Enhanced Just Noticeable Difference Model with Visual Regularity Consideration

Jinjian Wu[1], Guangming Shi[1], Weisi Lin[2], and C.C. Jay Kuo[3]

Abstract—**Just noticeable difference (JND) reveals the visibility of our human visual system (HVS), below which changes cannot be perceived by the human. Though dozens of JND estimation models have been introduced during the past decade, how to accurately estimate the JND thresholds for different content regions (e.g., edge and texture region) is still an open problem. Research on cognitive science indicates that the HVS is adaptive to extract the visual regularities from an input scene for content perception and understanding. Thus, we analyze the effect of content regularity on visual sensitivity, and suggest that the visual regularity is another important factor that determines the JND threshold. According to the orientation distributions of local regions, the content regularities are firstly calculated. Then, by considering the effect from content regularity, luminance adaptation, and contrast masking, a novel JND model is proposed. Experimental results demonstrate that the proposed model can effectively estimate the JND thresholds of regions with different visual contents.**

Index Terms—**Just Noticeable Difference (JND), Human Visual System (HVS), Content Regularity, Orientation Distribution**

I. INTRODUCTION

It is well known that the visual resolution of the human visual system (HVS) is limited, which can only perceive the changes above certain thresholds [1]. The Just Noticeable Difference (JND) technique estimates such thresholds, and it is effective to identify visual redundancies of images [2]. Therefore, the JND technique is useful in visual signal compression [3], watermarking [4], quality measurement [5], and so on.

During the past decade, a number of JND estimation models have been proposed, which are often divided into two classes: JND models in the subband-domain [6], [7] and the pixel-domain [8], [2]. For subband-domain JND models, the luminance adaption, contrast masking, and contrast sensitivity function (CSF) are considered [6]. When for pixel-domain JND models, the luminance adaption and spatial visual masking are often considered [2].

[1]Jinjian Wu (jinjian.wu@mail.xidian.edu.cn) and Guangming Shi are with School of Electronic Engineering, Xidian University, Xi'an, Shannxi, China

[2]Weisi Lin is with School of Computer Engineering, Nanyang Technological University, Singapore

[3]C.C. Jay Kuo is with Department of Electrical Engineering, University of Southern California, Los Angeles, USA

*This work was supported by the National Natural Science Foundation of China (Nos. 61401325, 61472301, 31300473), the Ph.D. Programs Foundation of Ministry of Education of China (20130203130001), the Fundamental Research Funds for The Central Universities (No. JB140227), the International cooperation project of Shaanxi science and technology R&D program (No. 2014KW01-02), and the Shaanxi province natural science foundation of China (No. 2014JQ8296).

We focus on pixel-domain JND estimation. Chou and Li [1] analyzed the masking effect from the background luminance and edge height, and proposed an early pixel-domain JND model. Yang et al. [9] indicated that Chou and Li's model overestimates the JND threshold of the edge regions. Thus they employed the Canny edge detector to protect the edge regions for JND estimation. Liu et al. [10] indicated that the JND thresholds of the texture regions are often underestimated, and they decomposed an image into texture and non-texture regions for JND estimation. Moreover, Wu et al. [8] introduced the disorderly concealment effect for JND estimation. All of these pixel-domain models improve the estimation accuracy of JND to some extent. However, there are still a large gap between these existing JND models and the human perception.

In order to design an effective JND estimation model which performs consistent with the human perception, we turn to investigate the subjective perception/cognition property. Research on cognition science shows that the HVS is adaptive to extract the visual regularities from an input scene for content perception and understanding [11]. Moreover, signals with obvious statistical regularities are easier to be understood than those do not have regularity [12]. In other words, the HVS is more sensitive to visual contents with regularity than those without. Inspired by this, we suggest that the regularity of visual content is another important factor which determines JND thresholds.

In this paper, by considering the effect of the content regularity on visual masking, we introduce a novel JND estimation model. Research on neurophysiology confirms that the primary visual cortex (i.e., V1) detects diverse orientations for visual perception [13]. Moreover, local regions with regularities (e.g., plain and edge) have similar orientations, while local regions without obvious regularities (e.g., disorderly texture) are more likely to possess dissimilar orientations. Thus, we firstly analyze the regularity of visual content with its orientation distribution. And then, by considering the effect from the luminance adaptation, contrast masking, and content regularity, a novel JND model is built. Finally, a subjective viewing test has been set up to verify the proposed JND model. Experimental results demonstrate that the proposed JND model outperforms the existing JND models.

The rest of this paper is organized as follows. The regularities of visual content is calculated and a novel JND model is proposed in Section II. In Section III, the subjective viewing test is set up for JND model comparison. Finally, conclusions are drawn in Section IV.

II. Visual Regularity based JND Estimation

In this section, the effect of regularity on visual sensitivity is firstly analyzed. Then, by combining the novel factor with luminance adaptation and contrast masking, a novel JND model is introduced.

A. Visual Regularity

It is well known that the HVS possesses an internal generative mechanism [14], within which the human brain actively infers the input scene for outside world cognition. Moreover, during the active inference, the brain is extremely adaptive to extract the regularity of the input scene for content perception and understanding [11]. As aforementioned, regions with obvious statistical regularities are easier to be understood than those otherwise [12]. The HVS is much more sensitive to these regular regions in comparison with others, and the JND thresholds of these regions are higher.Therefore, visual regularity plays an important role in visual masking.

However, how to measure visual regularity of image content is still an open problem. Evidence on neurophysiology indicates that the primary visual cortex (i.e., V1) is with obvious orientation selection mechanism to detects diverse orientations from input objects for visual perception [13]. Inspired by this, we have designed a subjective viewing test on a large number of images with different contents. During the subjective viewing experiment, we have observed that local regions with regularities have similar orientations, while local regions without obvious regularities are more likely to possess dissimilar orientations. We use the orientation distribution of a local region for its visual regularity computation.

The orientation is firstly calculated for regularity. In image domain, the orientation of each pixel is calculated as the gradient orientation,

$$\mathcal{O}(x,y) = \arctan \frac{G_v(x,y)}{G_h(x,y)}, \tag{1}$$

where $\mathcal{O}(x,y)$ is the orientation of pixel $I(x,y)$, G_h is the gradient magnitudes alone the horizontal direction and G_v is that alone vertical direction. G_h and G_v are computed with the Prewitt filters,

$$G_h = I * f_h, \quad G_v = I * f_v, \tag{2}$$

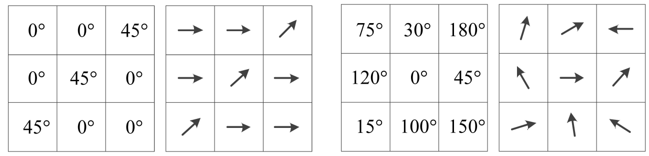
$$f_h = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix}, \tag{3}$$

$$f_v = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix}, \tag{4}$$

where $*$ denotes the convolution operation.

Then, the correlations of orientations in a local region is analyzed. Here, we directly compare the orientation difference ($\mathcal{D}$) between the central pixel (I_c) and the neighbor ($I_{c+\delta}$) for their correlation,

$$\mathcal{D}(I_c, I_{c+\delta}) = \mathcal{O}(x_c, y_c) - \mathcal{O}(x_{c+\delta}, y_{c+\delta}), \tag{5}$$



(a) Pattern with strong regularity (b) Pattern with weak regularity

Fig. 1: Two kinds of image patterns with different orientation regularities.

where $\mathcal{O}(x_c, y_c)$ is the orientation of the central pixel I_c, and $\mathcal{O}(x_{c+\delta}, y_{c+\delta})$ is that of the neighbor pixel $I_{c+\delta}$.

In order to calculate the visual regularity, we quantify the orientation differences into several bins for distribution computation. Subjective visual masking experiments demonstrate that the masking effect within two oriented gratings is strong under 12° [15]. Thus, we set the quantization step as $\mathcal{T}$=12°. Here, we define the number of individual types of the quantified orientation differences, $\mathcal{N}$, as a measure of orientation complexity. Fig. 1 shows two different patterns from different image regions. In Fig. 1 (a), there are two types of orientation (the pixels along the diagonal are with $45°$ orientation and the other ones with $0°$) for this local pattern (which is more likely to appear at the edge line region), and thus the orientation complexity $\mathcal{N}$=2. While the orientations for the pattern in Fig 1 (b) are quite different (this pattern is more likely to appear at irregular regions), and its orientation complexity $\mathcal{N}$=8. Moreover, the orientation complexity for a plain regions $\mathcal{N}$=1, and for other regions $\mathcal{N}$ can be numbers from 3-7.

B. JND Modeling

Intuitively, image regions with high luminance contrast and weak regularity present strong visual masking, while regions with low luminance contrast and strong regularity have weak visual masking. Thus, we use both regularity and luminance contrast to calculate the visual masking ($\mathcal{V}_M$),

$$\begin{aligned} \mathcal{V}_M(x,y) &= f(L_c(x,y), \mathcal{N}(x,y)) \\ &= f(L_c) f(\mathcal{N})(x,y) \\ &= \frac{1.84 \cdot L_c^{2.4}}{L_c^2 + 26^2} \cdot \frac{0.3 \cdot \mathcal{N}^{2.7}}{\mathcal{N}^2 + 1} \end{aligned} \tag{6}$$

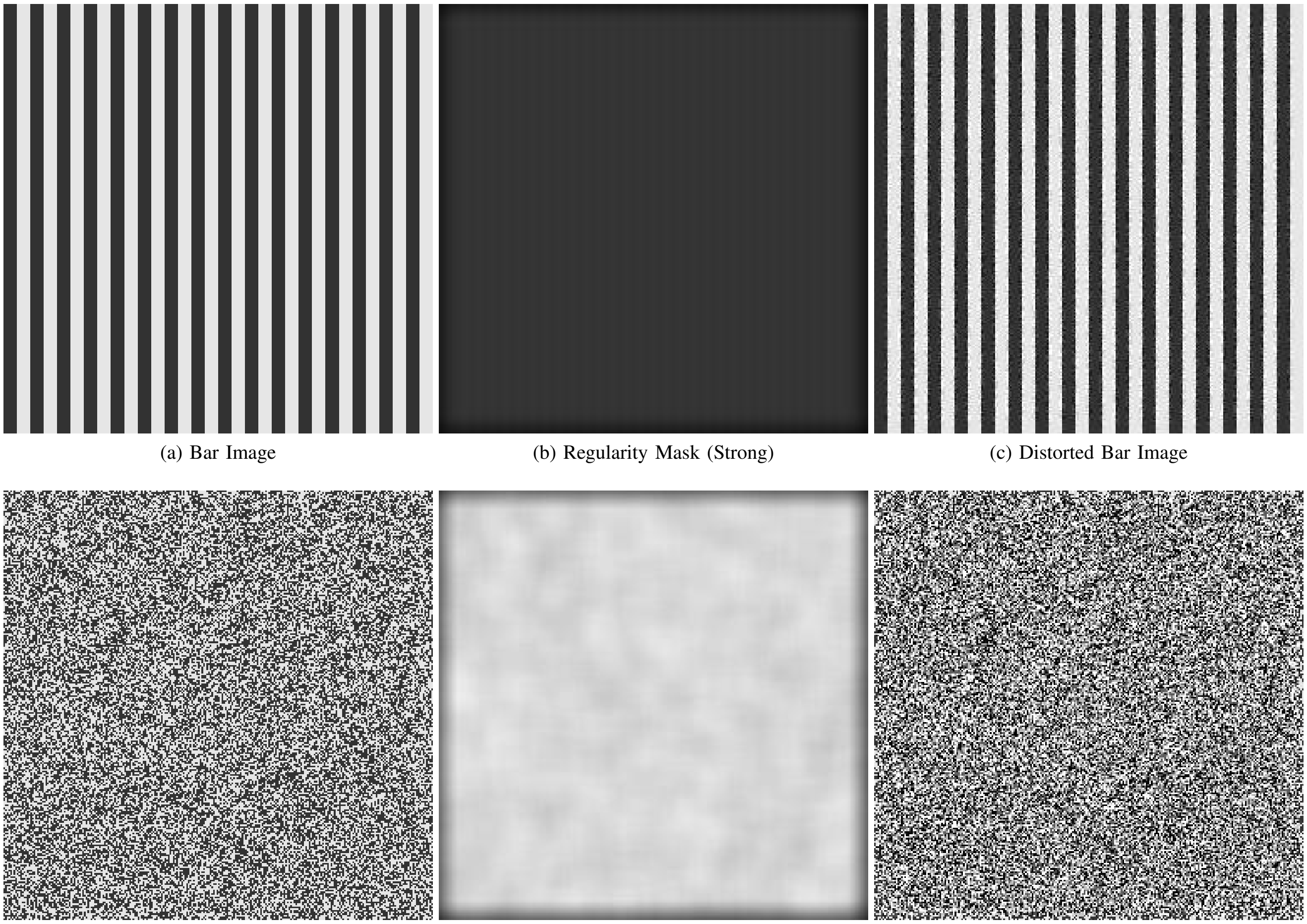
where $L_c(x,y)$ is the luminance contrast of pixel $I(x,y)$, and can be computed as,

$$L_c(x,y) = \sqrt{\frac{G_v^2(x,y) + G_h^2(x,y)}{2}}. \tag{7}$$

Another factor which determines the JND threshold is the luminance adaptation, which can be calculated as [1],

$$\mathrm{L}_A(x) = \begin{cases} 17 \times (1 - \sqrt{\frac{B(x)}{127}}) & \text{If } B(x) \le 127 \\ \frac{3}{128} \times (B(x) - 127) + 3 & \text{else} \end{cases} \tag{8}$$

where $B(x)$ is the mean luminance of a local region.



(a) Bar Image (b) Regularity Mask (Strong) (c) Distorted Bar Image

(d) Noise Image (e) Regularity Mask (Weak) (f) Distorted Noise Image

Fig. 2: Image regularity demonstration, (a) a bar image with strong regularity, and (d) a random noise image with weak regularity.

By combining Eqs. (6) and (8), the final JND threshold is acquired,

$$\begin{aligned}\mathrm{JND}(x,y) &= \mathrm{L}_A(x,y) + \mathrm{V}_M(x,y)\\ &\quad - 0.3\cdot\min\{\mathrm{L}_A(x,y), \mathrm{V}_M(x,y)\}.\end{aligned} \tag{9}$$

III. Experimental Results

In this section, the effect of the regularity on JND is firstly illustrated. Then, the proposed JND model is compared with the latest pixel-domain JND models (i.e., Yang [9], Liu [10], and Wu [8]) to demonstrate its effectiveness.

Fig. 2 shows two concept images, where (a) is a bar image and (d) is a random noise image. Though the two images are composed with same pixels (two types of pixels and each type represents 50%), their content is obviously different. Since Fig. 2 (a) presents regular structure, it is easy to extract its regularity for cognition. The computed regularity of Fig. 2 (a) is shown in Fig. 2 (b), where dark region means strong regularity (low orientation complexity) and bright region means weak regularity (high orientation complexity). As can be seen, the computed regularity value for Fig. 2 (a) is extremely small, which is consistent with the perception. While for Fig. 2 (d), the HVS can hardly extract any regularity from this image. In other words, the visual regularity of Fig. 2 (d) is quite weak. Fig. 2 (e) shows the computed regularity values of Fig. 2 (d), where almost all regions have high orientation complexity (i.e., small regularity values). According to the analysis above, the proposed regularity procedure can accurately calculate the regularity of visual contents.

Moreover, the regularity of visual content directly determines the visual sensitivity. Fig. 2 (c) is a noise contaminated image of Fig. 2 (a). Though the noise level (with MSE=150) in Fig. 2 (c) is weak, the HVS can easily sense the distortion (e.g., the distorted edge). While for Fig. 2 (f), though it is contaminated with strong noise (with MSE=1500, ten times larger than that in Fig. 2 (c)), we can hardly sense the distortion. That is because the HVS can hardly extract any regularity from Fig. 2 (d) for perception. Therefore, regularity is an important factor which determines the JND thresholds.

Fig. 3: Image set (I1-I9) for subjective viewing experiment, whose sizes are 512×512.

In order to further demonstrate the effectiveness of the proposed model, we compare the proposed JND model with three latest pixel-domain JND models, namely, Yang [9], Liu [10], and Wu [8]. A more accurate JND model should be able to distribute more noise into insensitive regions while less into the sensitive places. With a same level of noise energy, a better JND model should present better perceptual quality. Thus, we set up a subjective viewing experiment for comparison. Firstly, noise is injected into a set of test images (9 often-used images for JND comparison, as shown in Fig. 3) with the guidance of difference JND models under a same noise level,

$$\hat{I}(x, y) = I(x, y) + \beta \mathrm{rand}(x, y) \mathrm{JND}(x, y), \tag{10}$$

where $\hat{I}$ is the contaminated image distorted by JND noise, β regulates the energy of JND noise to make sure different JND models have the same level of noise, and $\mathrm{rand}(x)$ randomly takes $+1$ or -1.

Next, two JND noise contaminated images (from the proposed model and the compared model) are projected onto a scene for quality assessment. The viewing condition is set under the guidance of the ITU-R BT.500-11 standard. Twenty-six people (11 of them are experts and the left ones are naive) are invited in this subjective viewing experiment. During each test, subjects are asked to measure the quality of the two images following the rules below:

1) the left one has **much better (much worse)** quality than the right one, scored **+3 (-3)**;
2) the left one has **better (worse)** quality than the right one, scored **+2 (-2)**;
3) the left one has **slightly better (slightly worse)** quality than the right one, scored **+1 (-1)**;
4) the left one has **the same** quality with the right one, scored **0**.

TABLE I: Quality comparison between different JND noise contaminated images (Our JND model vs. three pixel-domain JND models (Yang [9], Liu [10], and Wu [8])).

Image	Our vs. Yang		Our vs. Liu		Our vs. Wu	
	Mean	Std	Mean	Std	Mean	Std
I1	0.06	0.43	0.12	1.11	0.04	0.21
I2	1.73	0.59	-0.33	0.82	1.35	0.61
I3	0.53	1.07	0.00	1.11	1.53	0.84
I4	0.06	0.75	0.53	0.87	0.79	0.78
I5	0.00	0.76	0.07	1.16	0.26	1.10
I6	0.71	0.75	1.38	0.92	0.39	1.03
I7	0.50	0.69	1.85	0.88	-0.17	0.96
I8	0.60	1.05	-0.10	0.97	0.02	1.64
I9	0.94	0.57	1.57	0.93	0.38	1.17
Mean	0.57	–	0.56	–	0.51	–

The mean subjective quality values and their corresponding standard deviations on the 9 testing images from the twenty-six subjects are listed in Table I. As can been seen, most values are positive, and this means the proposed JND model outperforms the compared JND models on most of these images. Moreover, on image *I1*, the proposed model performs quite similar with the three compared models. With further analysis we can see that most regions in image *I1* are plain and present strong regularity. Thus, for image *I1*, its JND is mainly determined by the background luminance and the edge strength, while the proposed regularity factor plays marginal role. On *I6-I9*, the proposed model performs better than the other models, because these images possess large disorderly regions (e.g., grassland) and the proposed regularity plays a dominant role in JND estimation. Regularity is an effect factor which determines JND thresholds, and the proposed model performs more consistent with the subjective perception than the existing JND models.

IV. Conclusion

In this paper, we have proposed a visual regularity based JND estimation model. Inspired by the research achievement on cognitive science, we have analyzed the effect of content regularity on visual sensitivity. As it has been demonstrated, visual regularity plays an important role in visual masking. How to calculate visual regularity is still an open problem. According to the orientation selectivity mechanism, we have proposed to calculate visual regularity of local regions based on their orientation distribution. Finally, by combining visual regularity with luminance adaptation and contrast masking, a novel JND model has been proposed. Experimental results demonstrate that the proposed model performs consistent with the subjective perception.

REFERENCES

[1] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 467–476, 1995.

[2] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4892–4904, 2013.

[3] H. R. Wu, A. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *PROCEEDINGS OF THE IEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[4] C.-H. Chou and K.-C. Liu, "A perceptually tuned watermarking scheme for color images," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2966 –2982, Nov. 2010.

[5] W. Lin and C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[6] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 820– 829, July 2006.

[7] Z. Wei and K. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in dct domain," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337–346, Mar. 2009.

[8] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1705–1710, 2013.

[9] X. K. Yang, W. S. Lin, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662–680, 2005.

[10] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.

[11] N. B. Turk-Browne, J. Jung, and B. J. Scholl, "The automaticity of visual statistical learning," *Journal of Experimental Psychology. General*, vol. 134, no. 4, pp. 552–564, Nov. 2005.

[12] J. R. Saffran and E. D. Thiessen, "Pattern induction by infant language learners," *Developmental Psychology*, vol. 39, pp. 484–494, 2003.

[13] H.-J. Park and K. Friston, "Structural and functional brain networks: From connections to cognition," *Science*, vol. 342, no. 6158, p. 1238411, Nov. 2013.

[14] K. Friston, "The free-energy principle: a unified brain theory?" *Nat Rev Neurosci*, vol. 11, no. 2, pp. 127–138, Feb. 2010.

[15] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *The Journal of Physiology*, vol. 187, no. 2, pp. 437–445, 1966.