ENHANCED JUST NOTICEABLE DIFFERENCE (JND) ESTIMATION WITH IMAGE DECOMPOSITION FOR SEPARATING EDGE AND TEXTURED REGIONS

Anmin Liu, Weisi Lin, Fan Zhang, Manoranjan Paul
{liua0002, wslin, m_paul}@ntu.edu.sg, zhangfan.hust@gmail.com
Nanyang Technological University, Singapore, 639798

ABSTRACT

Contrast masking (CM) on edge and textured regions have to be distinguished since distortions on edge regions are easier to be noticed than that on textured regions. Therefore, how to efficiently estimate the CM on edge and textured regions of an image is a key issue for accurate JND (Just Noticeable Difference) estimation. An enhanced image domain JND estimator is devised in this paper with new model for CM. We use the total variation method to obtain a structural image (which contains edge information) and a textural image (which contains texture information) from the input image, and then evaluate the CM for the two images separately rather than the whole image, and hence edge and texture are better distinguished and the under-estimation of JND on textured regions can be effectively avoided. Experimental results of subjective viewing confirm that the proposed model is capable of determining more accurate visibility thresholds.

Index Terms— JND, visibility threshold, contrast masking, image decomposition.

1. INTRODUCTION

It is well known that the human visual system (HVS) cannot sense all changes in an image due to its underlying physiological and psychological mechanisms. Just noticeable difference (JND), which accounts for the maximum sensory distortion that the HVS does not perceive (e.g., for 75% of the observers [1]), can be used to facilitate effective image/video compression, quality evaluation, watermarking, etc. The JND values can be estimated for the image domain or sub-bands (e.g., DCT, DWT). Sub-band based JND models only estimate the perceptual threshold in each sub-band, while the image domain JND models can provide a clear view of the threshold map for image pixels. In general, sub-band JNDs are popular for perceptual image/video compression, while pixel-based JND models are often used in motion estimation, visual quality evaluation, and video replenishment.

Luminance adaptation (LA) and contrast masking (CM) are the two major considerations in image domain JND. LA refers to the masking effect of the HVS toward the background luminance and CM accounts for the masking effect toward the spatial activities in the neighborhood. Image domain CM is of our interest in this paper. In Chou *et al.*'s model [2], CM is estimated with the maximum signal from four edge detectors with 45 degrees apart. In Chiu *et al.*'s model [3], CM is determined by the maximum grey-level difference of the central pixel and its neighbors. Yang *et al.* [4] modified Chou *et al.*'s model by detected edge pixels (by Canny detector [5] with a threshold) and suppressed the CM on the detected regions.

CM on textured region should be distinguished with that on edge region since distortion around edge is easier to be noticed than that on textured region due to the fact that edge structure attracts more attention from the typical HVS [4, 6-7]. In Chou *et al.*'s models, CM on textured region is under-estimated since they treated texture as the same as edge; In general, Yang *et al.*'s model outperforms Chou *et al.*'s models since edge and texture are distinguished by edge detection. However, for some images (especially the image with lots of texture), the performance of Yang *et al.*'s model can be worse than that of the Chou *et al.*'s model due to the inaccuracy of edge detector. One such example is shown in last row of Fig. 2.

To solve the above mentioned drawback in the existing models, an enhanced image domain JND estimator is devised in this paper with a new model for CM. CM on edge and textured region is to be estimated separately, and the separation can be done by image structure-texture decomposition based upon the total variation (TV) model. Our major contributions in this paper are twofold: (1) edge and texture are better distinguished by using the said image decomposition model; (2) we devised a more accurate CM model, therefore, a better JND model to mimic the masking effect of the HVS.

The rest of the paper is organized as follows: in Section 2, we describe the model of image decomposition; in Section 3, the detail of the proposed JND estimator is presented; the subjective test results and the results of visual quality evaluation based on the JND model are given in Section 4; finally, conclusions are drawn in Section 5.

2. IMAGE DECOMPOSITION MODEL

The task of separating texture from non-texture parts in images is of great interest and can be used in image compression, image denoising, image inpainting, image registration, etc. It can be assumed that $f = u + v$, where f is the observed image, u is the structural image and v is the textural image. Here the basic idea is to define appropriate features (norms) for structural (cartoon) component and textural component.

Vese *et al.* [8] proposed a TV based edge preserving method for image decomposition. The two major ingredients of this method are: the total variation minimization [9] and the space of oscillating functions introduced by Yves Meyer [10]. In [9], u is regarded as the image formed by homogeneous regions (i.e., piecewise smooth) and sharp edges along the contour. In [10], Meyer firstly characterized the texture as fine scale details, usually with some periodicity and oscillatory nature.

The standard TV model separates image u from an observed image f by solving the following [11]:

Fig.1 Block diagram of the JND models. (a) Proposed model, (b) Yang *et al.*'s model [4].

$$\min_u \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} |f - u| \quad (1)$$

where Ω is defined in R^2 , $f: \Omega \rightarrow R$ is the observed image, $f - u (=v)$ is the textural component of f , and $|\nabla u|$ is the first norm of the gradient vector $\nabla u \in R^2$.

3. PROPOSED JND ESTIMATION

The proposed JND estimator is shown in Fig. 1 (a), we also shown the latest image domain JND model (i.e., Yang *et al.*'s model [4]) in Fig. 1 (b) as a comparison. In Fig. 1, the parts enclosed with dash lines represent the modules for CM. As shown in Fig. 1, we updated the CM module in [4] with our proposed CM estimation scheme which is based on image decomposition model. The problem mentioned in Section 1 (under-estimation CM on textured region) in [4] can be effectively alleviated since we have already separated the texture information prior to Canny edge detection and no texture can be detected as edge. The other modules are the same as that in [4] and will be described later.

Following the notations in the previous section, u and v represent the structural and textural image of the input image, respectively, and are obtained by solving (1). One solution to this problem can be found in [11], which is adopted in our work.

3.1. Contrast masking (CM) model

CM is an important phenomenon in the HVS perception and is referred to as the reduction in the visibility of one visual component at the presence of another one [12, 1]. To be more specific, CM is due to spatial contrast/variation and there is little CM effect on smooth region. Therefore, we mainly consider the CM on edge region and textured region. As can be seen in Fig. 1(a), smooth region in u is eliminated via Canny operator.

In many approaches, CM on edge region is determined as a function of the luminance intensity gradient, and perturbation is increased until it becomes just discernible [2]. The relationship can be obtained by subjective tests where a perturbed edge and an unperturbed edge are shown to subjects (displayed on a monitor). Denote CM^e as CM on edge region, and we have [2]:

$$CM^e = \beta^e \cdot G_{le} \quad (2)$$

where G_{le} is the luminance intensity gradient. The parameter β^e is the slope of the linear function (2) which depends on the viewing distance, and increases slightly as the background luminance increases.

Fig.2. Contrast masking (CM) for different JND models (higher brightness means a larger masking value). (a1) original Lena image, (b1) CM in model [2], (c1) CM in model [3], (d1) Contrast in the proposed model; (a2)-(d2) are the corresponding results for Texmos3 image.

Fig.3. JND for Lena image (higher brightness means a larger JND value).

We propose the enhanced CM model as (3) below, considering: (1) there is little CM effect on smooth region; (2) u and v contain the edge and the texture information of the original image f , respectively [8-11]; (3) u contains the same edge information with or without Canny operator.

$$CM(f) = CM(u) + CM(v) \quad (3a)$$

and

$$\begin{cases} CM(u) = C_s(u) \cdot \beta \cdot W_e \\ CM(v) = C_s(v) \cdot \beta \cdot W_t \end{cases} \quad (3b)$$

where the parameter $C_s(\cdot)$ and β has the same meaning as G_{le} and β^e in (2), respectively; W_e and W_t are used to distinguish CM on edge and textured region; in the current work, we choose $W_e = 1$ and $W_t = 2$, since the CM on textured region should lead to a higher JND value than that on edge region for a same extent of spatial contrast($C_s(u)$ and $C_s(v)$).

We have shown the profile of CM for different JND models in Fig. 2. From Fig. 2 (b1) we can see that the CM on edge and textured region have nearly the same value (similar brightness) in Chou *et al.*'s model since it fails to distinguish between edge and texture; From Fig. 2 (c2) we can see that CM on textured region is heavily under-estimated (nearly zero brightness) in Yang *et al.*'s model since every pixel in Texmos3 (an image mainly with texture) tends to be detected as edge pixel when detected by using Canny operator. As can be seen via the comparison, our new model (refer to Fig. 2 (d1) and Fig. 2 (d2)) yields much better CM estimation, i.e., smaller CM on edge region and reasonable CM on textured region.

3.2. The overall JND model

The proposed JND profile ($JND(f)$) estimator for image f can be described mathematically as:

$$JND(f) = LA(f) + CM(f) - C^{lt} \times \min\{LA(f), CM(f)\} \quad (4)$$

and

$$LA(i) = \begin{cases} 17(1 - f(i)/127) + 3, & \text{if } f(i) \leq 127 \\ 3/128 \times (f(i) - 127) + 3, & \text{otherwise} \end{cases} \quad (5)$$

note that (4) and (5) are the same as in [4] since our aim in this paper is to improve $CM(\cdot)$ with (3), and they are the NAMM (nonlinear additivity model for masking) module and the luminance adaptation module as shown in Fig. 1; $LA(i)$ is the LA factor for the i^{th} pixel in the image f ; $f(i)$ is the intensity value of the i^{th} pixel; C^{lt} is the gain reduction factor in (4) due to the overlapping between $LA(\cdot)$ and $CM(\cdot)$. The bigger value of C^{lt} takes, the more significant overlapping effect it represents. In our work, C^{lt} is set as 0.3, the same as in [4]. In Fig. 3, we have shown the overall JND profile for Lena image generated by the proposed model.

4. EXPERIMENTAL RESULTS

In this section, we will evaluate the overall performance of the proposed image domain JND model and compare it with the other relevant models (i.e., Chou *et al.*'s model in [2] and Yang *et al.*'s model in [4]). Ten images (with different visual content and spatial complexity) were chosen for testing, all of them are of 512×512 and from USC-SIPI Image Database [13], as listed in the first column of Table II. Color images were converted to gray level by using Matlab function [rgb2gray].

A better JND model should be able to guide shaping more noise into an image at a certain level of resultant perceived images quality [14]. To evaluate the performance of JND models, noise is added, regulated by the yielded JND profiles. The noise associated with the JND threshold is randomly added or subtracted on each pixel of the original image.

$$f^{JND}(i) = f(i) + S^{random} \cdot JND(i) \quad (6)$$

where f and f^{JND} are the original and the noise contaminated image, respectively; S^{random} takes the value of +1 or -1 randomly, to avoid introducing a fixed pattern of changes. At the same perceptual quality, the higher the injected-noise energy (measured by PSNR–peak signal to noise ratio), the more accurate is the JND model. A better JND model can inject more noise at a given quality level.

For a comprehensive evaluation, the subjective viewing tests were conducted based on “Adjectival categorical judgement methods” recommended by ITU-R BT.500-11 standard [15]. In this experiment, the viewing equipment is a HP L1906 LCD display (contrast ratio is 500:1, maximum resolution is 1280×1024, and maximum brightness is 270cd/m²). The viewing distance is four times of the image height (about 60cm). In each test, two images of the same scene (i.e., an image processed by the proposed model and that processed by one of the other models) were juxtaposed on the screen. Twenty one subjects (their eyesight was either normal or had been corrected to be normal with spectacles) were asked to give quantitative scores for all the image pairs, using the continuous quality comparison scale shown in Table I. The order of the presentation of the image pairs was randomized in each session; the image processed by using the proposed model appeared randomly on the left- or right-hand side of the screen for each image pair to avoid the possible bias.

Table II (the left part) shows the mean of the subjective scores. A negative (positive) subjective score indicates that the image with the proposed model has better (worse) perceptual quality than that with the other model, and its magnitude

Table I
Scores for subjective quality evaluation

Subjective score	Description
-3	The right one 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 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 II

PSNR for different JND models and the subjective quality evaluation results (the proposed model against each of those in [2] and [4]) for 10 images with different visual content and spatial complexity

Image name	Subjective score		PSNR (dB)		
	Model [2]	Model [4]	Proposed	Model [2]	Model [4]
Airplane	0.048	0.286	32.65	32.74	34.80
Barbara	0.143	0.333	29.93	31.64	31.35
Boat	0.429	-0.095	31.37	32.55	32.62
Couple	0.190	-0.048	32.43	33.57	32.71
Gold	-0.619	0.000	31.16	32.32	31.97
Lena	-0.286	-0.048	31.89	32.80	32.72
Mandrill	-0.190	-0.333	29.77	32.99	32.54
Peppers	-0.190	0.000	30.12	30.78	30.79
Splash	0.190	-0.143	30.96	31.43	31.35
Tank	0.143	-0.048	34.22	35.22	34.88
average	-0.014	-0.010	31.45	32.60	32.58
average extra redundancy (in dB) of the proposed model			1.15	1.13	

represents the extent of quality improvement (degradation). With the reference to Table I, the range of magnitudes lies between 0 and 3. From this part of Table II, we can see that the mean overall subjective quality is -0.014 and -0.010, respectively. Thus, the subjective quality for the images noised with the proposed JND model is very close (in fact, slightly better) when compared to that with other models.

The PSNR of noised images is then used to measure the amount of noise being added by the JND models under comparison then. At a same level of perceived quality (as the cases demonstrated in Table II), a better model achieves higher JND thresholds and results in lower PSNR. Table II (the right part) shows the PSNRs with the three JND profiles. The proposed model is better than existing relevant JND models, with the evidence of an average additional PSNR redundancy of 1.15 dB and 1.13 dB from the model in [2] and the one in [4], respectively, without jeopardy of visual quality (in conjunction with the results in the left part of Table II).

5. CONCLUSION

In this paper, an enhanced JND estimator is proposed. The total variation based image decomposition model is used to improve the accuracy of contrast masking (CM) evaluation, as an attempt

for overcoming the shortcomings of the existing models. Extensive subjective tests confirm that the proposed scheme provides a more accurate JND profile. At the same level of perceptual quality, on average it allows more than 1.0 dB of extra data redundancy when compared with the most related existing methods.

The major contribution of this paper is a new JND estimator based upon a new and more accurate model for CM evaluation. With this, we are able to evaluate CM on edge and textured regions on the structural and textural components, respectively. Therefore, under-estimation of the visibility threshold on textured region is avoided, and a better and more comprehensive mimic to the masking effect of the HVS is achieved.

6. REFERENCES

- [1] X. Zhang, W. Lin, P. Xue, "Improved estimation for justnoticeable visual distortion," *Signal Processing*, vol. 85(4), pp. 795–808, Apr. 2005.
- [2] C. Chou and C. Chen, "A perceptually optimized 3-D subband image codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6(2), pp. 143–156, 1996.
- [3] Y. Chiu and T. Berger, "A software-only video codec using pixel wise conditional differential replenishment and perceptual enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9(3), pp. 438–450, 1999.
- [4] 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), pp. 662–680, 2005.
- [5] J. Canny, "A Computational Approach to Edge. Detection," *IEEE Trans. Pattern Analysis and Machine. Intelligence*, vol. 8(6), pp.679–698, 1986.
- [6] W. Lin, L. Dong, P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15(7), 900–909, 2005.
- [7] M.P. Eckert, A.P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70(3), 177–200, 1998.
- [8] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Science Computer*, 19, 553–577, 2003.
- [9] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* 60, 259–268, 1992.
- [10] Y. Meyer, "Oscillating Patterns in Image Processing and Nonlinear Evolution Equations," *University Lecture Series*, Vol. 22, Amer. Math. Soc., 2002.
- [11] D. Goldfarb, W. Yin, "Parametric Maximum Flow Algorithms for Fast Total Variation Minimization," *Rice CAAM Report TR07-09*, 2007.
- [12] G. Legge, J. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458–1471, 1980.
- [13] Image Database, <http://sipi.usc.edu/services/database>.
- [14] Z. Wei, K. Ngan, "Spatio-temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19(3), pp. 337–346, March 2009.
- [15] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.