Available online at www.sciencedirect.com
ScienceDirect
J. Vis. Commun. Image R. 19 (2008) 30–41
JOURNAL OF VISUAL Communication & IMAGE Representation
www.elsevier.com/locate/jvci

Just-noticeable difference estimation with pixels in images

Xiaohui Zhang [a], Weisi Lin [b], Ping Xue [a,*]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore

Received 26 May 2006; received in revised form 25 May 2007; accepted 5 June 2007
Available online 13 June 2007

Abstract

Perceptual visibility threshold estimation, based upon characteristics of the human visual system (HVS), is widely used in digital image and video processing. We propose in this paper a scheme for estimating JND (just-noticeable difference) with explicit formulation for image pixels, by summing the effects of the visual thresholds in sub-bands. The factors being considered include spatial contrast sensitivity function (CSF), luminance adaptation, and adaptive inter- and intra-band contrast masking. The proposed scheme demonstrates favorable results in noise shaping and perceptual visual distortion gauge for different images, in comparison with the relevant existing JND estimators.

Keywords: JND; Visual quality; HVS

1. Introduction

Visual just-noticeable difference (JND) refers to the minimum visibility threshold when visual contents are altered, and results from physiological and psychophysical phenomena in the human visual system (HVS). The use of JND facilitates effective image/video compression [1–8], quality evaluation [2,9], watermarking [10], etc. The JND values can be estimated for the domain of pixels or sub-bands (e.g., DCT, DWT). Substantial research has been performed in DCT sub-bands [1,2,5–7,11] due to DCT's popularity in visual signal compression (JPEG, H.261/3/4, and MPEG 1/2/4), and research was later extended in modeling with wavelet sub-bands [12–14]. On the other hand, pixel-based JND estimation models have been developed [3,4,8,15,31]. In general, sub-band JNDs are popular for perceptual image/video compression [2,5,16,17], while pixel-based JND models are often used in motion estimation [32], visual quality evaluation [9] and video replenishment [4] to avoid extra decomposition.

Sub-band JND estimators usually consider all the major affecting factors, namely, contrast sensitivity function (CSF), luminance adaptation, and contrast masking. A well-cited scheme for DCT thresholds was developed by Ahumada and Peterson [1], based upon the spatial CSF. It becomes the basis of a number of improved or simplified models [2,5,6,11], and has been extended for color images [18]. The basic scheme was improved into the DCTune model [2] by Watson after contrast masking [19,20] had been calculated. Hontsch and Karam [5] modified the DCTune model with a foveal region being considered instead of a single pixel. Tong and Venetsanopoulos classified the image into edge, plain, and texture regions [7] to evaluate the contrast masking. In [6], Cortex transform is used to map DCT coefficients for mimicking the HVS' contrast masking characteristics more closely. Since a *quasi-parabola* function[1] is more realistic for luminance adaptation in real-world digital image display [3,4], and it has been therefore incorporated in our earlier work [11].

Most pixel-wise models developed so far have merely focused on luminance adaptation and texture masking. In

* Corresponding author. Fax: +65 67933318.
E-mail addresses: p144238551@ntu.edu.sg (Xiaohui Zhang), wslin@ntu.edu.sg (Weisi Lin), epxue@ntu.edu.sg (Ping Xue).

[1] The HVS' sensitivity reaches the highest with medium background gray level in digital images.

1047-3203/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2007.06.001

Chou and Li's model [3], texture masking is estimated with the maximum signal from four edge detectors with 45° apart, and the JND is determined by the maximum effect of luminance adaptation and texture masking. In Chiu and Berger's model [4], texture masking is determined by the maximum grey-level difference of the central pixel and its neighbors in horizontal and vertical directions, and luminance adaptation is assumed as the dominant effect. Yang et al. [8] improved Chou and Li's model with a nonlinear additivity formula to integrate luminance adaptation and texture masking for more aggressive JND threshold estimation that matches the HVS' characteristics. Ramasubramanian et al.'s model [15] formulates the spatial CSF roughly for pixel-domain JND, via Laplacian pyramid image decomposition, since only 6 frequency points are used. Dumont et al. [34] developed their perceptual metric which is based upon the similar concept as Ramasubramanian's model and uses pyramid decomposition to calculate the spatial contribution to the threshold from each sub-band. More accurate pixel-based thresholds can be used for better motion estimation [32], video coding [4] and perceptual visual quality evaluation ([9], without the need of sub-band decomposition).

The existing models estimate JNDs in either sub-bands or pixels. In fact, the JNDs in different domains are due to the same underlying mechanism in the HVS, so there should be a way for conversion between the two JND estimations. In [3], pixel-based JND has been converted to sub-bands, without consideration of CSF (contrast sensitivity function). In [31], DCTune model for sub-bands has been converted to the pixel domain; however, the contrast masking and realistic luminance adaptation have not been addressed. In some approaches (e.g., [33]), perceptual visual error is evaluated between the original image and a processed image, and no explicit formulae are given for JND estimation.

In this paper, we propose explicit formulae to derive pixel-based JND by summing the effects of the visual thresholds in sub-bands, and considering all the relevant factors, i.e., CSF, luminance adaptation for digital images and contrast masking. Apart from this, the proposed scheme is also more efficient than that in [31] since only one IDCT process is required (in [31], IDCT is needed for both the noise injected image and the original image). Section 2 firstly presents a DCT-based JND model that incorporates all major affecting factors, namely, spatial CSF, *quasi-parabola* luminance adaptation, and adaptive inter- and intra-band contrast masking. Only the generalized formulae in the main parts are presented in this section, while the implementation specific details are included as Appendices. A methodology is then proposed in Section 3 to derive pixel-wise JNDs from the formulated DCT-based model. The major improvement of the resultant pixel-wise JND estimator over the existing relevant models is the more accurate spatial CSF consideration. The experimental performance has been demonstrated and compared with the existing relevant models in Section 4, with applications in noise shaping and visual quality evaluation. The last section gives the concluding remarks.

2. DCT-based JND estimation

The overall DCT-JND of the (i,j)-th DCT sub-band can be determined by the base visibility threshold T due to the spatial CSF, the luminance adaptation factor a_{Lum} and the contrast masking factor a_{e} [5,11]:

$$t_{\mathrm{JND}}(n_1,n_2,i,j) = T(i,j)\cdot a_{\mathrm{Lum}}(n_1,n_2)\times a_{\mathrm{e}}(n_1,n_2,i,j) \tag{1}$$

where (n_1,n_2) indicates the DCT-block location in an image.

For 8-bit image representation, the base visibility threshold can be determined as [1,18]:

$$T(i,j) = \chi_{i,j}\cdot N\cdot T^{o}(i,j) \tag{2}$$

where

$$\chi_{i,j} = \begin{cases} 1 & i=j=0 \\ 1/\sqrt{N} & i=0 \text{ or } j=0 \\ 1/2 & i,j\neq 0 \end{cases}$$

and N is the dimension of a DCT block; $T^{o}(i,j)$ is derived from the spatial contrast sensitivity function (CSF), as plotted in Fig. 1, which is based on experimental data obtained in [21]. The detailed formulation and parameter determination for $T^{o}(i,j)$ are presented in Appendix A.

Viewing experiments have been done [3,16] in order to determine the relationship between the threshold and the grey level of a digital image displayed on a monitor, and shown that the brightness adaptation is with *quasi-parabola* curves, i.e., a higher visibility threshold occurs in either very dark or very bright regions in an image, and a lower visibility threshold occurs in regions with medium brightness [3,16,22–24]. The *quasi-parabola* luminance adaptation is modeled as follows [11]:

$$a_{\mathrm{Lum}}(n_1,n_2) = \begin{cases} 2\left(1-\frac{C(n_1,n_2,0,0)}{128\cdot N}\right)^3+1 & \text{if } C(n_1,n_2,0,0)\leqslant 128\cdot N \\ 0.8\left(\frac{C(n_1,n_2,0,0)}{128\cdot N}-1\right)^2+1 & \text{otherwise} \end{cases} \tag{3}$$

where $C(n_1,n_2,0,0)$ is the DC component of the DCT block. Eq. (3) adapts well to the HVS characteristics in low and high luminance regions in digital images. The coefficients in the upper equation are determined based on the finding that the luminance adaptation factor decreases as grey-level increases as long as it is below 128. And the coefficients in the lower equation are determined in order to approximate Watson's DCTune model (for grey-level above 128).

Contrast masking is the reduction in the visibility of one visual component at the presence of another one [19,20]. Usually noise becomes less visible in the regions with high texture energy, and more visible in smooth areas. However, the HVS has acute sensitivity at or near edges in an image [25] (where texture energy is also high), because edge

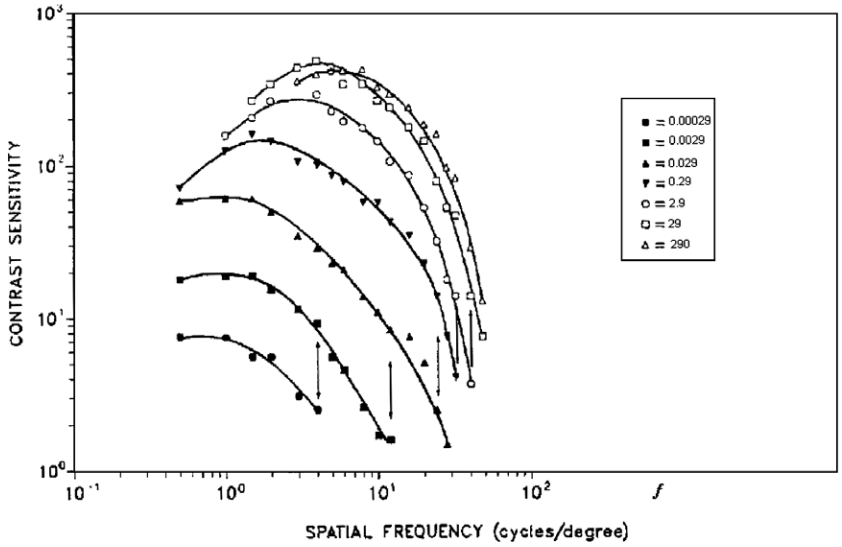


Fig. 1. Spatial contrast sensitivity function (Csf) [21] (different curves corresponding to different luminance values).

structure is simpler than a texture one and typical observers have some prior knowledge on what edges should look like [26]. Following the methodology in [7], each DCT block can be assigned to one of the three classes with descending order of the HVS sensitivity, namely, *PLAIN*, *EDGE*, and *TEXTURE*, according to the energy in low-frequency (LF), medium-frequency (MF) and high-frequency (HF) parts (as shown in Fig. 2).

Let $\xi(n_1,n_2)$ denote the extent of inter-band masking. In addition to the inter-band masking effect classified by Tong and Venetsanopoulos [7], the signal in the same sub-band also contributes to the masking factor [19]. The contrast masking factor can be therefore decided with the combined effect of inter- and intra-band masking [2,5]:

$$a_e(n_1,n_2,i,j) = \begin{cases} \xi(n_1,n_2) & \text{for } (i,j)\in \text{LF}\cup\text{MF in Edge block} \\ \xi(n_1,n_2)\times\max\left\{1,\left(\frac{C(n_1,n_2,i,j)}{T(n_1,n_2,i,j)}\right)^{0.36}\right\} & \text{otherwise} \end{cases} \tag{4}$$

where $C(n_1,n_2,i,j)$ represents the DCT coefficient. Since the HVS is sensitive for changes on edges, the LF and MF parts for an *EDGE* block are excluded from the intra-band masking evaluation to avoid over-estimation of JND. The determination of $\xi(n_1,n_2)$ is discussed in Appendix B.

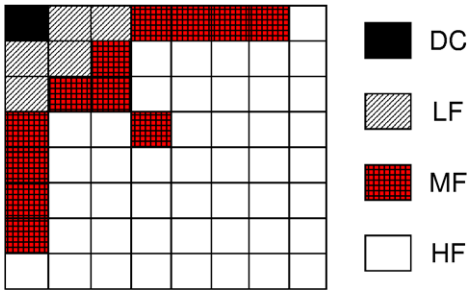


Fig. 2. DCT block classification [7].

3. Pixel-wise JND estimation

The JND in different domains results from the same HVS mechanism. The JND for each pixel can be obtained by proper summation of the influence from all DCT sub-bands within the same block. Any noise below the corresponding $t_{\text{JND}}(n_1,n_2,i,j)$ would not introduce any visible difference to the HVS; if the magnitude of the corresponding DCT coefficient, $C(n_1,n_2,i,j)$, is less than $t_{\text{JND}}(n_1,n_2,i,j)$, the contribution of the latter can be ignored for the resultant JND at the pixel. The contribution of $t_{\text{JND}}(n_1,n_2,i,j)$ can be therefore evaluated as:

$$t'(n_1,n_2,i,j) = \begin{cases} \text{sign}_{C(n_1,n_2,i,j)}\cdot t_{\text{JND}}(n_1,n_2,i,j) & \text{if } |C(n_1,n_2,i,j)| \geqslant t_{\text{JND}}(n_1,n_2,i,j) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\text{sign}_{C(n_1,n_2,i,j)}$ denotes the sign of $C(n_1,n_2,i,j)$, and its use avoids the possible discontinuity on boundaries of neighboring blocks or artificial patterns over blocks. The similar concept has been adopted by Walter et. al. [31] when DCTune model was used to estimate JND in the pixel domain.

The compound effect for an image pixel at (n_1,n_2,x,y) is obtained by summing (inverse-DCT transforming) the DCT-based thresholds in the same block:

$$\begin{aligned} t_{\text{P}}(n_1,n_2,x,y) &= \text{IDCT}(t'(n_1,n_2,i,j)) \\ &= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\phi_i\phi_j\cos\left(\frac{(2x+1)i\pi}{2N}\right)\right. \\ &\quad \left.\times\cos\left(\frac{(2y+1)j\pi}{2N}\right)t'(n_1,n_2,i,j)\right) \end{aligned} \tag{6}$$

where x and y are the pixel position indices in a block and vary from 0 to $N-1$, and

$$\phi_u = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & u \neq 0 \end{cases}.$$

As suggested by Chou and Li [3], the JND threshold at a pixel can be hence determined by the dominant factor between the spatial masking effect and the luminance adaptation effect:

$$T_{\mathrm{P}}(n_1, n_2, x, y) = \max\{|t_{\mathrm{P}}(n_1, n_2, x, y)|, t_l(n_1, n_2)\} \quad (7)$$

where $t_l(n_1, n_2)$ accounts for the local background luminance adaptation and can be directly obtained from the DC component in Eq. (1), $t_{\mathrm{JND}}(n_1, n_2, 0, 0)$:

$$t_l(n_1, n_2) = t_{\mathrm{JND}}(n_1, n_2, 0, 0)/N \quad (8)$$

The complete procedures of the proposed JND estimator are shown in Fig. 3. Both the JND calculation (Sections 2 and 3) and the JND performance evaluation (the next section) were carried out under the following viewing condition: a CRT monitor (NEC MultiSync FE770) with viewing distance of $\ell = 50$ cm in a room illuminated by fluorescent ceiling lights. The monitor in our experiments was set as 50% of the full scale for both contrast and brightness. Actually, the monitor setting does not have significant impact on the JND calculation, because in the normal viewing condition for digital images, if the global mean luminance L varies within a very small range (it is the case in digital images), the shift of contrast sensitivity curve is almost negligible.

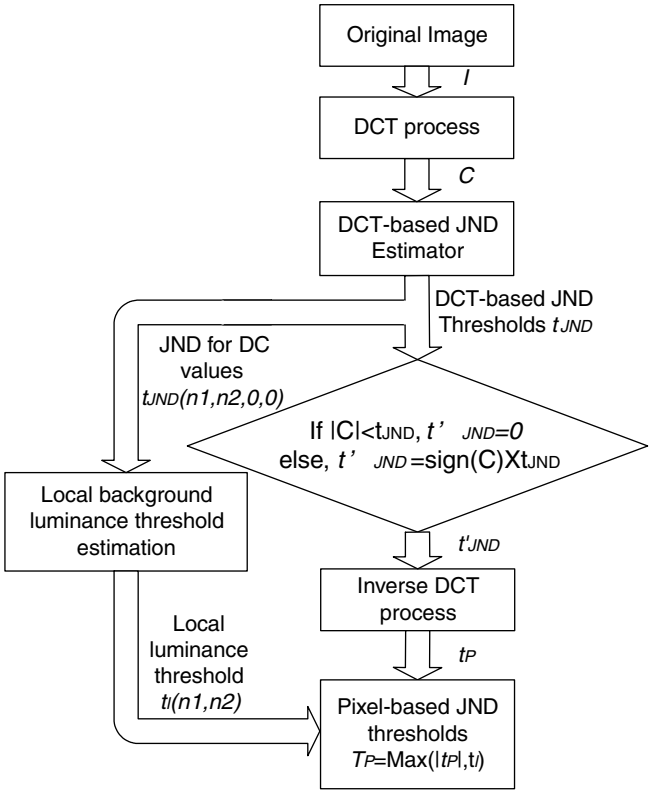


Fig. 3. Block diagram of the derivation for DCT based JND thresholds T_{JND} and pixel-based JND thresholds T_{P}.

4. Performance

A better sub-band-based JND estimator will lead to better JND results in the pixel domain, using the formulae in Section 3. In our earlier work [11], the scheme presented in Section 2 has demonstrated better performance against Watson's DCTune [2] in the DCT domain. As presented in Section 2 and Appendix B, new formulae for adjustment of luminance adaptation is used, and the block classification is modeled for more accurate JND estimations in *PLAIN*, *EDGE*, and *TEXTURE* regions.

In this section, we will evaluate the proposed scheme's performance against five existing relevant models in the pixel domain with explicit JND formulation: *Model I*—Chou and Li's model [3], *Model II*—Yang et al.'s model [8], *Model III*—Chiu and Berger's model [4], *Model IV*—the modified Ramasubramanian et al.'s model [15][2], and *Model V*—Walter's model [31], with different scenarios and images.

4.1. Noise shaping effect

The performance of a JND model can be evaluated by its effectiveness in noise shaping for images. For pixel-based JND models, a noise-contaminated image for an image I can be yielded as [3]:

$$I_n(X, Y) = I(X, Y) + S_{X,Y}^{\mathrm{random}} \cdot \tau \cdot T_{\mathrm{P}}(X, Y) \quad (9)$$

where (X, Y) represents a pixel position; $T_{\mathrm{P}}(X, Y)$ is $T_{\mathrm{P}}(n_1, n_2, x, y)$ at (X, Y) with $X = n_1 \cdot N + x$ and $Y = n_2 \cdot N + y$; $S_{X,Y}^{\mathrm{random}}$ takes $+1$ or -1 randomly regarding X and Y, to avoid introduction of fixed pattern of changes; and $\tau(\tau > 1)$ is an adjustable parameter to ensure a same amount of error energy (therefore same MSE or PSNR) among different JND estimators. Under the same error energy, better visual quality of the resultant image $I_n(X, Y)$ indicates a better JND estimator.

The noise-contaminated *Cameraman* images obtained from various pixel-based JND models are shown in Fig. 4, by Eq. (9) with the same noise energy (PSNR = 27.63 dB). *Model III* results in a poorer visual quality [Fig. 4(C)] due to simplicity on texture masking measurement; *Model I* [Fig. 4(A)] encounters the similar problem of underestimation of texture masking; its improved version, *Model II*, achieves a slightly better visual quality (Fig. 4(B)). With *Model IV* (Fig. 4(D)), less noise is

[2] With the 6 pyramid sub-bands set at [0.5 1.0 5 2.1 4.2 8.4 16.8] (cpd), the associated CSF parameters are determined as [1 1 1 1 1 5 4.3] using the CSF equation in [15], instead of [1 1 1 .0 2 1 .5 7 4 .2 0 3 1 .3 2] in the original model for the sub-bands at [1,2,4,8,16,32] (cpd). To avoid the overestimation of the thresholds at low frequencies, the CSF sensitivity factor below 4 cpd has been normalized as 1, as suggested by Ramasubramanian et al. [15].

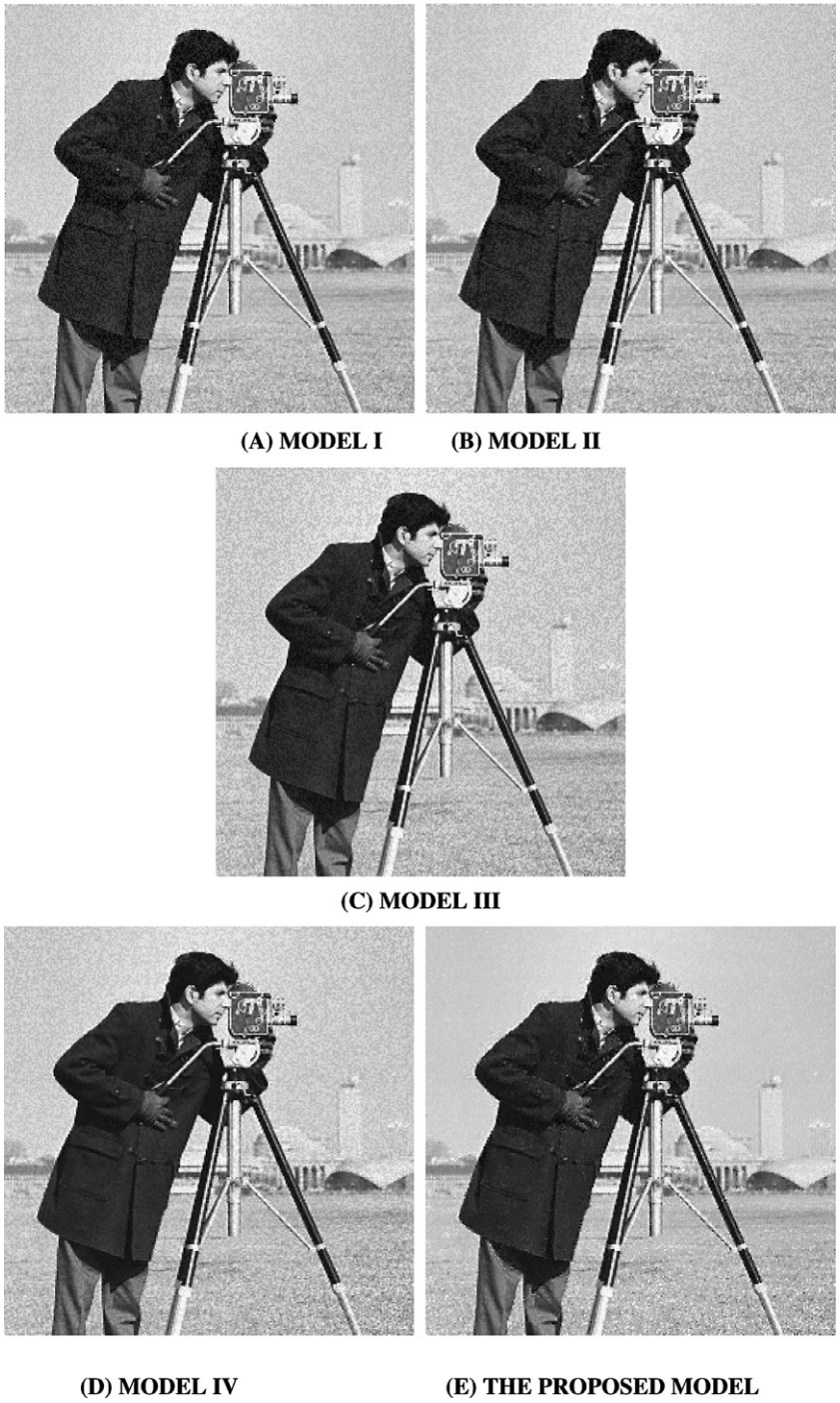


Fig. 4. Noise-contaminated images via different models (PSNR = 27.63 db).

injected to dark regions (e.g., the man's coat), and excessive noise is injected to smooth regions (e.g., the sky). The former is due to the use of the monotonically increasing function that leads to underestimating luminance adaptation in dark regions; the latter is due to the indiscrimination for smooth areas where the HVS is more sensitive and the coarse CSF implementation that fails to account for the rapid sensitivity decrease in those sub-bands above 10 cpd (cycles/degree).

The proposed model [as shown in Fig. 4(E)] achieves the highest visual quality by shaping more noise into the non-smooth and dark regions. This is because of the more accurate CSF implementation, as well as the discrimination of different (i.e., smooth, edge, and texture) regions and utilization of the more realistic *quasi-parabola* function for luminance adaptation.

The visual comparison results have been confirmed by the subjective tests, in which ten observers, 3 females and

7 males aged 20–27, were involved. The observer's eyesight was either normal or had been corrected to normal with spectacles. At each time two images were shown on the screen: the original image on the left as the reference, and the noise-injected image on the right. The observers gave the noise-injected image a quality score 0–5, defined as *perfect*, *very good*, *good*, *acceptable*, *poor*, and *unbearable*, after viewing the display at least 4 seconds. The order of the images to be displayed for each observer was randomized so that the possible bias due to viewing experience was minimized. We have three test images, ''Cameraman'', ''Actor'' and ''Mandrill'' contaminated with noise at different PSNR levels. The average distortion score is calculated from the average of the ten reviewers' scores. A lower average distortion score indicates a less distorted image. As shown in Table 1, the proposed JND model is associated with a lower perceptual distortion score for a PSNR level.

4.2. Visual quality gauging

Traditional error measures for images, such as mean square error (MSE) and peak signal-to-noise ratio (PSNR), do not provide a good gauge of the HVS' perceptual image quality in many cases [25–27]. Let $I(X, Y)$ and $I'(X, Y)$ represent pixel values at position (X, Y) in an image and the corresponding distorted image, respectively; a JND-scaled distortion measure can be defined as the difference of noticeable image details, for better perceptual error evaluation in pixel domain:

Table 1
Perceptual distortion measures for noise-contaminated images

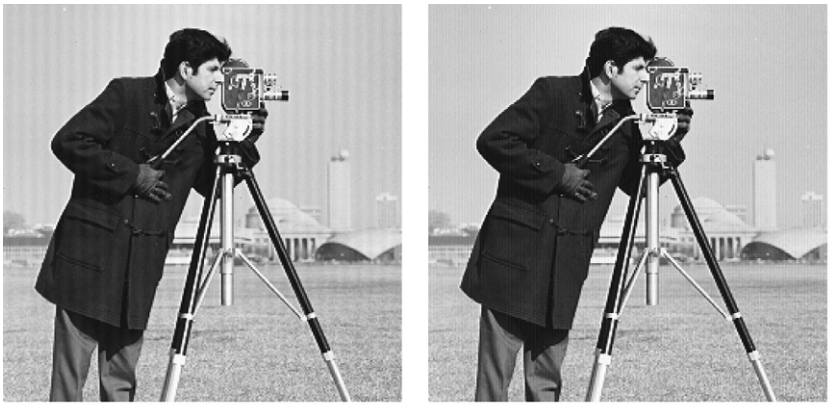
JND model	Average subjective distortion score		
	Cameraman (PSNR = 27.6 dB)	Actor (PSNR = 22.3 dB)	Mandrill (PSNR = 20.6 dB)
Model I	3.4	2.2	4.0
Model II	2.4	2.3	4.1
Model III	3.6	2.1	3.1
Model IV	2.8	3.1	3.9
Ours	1.1	1.5	2.8

$$p(X, Y) = \left| \frac{I(X, Y) - I'(X, Y)}{T_P(X, Y)} \right| \tag{10}$$

where $T_P(X, Y)$ is the pixel-wise JND thresholds for the image. The value of $p(X, Y)$ is expected to follow HVS perception more closely than the traditional MSE/PSNR distortion measure. Better JND estimation facilitates the perceptual distortion measurement to be closer to the HVS' perception for an image.

The main difference between the proposed pixel-based scheme and *Models I–IV* lies on whether and how the CSF is incorporated. In order to test all models regarding CSF, we superimposed vertical and horizontal sine wave gratings of different frequencies into various images, and evaluated how well $p(X, Y)$ predicts the visual quality changes. The similar evaluation method was proposed in [28] for comparing the accuracy of two quality metrics. Fig. 6 shows the *Cameraman* image contaminated by a vertical sine wave grating at 4 and 14.6 cpd, respectively, with same total noise energy, and these two cases are hereinafter noted as *NA4* and *NA14.6*. For human observers, it is easy to detect the grating in *NA4* (Fig. 5(A)), while the grating in *NA14.6* [Fig. 5(B)] is almost invisible. This consists with the spatial CSF since the HVS is less sensitive to high frequency signals. The corresponding JND-scaled distortion map calculated with [10] for each JND model is shown in Fig. 6. In the figure, bright areas represent highly visible distortions predicted with a JND model and Eq. (10), while dark areas correspond to the invisible distortions predicted in the same way.

Models I–III cannot differentiate the sensitivity for spatial frequencies; hence they cannot predict the dominant visual distortion well with *NA14.6*. This can be observed in Fig. 6(A2, B2, C2) where the high frequency noise is identified incorrectly (in contradiction with the human observation) as the dominant distortion. $T_P(X, Y)$ takes almost same values for *NA4* and *NA14.6*, since the three models are not frequency dependent. Due to the absolute-value operation in Eq. (10), the vertical patterns for *NA4* [Fig. 6(A2, B2, C2)] appear with a frequency higher than



(A) NOISE ADDED AT 4 CPD (NA4) **(B) NOISE ADDED AT 14.6 CPD (NA14.6)**

Fig. 5. Noise-contaminated images for cameraman (MSE = 6.25).

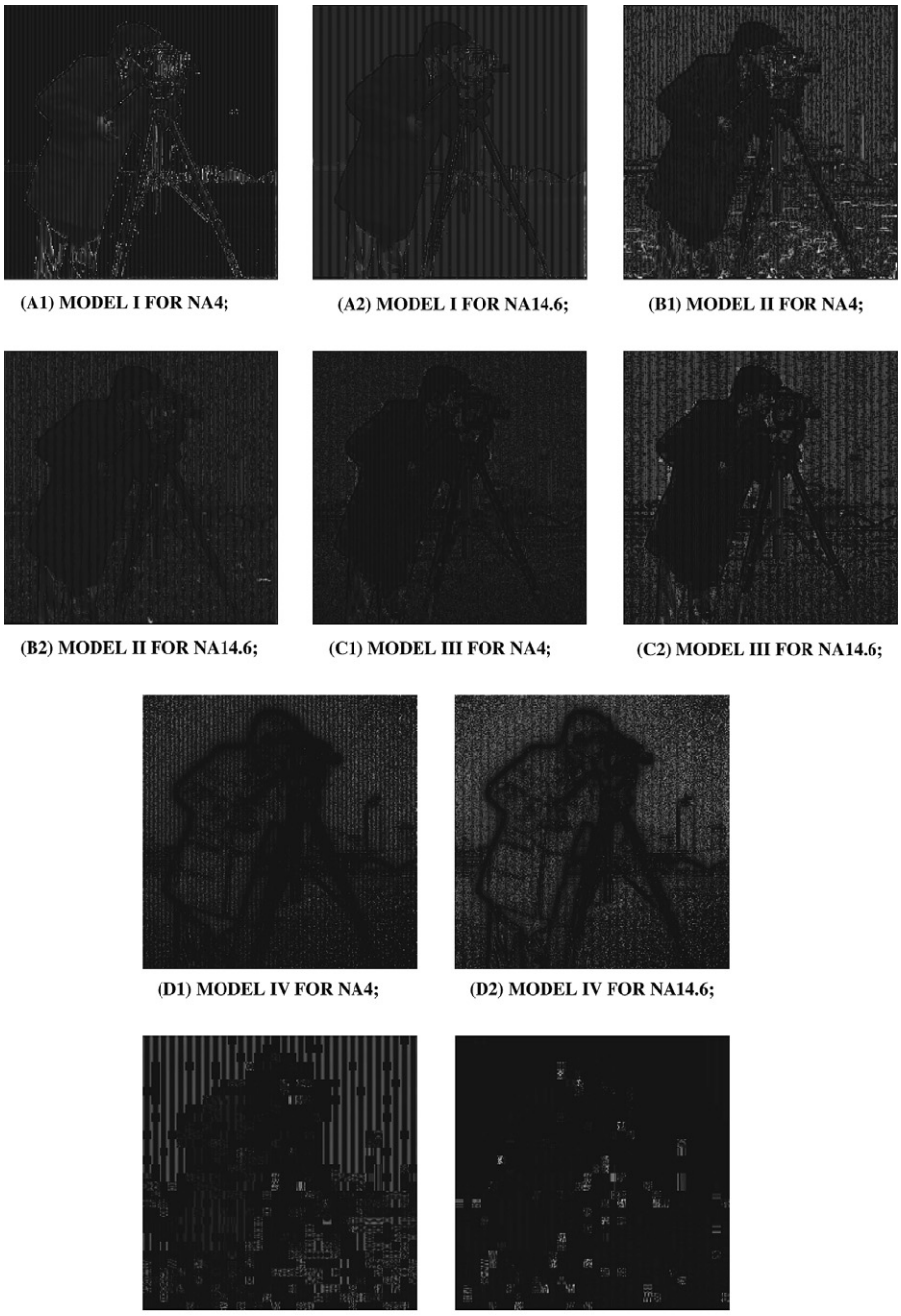


Fig. 6. JND-scaled distortion maps, with a brighter grey level corresponding to higher predicted distortion.

4 cpd, whilst those for *NA14.6* [Fig. 6(A1, B1, C1)] appear with a frequency lower than 14.6 cpd (because of the *smoothness* effect similar to that described in [29]). *Model II* works slightly better by predicting relatively small distortion for *NA14.6* [Fig. 6(B2)] than *Model I*. *Model IV* also fails to predict the human observation, because of its poor resolution for CSF (and the noise in *NA14.6* is not with a same peak frequency as any sub-band).

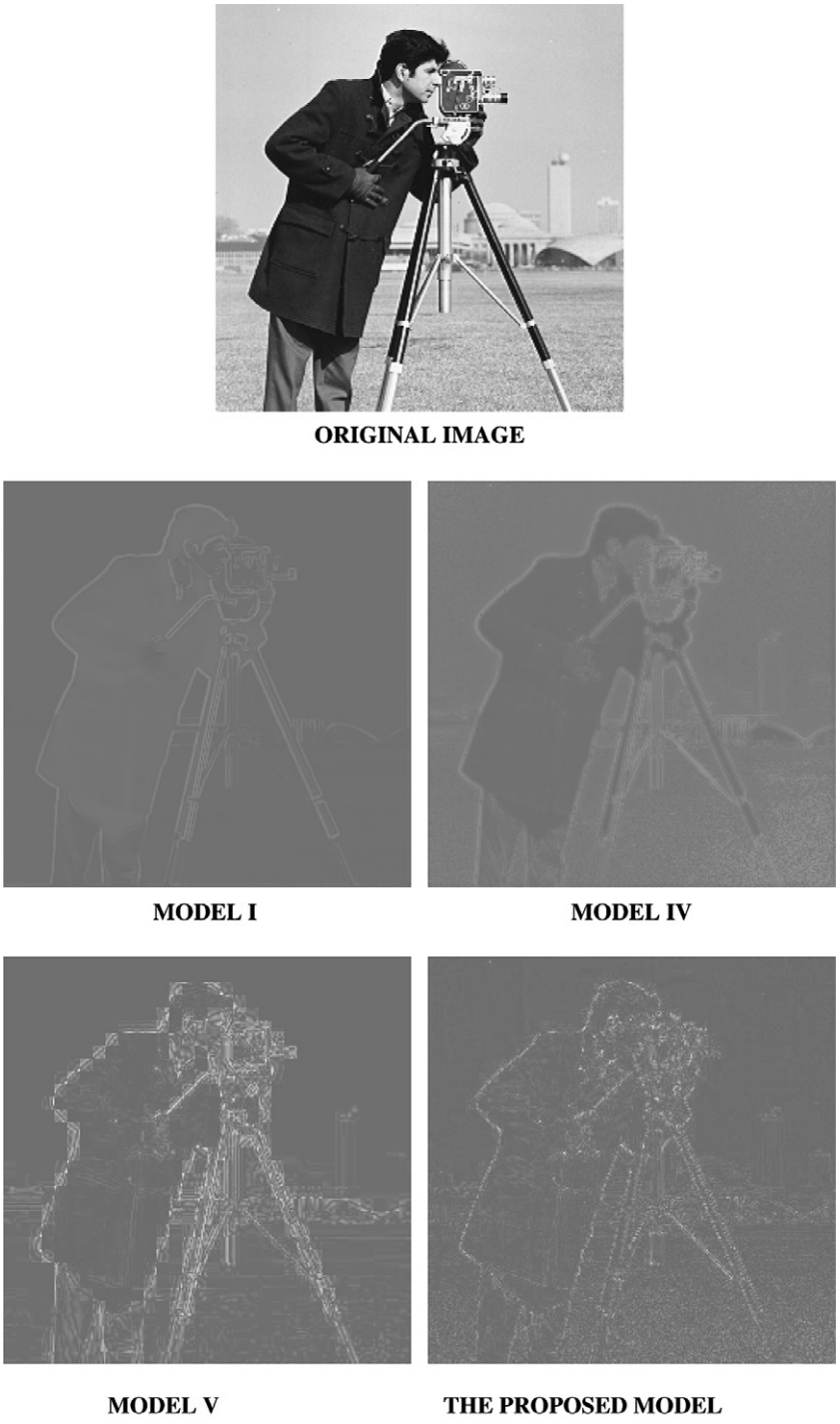
As can be seen in Fig. 6(E1, E2), the proposed model outperforms all above models by giving the closest predicted distortion map in comparison with human observation: rightfully highlighting the visual disturbance from the 4 cpd sine wave in Fig. 6(E1), and successfully excluding the influence from the 14.6 cpd sine wave in Fig. 6(E2). The sparse bright blocks in Fig. 6(E2) are caused by the different results of block classification between the original image and the distorted image in DCT JND evaluation.

4.3. Just-noticeable distortion map

To further demonstrate the JND thresholds calculated from different models, we show the JND maps for four

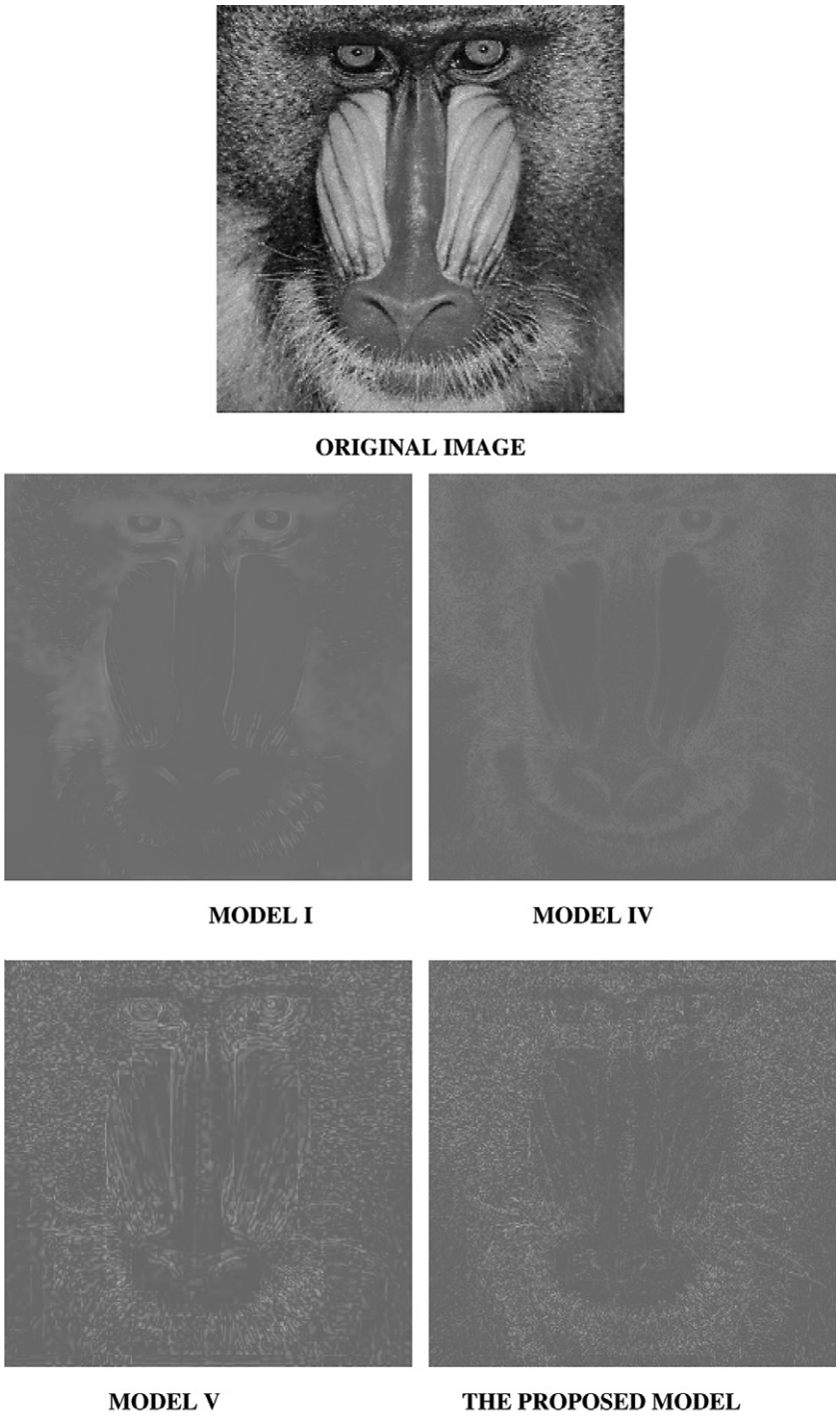
models, *Model I* (purely operated in pixel domain), *Model IV* (pyramid decomposition to several sub-bands), *Model V* (converted from sub-band domain), and our model (converted from sub-band domain), in Fig. 7 for three images (Cameraman, Mandrill, and Actor).

From the JND maps, it can be seen that *Model I* has under-estimated the threshold in texture areas, and this is a common drawback of JND models purely operated in pixel domain without incorporation of CSF. It is also observed that *Model IV* presents a relatively high threshold



(a) Cameraman

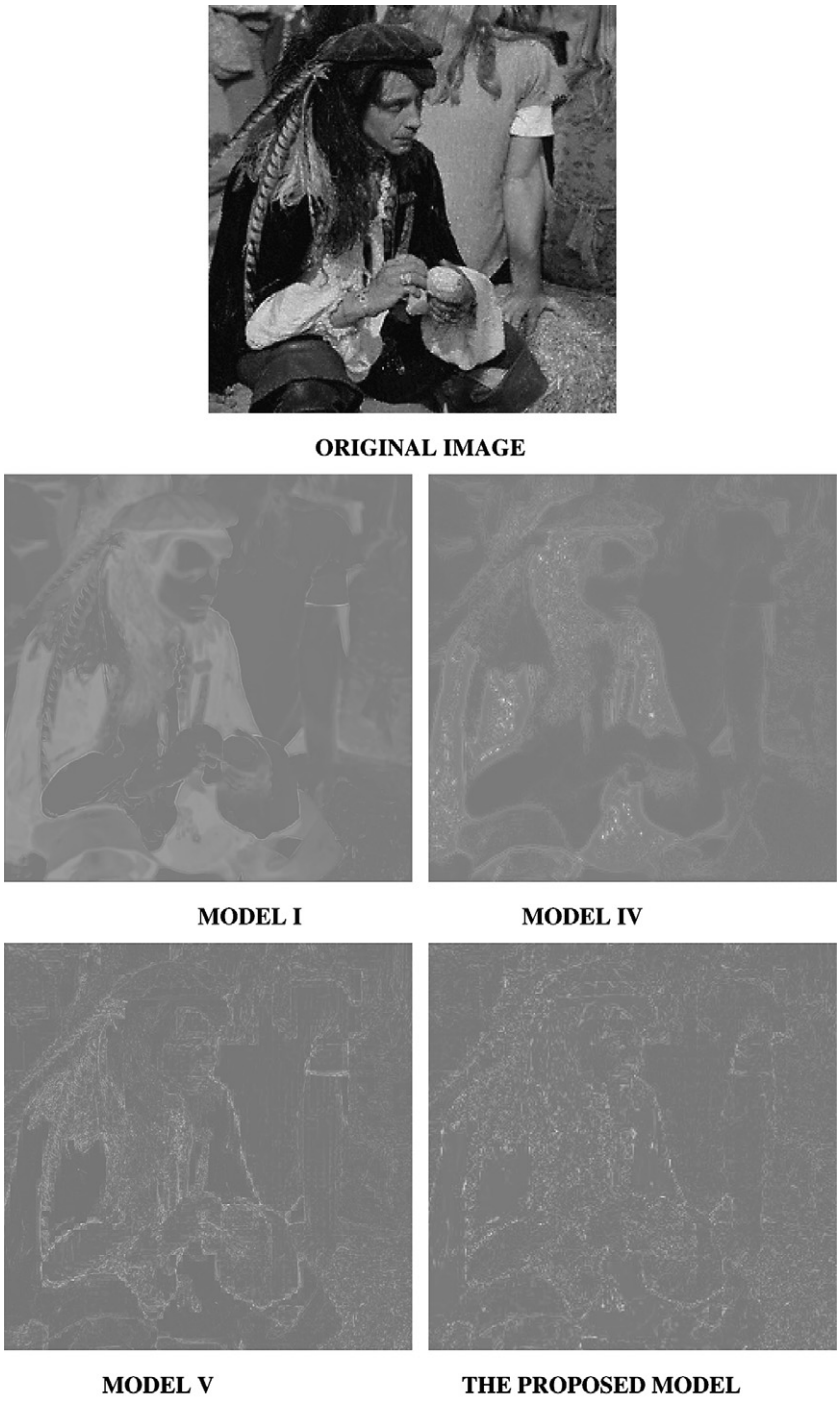
Fig. 7. Scaled JND threshold maps (The JND threshold maps have been scaled at a same ratio in order to have a better view, brighter areas represent higher threshold values).



(b) MANDRILL

Fig. 7 (*continued*)

values in some smooth areas (such as the sky) and low threshold values in dark regions. These observations are consistent with discussions in the previous sections. Although *Model V* has a similar concept as ours, in which both threshold maps in pixel domain have been converted from JND thresholds calculated in DCT domain by inverse DCT transform, we can observe that *Model V* exhibits some edge distortion near the edges [e.g. the edge of Cameraman's clothes in Fig. 7(a)]. This is because traditional JPEG quantization matrix is not adaptive enough for edge regions, while in our model the edge regions have been classified adaptively to avoid overestimation of threshold in edge regions.



(c) ACTOR

Fig. 7 (*continued*)

5. Conclusion

This paper devised a scheme for estimating JND (just-noticeable difference) in the pixel domain via conversion from that in sub-bands, based upon the fact that JNDs in different domains result from the same HVS characteristics. In this work, all the major affecting factors have been formulated, inclusive of spatial contrast sensitivity function (CSF), luminance adaptation for digital images, and intra- and inter-band contrast masking.

The proposed scheme outperforms the relevant existing models in deriving a JND value closer to the HVS visibility bound for each image sub-band or pixel, as demonstrated in different scenarios (i.e., different types of experiments, different test images, and different distortion levels). It has yielded more favorable results than the other relevant

existing models in applications like noise shaping and visual distortion measurement. The benefits of more accurate visibility threshold determination can be translated into resource (computational power, bitrate, etc.) saving and performance (e.g., resultant visual quality) enhancement in various other applications.

Appendix A. Approximation of spatial CSF

$T^o(i,j)$ is calculated as an approximation of the spatial CSF curves as shown in Fig. 1 [1,18]:

$$\log T^o(i,j) = \log \frac{s \cdot T_{\min}}{r + (1-r)\cos^2 \theta_{i,j}} + K(\log f_{i,j} - \log f_p)^2 \tag{A1}$$

being the function of background luminance L and spatial frequency $f_{i,j}$. Note that the visibility threshold $T^o(i,j)$ is related to the reciprocal of the sensitivity shown in Fig. 1.

$$f_{i,j} = \frac{1}{2N_{\mathrm{DCT}}} \sqrt{\frac{i^2}{\omega_x^2} + \frac{j^2}{\omega_y^2}} \tag{A2}$$

where ω_x and ω_y are the horizontal and vertical visual angles of a pixel, and can be calculated based on viewing distance ℓ and the display width of a pixel Λ on the monitor:

$$\omega_\hbar = 2 \cdot \arctan\left(\frac{\Lambda_\hbar}{2 \cdot \ell}\right), \quad \hbar = x, y \tag{A3}$$

The minimum threshold $T_{\min}$ and the corresponding frequency f_{p} are calculated as:

$$T_{\min} = \begin{cases} 0.142 \times \left(\frac{L}{13.45}\right)^{0.649} & L \leqslant 13.45\ \mathrm{cd/m^2} \\ \frac{L}{94.7} & \text{otherwise} \end{cases} \tag{A4}$$

$$f_{\mathrm{p}} = \begin{cases} 6.78 \times \left(\frac{L}{300}\right)^{0.182} & L \leqslant 300\ \mathrm{cd/m^2} \\ 6.78 & \text{otherwise} \end{cases} \tag{A5}$$

and

$$K = \begin{cases} 3.125 \times \left(\frac{L}{300}\right)^{0.0706} & L \leqslant 300\ \mathrm{cd/m^2} \\ 3.125 & \text{otherwise} \end{cases} \tag{A6}$$

$$\theta_{i,j} = \arcsin \frac{2 f_{i,0} f_{0,j}}{f_{i,j}^2} \tag{A7}$$

In Eq. (A1), s and r are set to 0.25 and 0.6 to account for the *spatial summation* effect [18] and the *oblique* effect [1,18], respectively. The coefficients used in the equations are determined based on standard CSF curves.

Appendix B. Inter-band contrast masking

The inter-band contrast masking is determined with the methodology in [7]. Let L, M, and H represent the sums of the absolute DCT coefficient values in LF, MF, and HF parts (as shown in Fig. 2), respectively. The texture energy for a DCT block is approximated by:

$$TexE = M + H \tag{B1}$$

Table 2
Conditions for block classification

Conditions		Classification
TexE	Inequality (B4)	
$TexE \leqslant 125$	—	*PLAIN*
$125 < TexE \leqslant 290$	if (B4) is met for $\kappa = 1$	*EDGE*
	otherwise	*PLAIN*
$290 < TexE \leqslant 900$	if (B4) is met for $\kappa = 1$	*EDGE*
	otherwise	*TEXTURE*
$TexE > 900$	if (B4) is met for $\kappa = 0.1$	*EDGE*
	otherwise	*TEXTURE*

The presence of edge can be indicated by [7,30]:

$$E_1 = (\bar{L} + \bar{M})/\bar{H} \tag{B2}$$

and

$$E_2 = \bar{L}/\bar{M} \tag{B3}$$

where $\bar{L}$, $\bar{M}$ and $\bar{H}$ denote $L/5$, $M/12$ and $H/46$, respectively.

A block is classified as an *EDGE* one if

$$E_1 \geqslant 16 \tag{B4-a}$$

or

$$\max\{E_1, E_2\} \geqslant 7\kappa \ \&\ \min\{E_1, E_2\} \geqslant 5\kappa \quad \text{where } \kappa \leqslant 1 \tag{B4-b}$$

The overall block classification is performed according to Table 2, and the extent of inter-band masking is then measured as:

$$\xi(n_1,n_2) = \begin{cases} 1 + [(TexE(n_1,n_2) - \mu_2)/(2\mu_3 - \mu_2)] \cdot 1.25 & \text{for } \textit{TEXTURE} \text{ block} \\ 1.25 & \text{for } \textit{EDGE} \text{ block and } L + M > 400 \\ 1.125 & \text{for } \textit{EDGE} \text{ block and } L + M \leqslant 400 \\ 1 & \text{for } \textit{PLAIN} \text{ block} \end{cases} \tag{B5}$$

More details of the block-based contrast masking model and parameterization can be found in [7,11].

References

[1] A.J. Ahumada, H.A. Peterson, Luminance-model-based DCT quantization for color image compression, Proceedings of the SPIE, Human Vision, Visual Processing, and Digital Display III 1666 (1992) 365–374.
[2] A.B. Watson, DCTune: a technique for visual optimization of DCT quantization matrices for individual images, Society for Information Display (SID) Digest 24 (1993) 946–949.
[3] C.H. Chou, Y.C. Li, '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) (1995) 467–476.
[4] Y.J. Chiu, T. Berger, A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancement, IEEE Transaction on Circuits and Systems for Video Technology 9 (3) (1999) 438–450.
[5] I. Hontsch, L.J. Karam, Adaptive image coding with perceptual distortion control, IEEE Transaction on Image Processing 11 (3) (2002) 213–222.
[6] T.D. Tran, R. Safranek, A locally adaptive perceptual masking threshold model for image coding, Proceedings of International

Conferrence on Acoustics, Speech, and Signal Processing (ICASSP) 4 (1996) 1883–1886.
[7] H.Y. Tong, A.N. Venetsanopoulos, A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking, Proceedings of IEEE Internatiopnal Conference Image Processing (ICIP) (1998).
[8] X.K. Yang, W.S. Lin, Z.K. Lu, E.P. Ong, S.S. Yao, Rate control for videophone using local perceptual cues, IEEE Transaction on Circuits and Systems for Video Technology 15 (4) (2005) 496–507.
[9] W. Lin, L. Dong, P. Xue, Visual distortion gauge based on discrimination of noticeable contrast changes, IEEE Transaction on Circuits and Systems for Video Technology 15 (7) (2005) 900–909.
[10] R.B. Wolfgang, C.I. Podilchuk, E.J. Delp, Perceptual watermarks for digital images and video, Proceedings of IEEE 87 (7) (1999).
[11] X.H. Zhang, W.S. Lin, P. Xue, Improved estimation for just-noticeable visual distortion, Signal Processing 85 (4) (2005) 795–808.
[12] W. Zeng, S. Lei, Digital watermarking in a perceptually normalized domain, in: Proceedings of 33rd Annual Asilomar Conference On Signals, Systems and computers, 1999.
[13] A.B. Watson, G.Y. Yang, J.A. Solomon, J. Villasenor, Visibility of wavelet quantization noise, IEEE Transaction on Image Processing 6 (8) (1997) 1164–1175.
[14] S. Kuo, J.D. Johnston, Spatial noise shaping based on human visual sensitivity and its application to image coding, IEEE Transaction on Image Processing 11 (5) (2002) 509–517.
[15] M. Ramasubramanian, S.N. Pattanaik, D.P. Greenberg, A perceptual based physical error metric for realistic image synthesis, Proceedings of Computer Graphics (SIGGRAPH'99) Conference 4 (33) (1999) 73–82.
[16] R.J. Safranek, J.D. Johnston, A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression, Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1989) 1945–1948.
[17] R.J. Safranek, JPEG compliant encoder using perceptually based quantization, in: Proceedings of SPIE Conference on Human Vision, Visual Processing, and Digital Display V, 1994, pp. 117–126.
[18] H.A. Peterson, A.J. Ahumuda, A.B. Watson, An improved detection model for DCT coefficient quantization, in: Human Vision, Visual Processing, and Digital Display VI, 1993, pp. 191–201.
[19] G.E. Legge, J.M. Foley, Contrast masking in human vision, Journal of the Optical Society of America 70 (1980) 1458–1471.
[20] G.E. Legge, A power law for contrast discrimination, Vision Research 21 (1981) 457–467.
[21] F.L. van Nes, M.A. Bouman, Spatial modulation transfer in the human eye, Journal of the Optical Society of America 57 (1967) 401–406.
[22] M.J. Nadenau, Integration of Human Color Vision Models into High Quality Image Compression, PhD's Thesis, Lausanne, EPFL (2000).
[23] A.N. Netravali, B.G. Haskell, Digital Pictures: Representation and Compression, Plenum, New York, 1988.
[24] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models of human perception, in: Proc. IEEE, 1993, pp. 1385–1422.
[25] B. Girod, What's wrong with mean-squared error? in: A.B. Watson (Ed.), Digital Images and Human Vision, The MIT Press, 1993, pp. 207–220.
[26] M.P. Eckert, A.P. Bradley, Perceptual quality metrics applied to still image compression, Signal Processing 70 (1998) 177–200.
[27] Sarnoff Corporation, Measuring Image Quality: Sarnoff's JNDmetrix Technology, Sarnoff JNDmetrix Technology Overview (2002).
[28] B. Li, G.W. Meyer, R.V. Klassen, A comparison of two image quality models, Proceedings of SPIE, Human vision and Electronic Imaging III 3299 (1998) 98–109.
[29] S.A. Klein, A.D. Silverstein, T. Carney, Relevance of human vision to JPEG-DCT compression, Proceedings of SPIE, Human Vision, Visual Processing, and Digital Display III 1666 (1992) 200–215.
[30] J. Park, J.M. Jo, J. Jeong, Some adaptive quantizers for HDTV image compression, in: L. Stenger et al. (Ed.), Signal Processing of HDTV V (1994).
[31] B.J. Walter, S.N. Pattanaik, D.P. Greenberg, Using perceptual texture masking for efficient image synthesis, Computer Graphics Forum 21 (3) (2003) 393–400.
[32] X. Yang, W. Lin, Z. Lu, E. Ong, S. Yao, Just noticeable distortion model and its applications in video coding, Signal Processing: Image Communication 20 (7) (2005) 662–680.
[33] J. Lubin, A visual discrimination model for imaging system design and evaluation, in: E. Peli (Ed.), Vision Models for Target Detection and Recognition, World Scientific, 1995, pp. 245–283.
[34] R. Dumont, F. Pellacini, J.A. Ferwerda, Perceptually-driven decision theory for interactive realistic rendering, ACM Transaction on Graphics 22 (2) (2003) 152–181.