edu/psyc301/WebersLaw.htm
2. Ahumada, A.J., Peterson, H.A.: Luminance-model-based DCT Quantization for Color Image Compression. In: Proceedings of the SPIE, Human Vision, Visual Porcessing, and Digital Display III, vol. 1666, pp. 365–374 (1992)
3. Watson, A.B.: DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images. Society for Information Display (SID) Digest 24, 946–949 (1993)
4. Hontsch, I., Karam, L.J.: Adaptive Image Coding with Perceptual Distortion Control. IEEE Transactions on Image Processing 11(3), 213–222 (2002)
5. Lin, W., Dong, L., Xue, P.: Visual Distortion Gauge Based on Discrimination of Noticeable Contrast Changes. IEEE Transactions on Circuits and Systems for Video Technology 15(7), 900–909 (2005)
6. Lu, Z., Lin, W., Yang, X., Ong, E., Yao, S.: Modeling Visual Attention's Modulatory Aftereffects on Visual Sensitivity and Quality Evaluation. IEEE Transactions on Image Processing 14(11), 1928–1942 (2005)
7. Wolfgang, R.B., Podilchuk, C.I., Delp, E.J.: Perceptual Watermarks for Digital Images and Video. Proceedings of IEEE 87(7), 1108–1126 (1999)
8. Chou, C., Chen, C.: A Perceptual Optimized 3-D Subband Codec for Video Communication Over Wireless Channels. IEEE Transactions on Circuits and Systems for Video Technology 6, 143–156 (1996)
9. Chin, Y., Berger, T.: A Software-only Videocodec Using Pixelwise Conditional Differentialreplenishment and Perceptual Enhancements. IEEE Transactions on Circuits and Systems for Video Technology 9, 438–450 (1999)

10. Yang, X., Lin, W., Lu, Z., Ong, E., Yao, S.: Motion-Compensated Residue Pre-processing in Video Coding Based on Just-Noticeable-Distortion Profile. IEEE Transactions on Circuits and Systems for Video Technology 15, 742–750 (2005)
11. Zhang, X., Lin, W., Xue, P.: Improved Estimation for Just-Noticeable Visual Distortion. Signal Processing 85, 795–808 (2005)
12. Wei, Z., Ngan, N.K.: Spatial-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain, vol. 19, pp. 337–346 (2009)
13. Zhang, X., Lin, W., Xue, P.: Just-Noticeable Difference Estimation with Pixels in Images. Journal of Visual Communication and Image Representation 19, 30–41 (2008)
14. Dong, J., Lou, J., Zhang, C., Yu, L.: A New Approach to Compatible Adaptive Block-Size Transforms. In: Proceedings of SPIE Visual Communications and Image Processing, Bellingham, vol. 5960 (2005)
15. Qi, H., Gao, W., Ma, S., Zhao, D.: Adaptive Block-Size Transform Based on Extended Integer 8×8/4x4 Transforms for H.264/AVC. In: IEEE International Conference on Image Processing, pp. 1341–1344 (2006)
16. Ngan, K.N., Leong, K.S., Singh, H.: Adaptive Cosine Transform Coding of Image in Perceptual Domain. IEEE Transactions on Acoustics, Speech, and Signal Processing 37, 1743–1750 (1989)