JUST-NOTICEABLE-DISTORTION PROFILE WITH NONLINEAR ADDITIVITY MODEL FOR PERCEPTUAL MASKING IN COLOR IMAGES

X. K. Yang, W. S. Lin, Zhongkang Lu, E. P. Ong and Susu Yao

Laboratories for Information Technology
Agency for Science, Technology and Research, Singapore 119613.
Email: xkyang@ieee.org, {xkyang,wslin,zklu,epong,ssyao}@lit.a-star.edu.sg

ABSTRACT

We propose a new spatial just noticeable distortion (JND) profile for color image processing. The JND threshold depends on various masking effects underlying existing in human vision system (HVS). How to efficiently integrate different masking effects together is the key issue of modelling JND profile. Based on recent vision research results, we model the masking effects in different stimulus dimensions as a nonlinear additivity model for masking (NAMM). It applies to all color components and accounts for the compound impact of luminance masking and texture masking to estimate the just noticeable distortion (JND) threshold in images. In our PSNR and subjective comparison to the related work, the proposed NAMM scheme provides a more accurate JND profile towards the actual JND bound in the HVS.

1. INTRODUCTION

Human eyes cannot sense any changes in an image that are below the just noticeable distortion (JND) threshold [1]. Methods have been proposed for JND profiling in subbands [2] [3] and images [4] [5], and spatial JND profile is particularly useful for perceptually optimized subband coding, motion estimation, conditional replenish and data hiding. The JND threshold derives from various masking effects in human vision system (HVS), and there are primarily two factors affecting the spatial JND: luminance masking and texture masking. Since they co-exist in most images, how to efficiently integrate these two masking effects is an important issue in obtaining accurate JND profile.

Although useful results had been reported in [4] for image-domain spatial JND profile, there are two drawbacks in the approach: only the JND threshold for the luminance component in an image is considered; integration of different spatial masking effects is simplified as seeking the maximum value between the visibility thresholds for texture masking and luminance masking. We believe that masking effect in chrominance channels should be also exploited to improve compression or watermarking performance; furthermore, combinative effect of multiple maskings should take some form of addition (not linear addition though) of individual factors, in analogy with the saliency effect from different stimuli in the recent vision research results [6]. Based upon these two ideas, this paper proposes a nonlinear additivity model for masking (NAMM) to enable more accurate JND calculation towards the actual JND bound existing in the HVS.

2. NAMM-BASED JND PROFILE WITH COLOR IMAGES

The proper choice of a suitable color space for HVS-based manipulations is important. YC_BC_R color space is used in our scheme because it not only facilitates good compression, but also performs well in prediction of visual distortion between two colors [7]. In addition, it has been adopted by the prevalent compression standards, such as JPEG, MPEG and H.26X.

Let $I_\theta(x, y)$ denote the intensity component for a pixel located at (x, y) for a color channel θ, and θ=Y, C_b, C_r. Our objective is to determine JND_θ(x,y), the JND value for the said component, according to the HVS characteristics.

2.1. Nonlinear-additivity model for masking

Simultaneous existence of multiple masking factors in a neighborhood makes targets (e.g., coding artifacts in a decoded image) more difficult to be noticed, when compared with the case of one source of masking alone. However, the masking effects do not add linearly. Thus, the combination of various masking effects can be mathematically described in the following *nonlinear-additivity model for masking* (NAMM):

$$T = \sum_{i=1}^{N} T^i - \sum_{i=1}^{N} \sum_{j=i+1}^{N} C^{i,j} \cdot \gamma(T^i, T^j) \qquad (1)$$

where T is the visibility threshold due to the overall masking effect. T^i is the visibility threshold due to the individual

masking stimulus i. $C^{i,j}$ is the gain reduction factor due to overlapping between two masking stimuli. $\gamma(\cdot,\cdot)$ is an appropriate nonlinear function.

2.2. JND Profile via NAMM with Color Images

In the spatial domain, there are primarily two factors affecting the JND of each pixel. One is luminance masking, i.e., human visual perception is sensitive to luminance contrast rather than the absolute luminance value. The other is texture masking, i.e., the reduction of visibility of stimuli is caused by the increase in the texture non-uniformity of the background. As an approximation of the nonlinear model described in Equation (1), the overall spatial domain JND, $JND_\theta(x,y)$, with combined texture and luminance masking is given by the addition of the individual masking components minus the proportion of overlapping effects in both maskings:

$$JND_\theta(x,y) = T^l(x,y) + T^t_\theta(x,y) - C^{lt}_\theta \cdot \min\{T^l(x,y), T^t_\theta(x,y)\} \quad (2)$$

where $T^l(x,y)$ and $T^t_\theta(x,y)$ are the visibility thresholds due to luminance masking and texture masking for a color channel, respectively; the computation of $T^l(x,y)$ follows the relevant methodology proposed in [4]; the computation of $T^t_\theta(x,y)$ will be addressed in the following subsection; and C^{lt}_θ accounts for the overlapping effect in masking for a color channel θ. C^{lt}_θ is quantitated as C^{lt}_Y = 0.3, $C^{lt}_{C_b}$ = 0.25 and $C^{lt}_{C_r}$ = 0.2, and our extensive subjective testing demonstrates that this parameter setting maintains good image quality while results in good performance for HVS-based compression.

2.3. Edge-adaptive visibility threshold of textual masking

The computation of $T^t_\theta(x,y)$ uses a new method given by Equation (3). Comparing with the method in [4], our new method not only considers all three channels in YC_bC_r color space, but also takes the importance of object edges into account.

In contrast to pixels, edges are directly related to image content in that they demarcate object boundaries, surface crease, reflectance change and other significant visual events. Moreover, there is a substantial body of literature attesting to the importance of edges to primate perception (e.g., [8][9]). The distortion at edge introduced by image processing is more visible than that in other texture region. To shape noise beyond edges, we propose the edge-adaptive visibility threshold of textual masking as follows:

$$T^t_\theta(x,y) = G_\theta(x,y) \cdot \beta_\theta \cdot W_\theta(x,y) \quad (3)$$

where $G_\theta(x,y)$ denotes the maximal weigthed average of gradients around the pixel at (x,y) and its computation is described in [10]. β_θ is the empirical parameter for each color channel. β_θ = 0.117, 0.65 and 0.45 is determined for $\theta = Y$, Cb and Cr respectively according to our perceptual experiment according to our extensive subjective testing under the viewing condition addressed in the following Subsection 3.1. $W_\theta(x,y)$ is an edge-adaptive weight of the pixel at (x,y), and its corresponding matrix $\mathbf{W}_\theta$ is computed by edge detection followed with a Gaussian low-pass filter:

$$\mathbf{W}_\theta = \mathbf{E}_\theta * \mathbf{h} \quad (4)$$

where $\mathbf{E}_\theta$ is the edge matrix of the original video frame for each color component, with element values of ε_1 and ε_0 for edge and non-edge pixel respectively. $\varepsilon_1 = 0.1$ and $\varepsilon_0 = 1$ is used in this paper. For edge detection, the Canny method[11] with the sensitivity thresholds of 0.5, 0.175 and 0.175 are used for Y, C_b and Cr respectively in this paper. $\mathbf{h}$ is a $k \times k$ Gaussian low pass filter with standard deviation σ. $k = 7$ and $\sigma = 0.8$ yields good performance for the images with common size, such as 512×512, in the same subjective testing condition as that of determining β_θ.

3. EXPERIMENTAL RESULTS

The proposed model can be tested by comparing an image $I_\theta(x,y)$ with its variation, $I^{JND}_\theta(x,y)$, which is formed via:

$$I^{JND}_\theta(x,y) = I_\theta(x,y) + s_{rand}(x,y,\theta) \cdot JND_\theta(x,y) \quad (5)$$

where $s_{rand}(x,y,\theta)$ takes value of either +1 or -1 at random regarding x, y and θ, and this is to avoid fixed artifact patterns introduced to the image. If JND_θ(x,y) determined in (1) is close to the JND bound in the HVS, it should take the largest possible value while perceptual distortion in the image constructed by (5) is minimized.

For comparison purpose, random noise is added to the same image:

$$I^{Non-JND}_\theta(x,y) = I_\theta(x,y) + \alpha \cdot s_{rand}(x,y,\theta) \cdot rand(x,y,\theta) \quad (6)$$

where $rand(x,y,\theta)$ takes a random value in (0.0, 1.0) and α is a control factor to maintain a similar amount of error energy (therefore similar PSNR) between I^{JND}_θ(x,y) and $I^{Non-JND}_\theta$(x,y).

The image *Lenna* (Figure 1(a)) is processed by (5) and (6), and the results are shown in Figure 1(b) and (c), respectively. With similar PSNR, as can be seen, Figure 1(b) has much less visual distortion than Figure 1(c). In fact, human viewers can hardly discern any difference between the original image and the image modified by the proposed JND profile, because the proposed JND profile effectively

Table 1: Comparison of Average PSNR (dB) for 30 Images

Method	Average PSNR (dB)
Noise Injection with NAMM	29.04
Random Noise Injection	29.13
Noise Injection with Grey level NAMM	30.37
Noise Injection with Model in [4]	31.44

shapes the added noise to perceptually insensitive regions. Although the whole full-color image is processed, only the central part of the image is displayed in Figure 1 for clearer comparison of the printed versions. Figure 1(d) and (e) show the results when only JND_Y(x,y) is incorporated in (2) and when the JND derived in [4] for grey level images is used, respectively, and human viewers rate their visual quality to be similar (supported by the subjective testing in the following subsection); this therefore indicates that the proposed model allows more data redundancy (of 1.35 dB) for a same picture quality level in grey level image.

The same experiments have been performed for 30 standard testing color images, most of which can be obtained from http://www.ipl.rpi.edu/resource/stills/kodak.html. The results are found to be consistent and the average PSNR is shown in Table 1.

3.1. The comparative subjective quality assessment

To evaluate the performance of the proposed JND profile via NAMM with color images in comparison to that in [4], the comparative subjective quality assessment of the noised images was performed. For fairness of comparison between the proposed JND for color image and the JND only for grey-scale image in [4], the noises are injected into Y components in the case of the JND in [4] while the noises are injected into all YC_bC_r components in the case of the proposed JND.

The subjective assessment setup is similar to that in [12]. The associated subjective visual quality assessment was performed in a dark room by 8 subjects (five of them are with average image processing knowledge and the rest are naive), using a 21" EIZO T965 professional color monitor in image model with resolution on 1600×1200. The viewing distance is approximately six times of the image height. On each trial of the experiment, subjects viewed two images of the same scene (see Figure 2). Subjects were then given time to vote on the comparative quality of two images [13], using the continuous quality comparison scale shown in Table 2. Subjects were not allowed to respond until after they had viewed the images for at least two seconds. The order of presentation of the 30 possible trials (the above-mentioned testing images) was randomized in each session. On each trial, which image was presented on which side of the display was also chosen randomly.

Table 2: Comparison scale for subjective quality evaluation

-3	the left one much worse than the right one
-2	the left one worse than the right one
-1	the left one sightly worse than the right one
0	the same
+1	the left one sightly better than the right one
+2	the left one better than the right one
+3	the left one much better than the right one

Table 3: The comparative subjective quality ("+": the proposed JND better, "-": the JND in [4] better.)

Subject index	Mean	Standard Deviation
1	-0.173	0.888
2	-0.056	0.348
3	-0.130	1.548
4	-0.280	1.160
5	+0.221	1.370
6	-0.019	0.590
7	-0.319	1.005
8	+0.312	2.361
Average	-0.055	1.159

The results which show the comparative subjective quality are listed in Table 3, where Mean and Standard Deviation are computed over all 30 possible trials. From Table 3, we can see that the overall comparative subjective quality tends to a near zero mean of -0.055 with its associated standard deviation of 1.159. Thus, the subjective quality for the images noised with the proposed JND profile is very close to that noised with the JND profile in [4].

4. CONCLUSION

The proposed NAMM scheme provides a more accurate JND profile towards the actual JND bound in the HVS, since it is capable of exploiting larger JND values without jeopardizing the visual quality. As indicated in Table 1 (the first and last rows), it is expected to outperform the approach in [4] by more than 2 dB (in terms of PSNR) of permitted data redundancy on an average for a same level of visual quality, because of the due consideration for the compound masking effect and full color impacts. Consequently, it enables better visual compression and watermarking.

5. REFERENCES

[1] N. S. Jayant, J. D. Johnston, and R. J. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, pp. 1385–1422, 1993.

[2] A. B. Watson, Gloria Y. Yang, Joshua A. Solomon, and John Villasenor, "Visibility of wavelet quantization

noise," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1164–1175, August 1997.

[3] Ingo Hntsch and Lina J. Karam, "Locally adaptive perceptual image coding," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1472–1483, Sept. 2000.

[4] Chun-Hsien Chou and Yun-Chin Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, 1995.

[5] Y. J. Chiu and T. Berger, "A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 3, pp. 438–450, April 1999.

[6] H. C. Nothdurft, "Salience from feature contrast: additivity across dimensions," *Vision Research*, vol. 40, pp. 1183–1201, 2000.

[7] M. J. Nadenau, *Integration of Human Color Vision Models into High Quality Image Compression*, Ph.D. thesis, EPFL, Switherland, 2000.

[8] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freemanand Co., 1982.

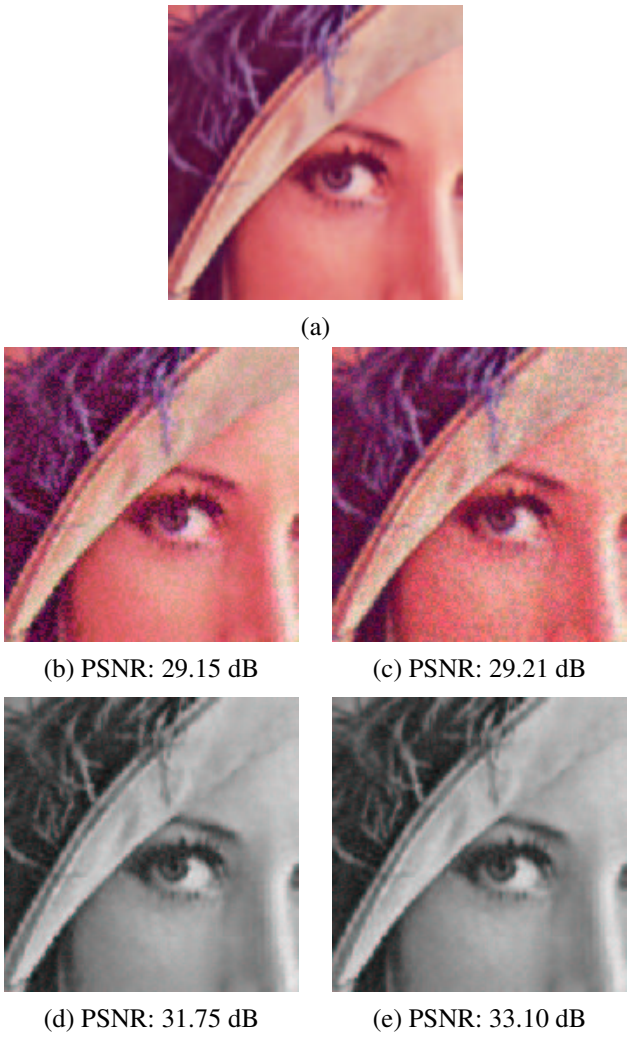
[9] J. H. Elder and R. M. Goldberg, "Local scale control for edge detection and blur estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 699–716, July 1998.

[10] Chun-Hsien Chou and Chi-Wei Chen, "A perceptually optimized 3-d subband image codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 143–156, 1996.

[11] Canny John, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[12] P. Longere, X. Zhang, P. B. Delahunt, and D. H. Brainaro, "Perceptual assessment of demosaicing algorithm performance," *Proc. IEEE*, vol. 90, no. 1, pp. 123–132, Jan. 2002.

[13] W. Osberger, *Perceptual Vision Models for Picture Quality Assessment and Compression Applications*, Ph.D. thesis, Queensland University of Technology,Australia, March 1999.



(a)

(b) PSNR: 29.15 dB

(c) PSNR: 29.21 dB

(d) PSNR: 31.75 dB

(e) PSNR: 33.10 dB

Figure 1: Tests and Comparison for the Proposed NAMM: (a) Original Image; (b) Noise Injection with NAMM; (c) Random Noise Injection; (d) Noise Injection with Grey-level NAMM; (e) Noise Injection with the Model in [4]

Figure 2: Experiment setup for the comparative subjective quality assessment