Self-similarity based structural regularity for just noticeable difference estimation[☆]

Jinjian Wu, Fei Qi[*], Guangming Shi[**]

[a]*School of Electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, PR China*

Abstract

In this paper, we introduce a novel just noticeable difference (JND) threshold estimation model based on a spatial masking function taking both luminance difference and structural regularity into account. Existing spatial masking functions underestimate the JND threshold for irregular textural regions, because they mainly consider the amplitude of luminance change for simplicity. As regular areas show weak masking effect due to their self-similar structures while irregular regions present strong masking effect, the spatial structure directly determines spatial masking. To effectively measure structural regularity in images under different contents, we propose an adaptive non-local self-similarity analysis based procedure. Then we weight luminance differences with similarity coefficients and deduce a new spatial masking function. Finally, an accurate JND estimation model is introduced. Experimental results demonstrate that the proposed JND model has a better visual effect than other models: it injects much noise into the insensitive regions, whereas little into the sensitive regions.

Keywords: Human visual system (HVS), Just noticeable difference (JND), Visual sensitivity, Spatial masking, Self-similarity, Non-local, Structural regularity, Image quality

[☆]This work is supported by National Natural Science Foundation of China under grant NO. 60805012, 61033004, and 61070138, and by RFDP under grant NO. 20090203110003.

[*]Corresponding author.

[**]Principal corresponding author.

Email addresses: jinjian.wu@mail.xidian.edu.cn (Jinjian Wu), fred.qi@ieee.org (Fei Qi), gmshi@xidian.edu.cn (Guangming Shi)

1. Introduction

The just noticeable difference (JND) threshold [1], which reveals the limitation of the human visual perception, is useful in image and video compression [2, 3, 4], image quality evaluation [5], watermarking [6], and so on. A lot of models have been proposed to estimate the JND threshold in the transform or spatial domains in the last few years.

Transform domain JND threshold estimations are performed in DWT [7] and DCT [8, 9] sub-bands, which usually consider the effects of spatial contrast sensitivity function (CSF), luminance adaptation, and spatial masking. In sub-band JND estimators, an input image is divided into blocks with small sizes to perform domain transformation. Further processing is applied on each block individually. As the block division breaks up the spatial relationship to some extent, sub-band based JND estimation models do not deal well with spatial structures for accurate spatial masking estimation.

Considering the relationship among pixels, spatial domain models directly compute the JND threshold for each pixel based on the spatial correlation between the central pixel and its local surroundings. In these models [10, 11, 12], two factors, which are luminance adaptation and spatial masking, are mainly taken into account, as shown in Fig. 1. The luminance adaptation models the fact that the HVS is sensitive to the luminance difference [10]. Generally, the luminance adaptation function is deduced according to Weber's law [10], which adapts to the perceptual vision and is widely used in spatial domain JND estimation.

Spatial masking is caused by interactions among stimuli [13]. It is a complicated visual perceptual mechanism without a widely accepted analytic model. Most existing models intend to estimate the spatial masking effect based on luminance contrast, which is a fundamental factor affecting spatial masking. In Chou and Li's model [10], the effect of spatial masking is computed based on the maximum signal along four directions. In Chiu and Berger's model [11], spatial masking is determined by the bigger luminance contrast in horizontal and vertical directions. Under these masking functions, places with high luminance contrast, e.g., edge regions, gain high JND thresholds.

However, edges can hide little noise since the human visual system (HVS) is very sensitive to

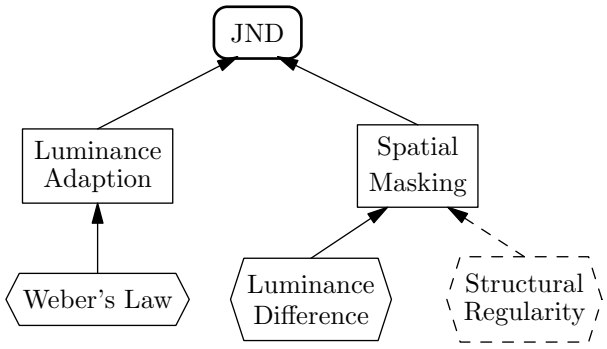


Figure 1: Architecture of the JND estimation model in spatial domain. Two factors, luminance adaptation and spatial masking, are mainly taken into account. It is widely accepted that luminance adaptation is deduced by Weber's law. While the spatial masking effect is complicated and existing models estimate it mainly based on luminance difference for simplicity. We find that structural regularity is another determination on spatial masking. So we estimate the spatial masking effect based on both luminance difference and structural regularity.

them [14]. To protect edge regions, Yang et al. [12] improved Chou and Li's model. In their model, the Canny edge detector is employed to detect edges firstly. Then these places are suppressed in computing the JND threshold of spatial masking. The intrinsic mechanism of the spatial masking function in Yang et al.'s model is still based on luminance contrast. Though the primary edge regions are protected, the secondary edge regions emerge and regions with irregular textures are still underestimated [15]. Therefore, further analysis on spatial masking is demanded to build an accurate spatial masking function.

We have found that, besides luminance difference, structural regularity is another decisive factor of the spatial masking effect, especially for natural images with complex textures. In fact, the HVS is adapted to extract structural information in images [16]. Therefore, it is highly sensitive to distortions in areas with regular structures, which means distortions in a region with regular texture are easy to be found out (detailed analysis is provided in section 2.1). Furthermore, our quantitative experiments certify that the JND thresholds are low in places with regular structures, while high in irregular regions [8, 17]. This motivates us to develop a structural regularity based spatial masking function for accurate JND threshold estimation.

In this paper, we propose a novel spatial masking function based on the structural regularity weighted luminance difference. Firstly, to effectively estimate the structural regularity of images in spatial domain, i.e., the periodically self-repeating contents, we introduce a non-local procedure. Considering the complexity of image background, a content adaptive similarity metric is introduced. Then, the spatial masking effect is measured via weighting luminance differences by similarity coefficients in a non-local region. Thus, an accurate spatial masking function is created. Finally, a new JND estimation model is obtained by non-linearly combining luminance adaptation and spatial masking.

The rest part of this paper is organized as follows. In section 2, we develop a novel spatial masking function based on the analysis of structural regularity to create a spatial domain JND estimation model. Experimental results on the spatial masking function and the overall JND threshold are provided in section 3. In section 4, we conclude the proposed approach.

2. JND threshold computation

In this section, we introduce the computational model of JND threshold. Firstly, we analyze the effect of structural regularity on spatial masking. Then we derive a new spatial masking function via weighting luminance differences with similarity coefficients. Finally, a new JND threshold estimator is proposed based on the new spatial masking function.

2.1. Spatial masking and image structure

The structural character of an image influences the spatial masking effect. Being sophisticated to extract structural information from an input scene [18], the HVS is highly sensitive to distortions in the place with regular structure, while insensitive to that in the irregular region. This is because a regular image presents self-repeating structure, i.e., its patches are much similar to patches nearby. With the comparison among these similar patches, it is easy to detect the distortion. So we should consider the structural regularity when estimating the spatial masking effect.

To analyze the sensation of the HVS on image structural regularity, we adopt four representative images, as shown in Fig. 2(a)–(d), which are formed with equivalent number of black and white pixels. Since the arrangements of the two kinds of pixels are dissimilar, the four images

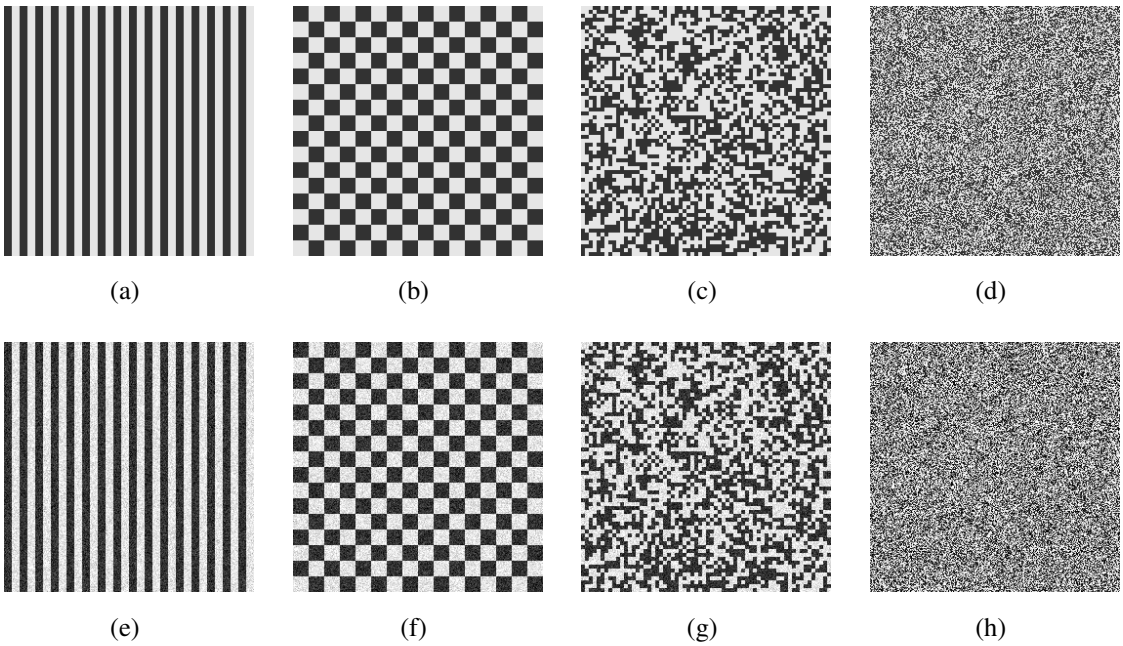


Figure 2: Image structural regularity influences the spatial masking effect. The top row shows original images and the bottom row corresponding contaminated images. From (a) to (d), the luminance contrast (represented by edge height) and the structural regularity are decreasing. Accompanying with the decrease of structural regularity, spatial masking effects in these images are gradually increased and the noise becomes less sensitive.

present different structures and textures. Fig. 2(e)–(h) are their corresponding noise contaminated images under a same white noise. It is easy to sense the noise in (e) and (f), while very hard in (g) and (h). Therefore, the spatial masking effects in (c) and (d) are stronger than that (a) and (b).

From the perspective of luminance edge height, Fig. 2 (a) and (b) are with high luminance contrast, while Fig. 2 (c) and (d) have low or even approximately no obvious edge. Since the existing spatial masking functions are based on luminance contrast directly [10] or indirectly [12, 15], it is hard for them to estimate the JND threshold accurately due to spatial masking effect, especially to (d) which has approximately no edge. For example, according to Chou and Li's model [10], Fig. 2 (a) and (b) will get higher JND thresholds than (c) and (d), which contradicts the perception of the HVS as mentioned in the above paragraph. Therefore, besides the luminance contrast, we should take the spatial arrangement of pixels into account for masking effect estimation.

The regularity of image structure has a dominant effect on the sensitivity of distortion for human visual perception. Since dependent pixels jointly carry structural information [16], the image structure appears as the arrangements of and relations among pixels. Fig. 2 (a) and (b) are with regular structures, as their pixels are strongly dependent and they possess high inter-pixel redundancy. Though both images have large luminance edge heights, it is easy to predict the value of a pixel from its surroundings, hence the HVS is highly sensitive to the distortion caused by noise. Therefore, the spatial masking effects of the two images are weak and they can hide little noise, as shown in Fig. 2 (e) and (f). With gradual decrease of the structural regularity, the dependence among pixels in Fig. 2 (c) and (d) becomes weak, so it is hard to accurately predict the value of a pixel from its neighbors. With non-uniform and irregular structural background, the spatial masking effect is strong. As a result, both images can hide much noise, as shown in Fig. 2 (g) and (h). In summary, structural regularity determines spatial masking, which indicates that places with irregular structures appear stronger spatial masking effects than regular regions do.

2.2. Spatial masking estimation

The spatial masking effect is caused by luminance change [19], which includes both structural regularity and luminance difference. A region with small luminance difference has weak spatial

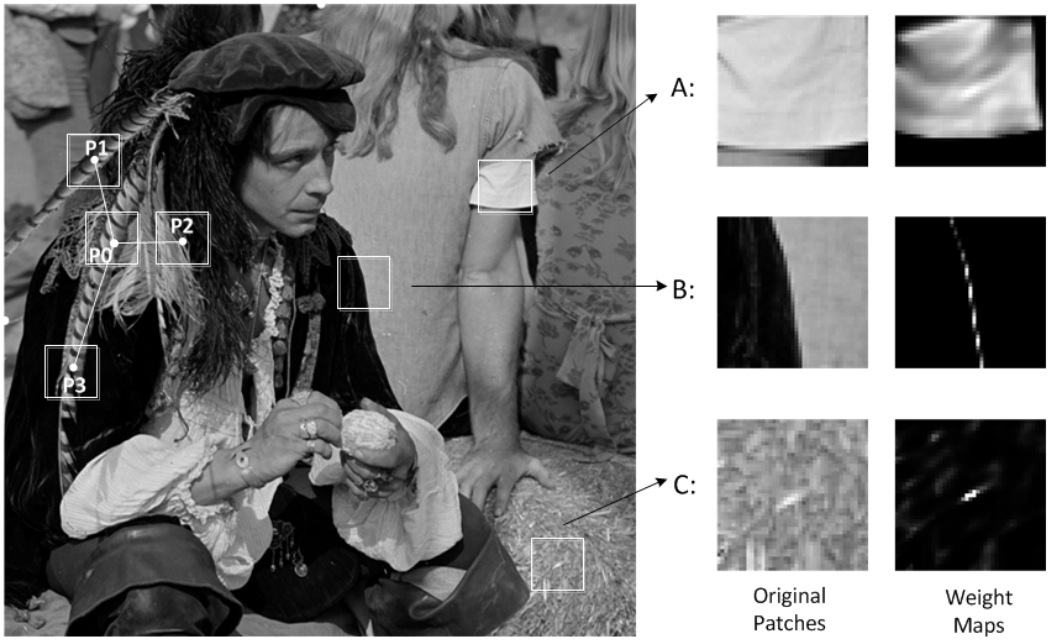


Figure 3: Illustration for the non-local estimation weights. The estimation weights between the estimation pixels (P_0) and the observations (P_1, P_2, and P_3) are based on their correlations rather than geometric distances. Though P_3 is the farthest one from P_0, it is the most similar one to P_0, and has the largest weight $w(P_0, P_3)$. Further more, $w(P_0, P_1)$ is larger than $w(P_0, P_2)$. Patches (A as smooth patch, B as edge patch, and C as texture patch) locate at the right side show the weights between the central pixels and their surroundings, and high luminance in the weight maps mean their corresponding pixels are similar with the central pixels.

masking effect, since it is uniform and presents regular structure. On the contrary, for a region with big luminance difference, its spatial masking effect is also weak if it has regular structure. Only when the region is with both big luminance difference and irregular structure, it presents strong masking effect. To model the mutual effect between the two factors, we weight luminance difference with structural regularity in the spatial masking function.

In regions with regular structure, pixels are strongly dependent with their surroundings. Therefore, we try to analyze the structural regularity based on self-similarity, which refers to the similarities among pixels of a region. As images represent some periodically self-repeating structures, we ought to find out these similar pixels for measuring the content's self-similarity. In this paper,

we adopt the non-local procedure to estimate the self-similarity.

Non-local is a scheme to exploit the structural similarity of image contents in spatial domain. With this scheme, we can average out noise for image denoising [20], or seek the structural character for super-resolution [21] and image deblurring [22]. Within non-local means, the estimation weights are based on the correlations between the estimation point and the observations [20, 23], rather than the geometric distance between two pixels. In other words, observations similar to the estimation point are highlighted with big weights in the non-local approach. As shown in Fig. 3, though P_3 is the farthest one from P_0, it is most similar to P_0 and has the largest weight $w(P_0, P_3)$. Moreover, P_1 is more similar to P_0 than P_2 does, and $w(P_0, P_1)$ is larger than $w(P_0, P_2)$. Therefore, the non-local approach exploits structural similarity of pixels.

With the non-local procedure, we measure the structural regularity by exploiting the self-similarity in images. If the central pixel is highly similar to pixels in its non-local neighborhood, it means the pixel locates in a region with regular structure, and vice versa. Therefore, we estimate the structural regularity via computing similarity coefficients between the center pixel and its surroundings. As illustrated by Fig. 3, the similarity coefficient between two pixels x and y is given by

$$w(x, y) = \frac{1}{\alpha(x)} \exp\left(-\frac{d(x, y)}{2\sigma_x^2}\right), \tag{1}$$

where $\alpha(x) = \sum_y \exp\left(-\frac{d(x,y)}{2\sigma_x^2}\right)$ is a normalizing constant, the parameter σ_x controls the decay rate of the similarity coefficient, and $d(x, y)$ denotes the distance between two rectangular local regions $\Omega(x)$ and $\Omega(y)$ centering at x and y, respectively. There are several ways to define the distance $d(x, y)$. Here, we take the sum of squared differences between corresponding image patches $\Omega(x)$ and $\Omega(y)$,

$$d(x, y) = \left\|\Omega(x) - \Omega(y)\right\|_2^2, \tag{2}$$

where $\|\cdot\|_2^2$ denotes the standard ℓ_2 norm.

Since images are complex and variable, the parameter σ_x is expected to be content adaptive. In denoising, Dore and Cheriet have provided a systematic analysis on setting up non-local parameters [24]. However, in spatial masking, the behavior of the parameter σ_x is slightly different. So we provide another adaptive scheme. The contents in smooth regions, e.g., patch A as shown

in Fig. 3, are homogeneous and most of the pixels are similar, while only a small part of pixels are similar in rough regions, e.g., patches B and C as shown in Fig. 3. To effectively estimate structural similarity, the similarity metric (1) is expected to restrain dissimilar pixels and highlight similar ones. To achieve this goal, a fast decay rate in (1) is expected for rough region to find out its structural character, while a slow one is demanded in smooth region for robustness and preserving the contributions from most of pixels. Therefore, a sharp Gaussian function with a small σ_x is assigned to a rough region, while a big σ_x to smooth region. As an effective metric of the roughness of a region, the variance var$(\cdot)$ is adopted to value the variable σ_x,

$$\sigma_x = \begin{cases} \sigma_0 & \text{if } \mathrm{var}(x) \le \sigma_0 \\ \sigma_0 \left(\frac{\sigma_0}{\mathrm{var}(x)}\right)^{0.5} & \text{else} \end{cases}, \tag{3}$$

where var(x) is the variance of the local region $\Omega(x)$, and the threshold σ_0 represents the minimum variance, under which the image content is considered to be smooth. In this paper, we set $\sigma_0 = 10$.

Then, weighting the luminance differences between the central pixel and pixels in its non-local surroundings by their similarity coefficients, the spatial masking function is deduced,

$$T_{\mathrm{tex}}(x) = \sum_{y \in \mathcal{R}(x)} w(x, y)(I(x) - I(y)), \tag{4}$$

where I is the original image, and $\mathcal{R}$ is a non-local surroundings. Since pixels have strong correlations with their nearby pixels [20], the correlations will always decay when the distance increasing. For simplicity, we employ a big enough rectangular neighboring region $\mathcal{R}$.

2.3. JND threshold computation

In spatial domain, the JND threshold of each pixel from an achromatic image is primarily affected by two factors [10, 12]. One is the spatial masking effect, which has been analyzed in detail in the previous subsection. The other is the background luminance adaptation. Inspired by Weber's law, which indicates that the ratio of the just noticeable illuminance change to the background illuminance is approximate constant, a perceptual experiment is designed to test the JND threshold due to background luminance, and the luminance adaptation function is deduced [10],

$$T_{\mathrm{lum}}(x) = \begin{cases} a_0 + a_1 \sqrt{\frac{B(x)}{B_0}} & \text{If } B(x) < B_0 \\ \gamma + \gamma B(x) & \text{else} \end{cases}, \tag{5}$$

where $a_0 = 20$, $a_1 = -17$, $\gamma = 3/128$, $B_0 = 127$, and $B(x)$ is the background luminance of the pixel x.

The visibility threshold is determined by luminance adaptation and spatial masking simultaneously. The luminance adaptation of a region is decided by its background, and that the spatial masking effect is caused by its foreground content. The two parts jointly operate on the human visual perception. So we compute the overall JND threshold based on a combination of the two parts rather than getting the bigger one with the winner-take-all scheme. Furthermore, there exists overlapping between the two parts, thus we cannot directly add the two parts for combination. Since the bigger part plays a more important rule in the overall JND value, we should weight the two parts with their own values for combination. Here, we combine the two derivations, the luminance adaptation function (5) and the spatial masking function (4), with a nonlinear procedure [17] to acquire the overall JND value of a pixel.

$$T_{\text{jnd}}(x) = \theta(x)T_{\text{lum}}(x) + (1 - \theta(x))T_{\text{tex}}(x), \tag{6}$$

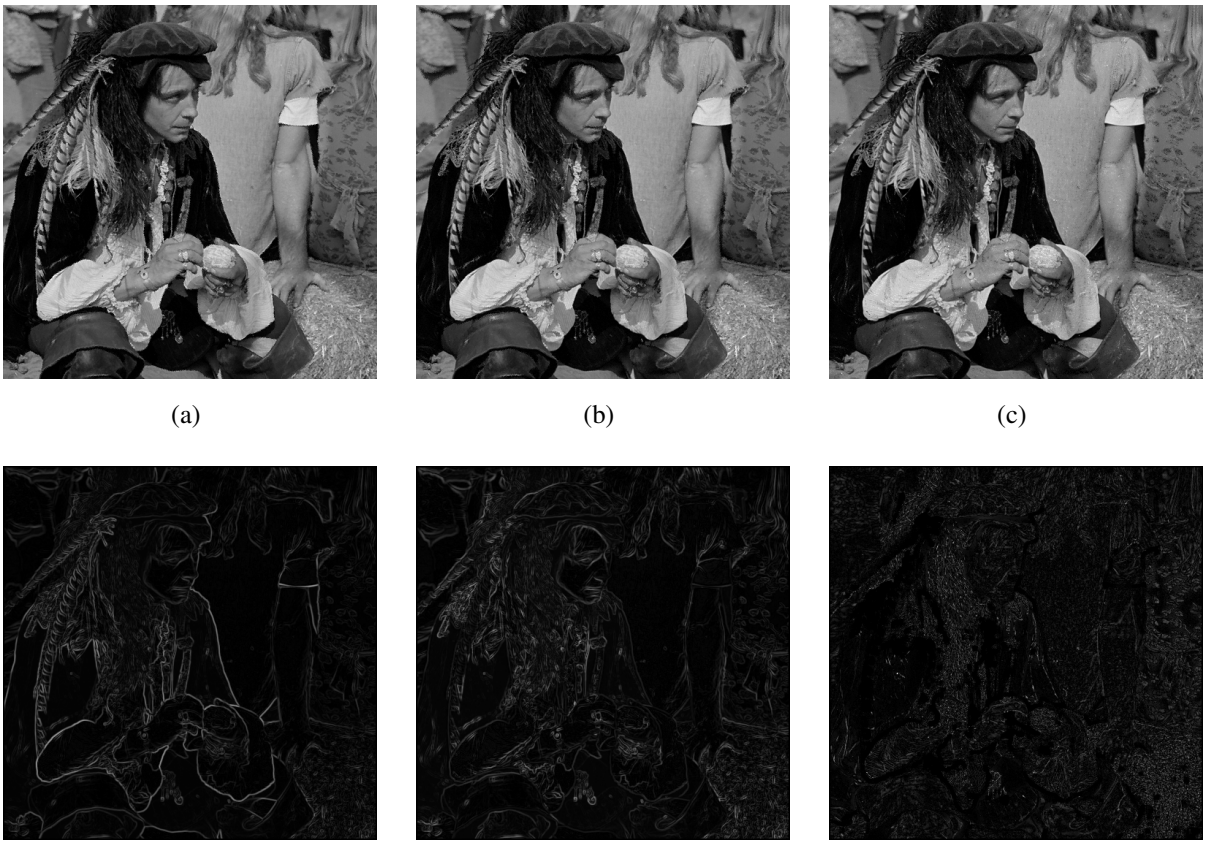
where $T_{\text{jnd}}(x)$ is the JND threshold of pixel x, the weight is $\theta(x) = \frac{T_{\text{lum}}(x)}{T_{\text{lum}}(x)+T_{\text{tex}}(x)}$, which is normalized to highlight the important part.

3. Experimental results and discussion

In this section, experiments are demonstrated to evaluate the performance of the proposed content adaptive JND estimation model. According to the JND estimation functions derived in section 2, we firstly compute the JND threshold of an input image. Then, we inject JND guided noise into images to make a comparison with other JND models. The JND guided noise shaping equation is

$$I'(x) = I(x) + \beta s(x)N(x), \tag{7}$$

where I' is the image contaminated by JND noise N. The parameter β regulates the energy of JND noise among different models, and s takes $+1$ or -1 randomly. In our experiment, we take the size of $\Omega(\cdot)$ as 7×7 and the size of $\mathcal{R}$ as 21×21.



(a) (b) (c)

(d) (e) (f)

Figure 4: Indian images, in the size of 512×512, are contaminated by the spatial masking JND noises with a same MSE of 41.18. The top row shows contaminated images and the bottom row the maps of the spatial masking noises. (a) and (d) Chou et al.'s model. (b) and (e) Yang et al.'s model. (c) and (f) The proposed model.

3.1. Spatial masking effect

Different to conventional computational functions based on luminance edge height [19, 10], the proposed spatial masking function considers both luminance difference and structural regularity. According to (4) and (7), the JND noise due to spatial masking, i.e., set $N = T_{tex}$, is injected into the Indian image. Then, we make a comparison with Chou et al.'s [10] and Yang et al.'s [12] spatial masking procedures (the energy of noise is the same), as shown in Fig. 4. Since spatial masking is directly computed based on luminance edge height in Chou et al.'s model, most noise is injected into primary edge regions. Moreover, the noise mask of Chou et al.'s model looks like

an edge detection result, as shown in Fig. 4 (d). We can easily sense the distortion in edge regions, i.e., the black coat edge region of the Indian in Fig. 4 (a), as the HVS is sensitive to it. Though little noise is injected into the primary edge regions by Yang's model, much noise moves to the secondary edge regions, i.e., the edge of the face in Fig. 4 (b), which we can also easily sense. This is because the edge regions, including both primary and secondary edges, present strong self-repeating structures, the spatial masking effect is weak in this kind of regular regions. With the proposed spatial masking procedure, the noise is mainly distributed into the irregular textural regions, while the smooth and edge regions are effectively protected, as shown in Fig. 4 (c) and (f).

For further analysis of the spatial masking functions, we extract some representative patches from the Indian image to make detailed comparisons. As shown in Fig. 5, in the first column, four image patches are extracted from the original Indian image, each with a size of 60×60. From top to bottom, we denote them as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively, for the convenience of discussion. Patches $\mathcal{A}$ and $\mathcal{B}$ possess some prominent edge regions. They are severely distorted with Chou et al.'s model. The distortions are small in Yang et al.'s model, while tiny in the proposed model. Patch $\mathcal{C}$ contains some secondary edge regions, such as the mouth and jaw. Yang et al.'s model produces the most severely distorted patch, and the proposed model outputs the best result in all three models. As patch $\mathcal{D}$ shows, it is an irregular textural region which is insensitive to the HVS. The proposed model injects most of the noise into it, while Chou et al.'s and Yang et al.'s models inject less.

In summary, the proposed spatial masking function effectively protects the HVS sensitive regions (e.g., the smooth and edge places), while injects much noise into insensitive regions (e.g., the irregular places) at the same time. The proposed spatial masking function produces a better perceptual spatial masking JND result.

3.2. JND threshold

By taking $N(x) = T_{\text{jnd}}(x)$ in the JND guided noise shaping equation (7), we inject the overall JND noise into the Cameraman image and make a comparison among the proposed model, Chou et al.'s model [10], and Zhang et al.'s model [8]. The contaminated images and their corresponding

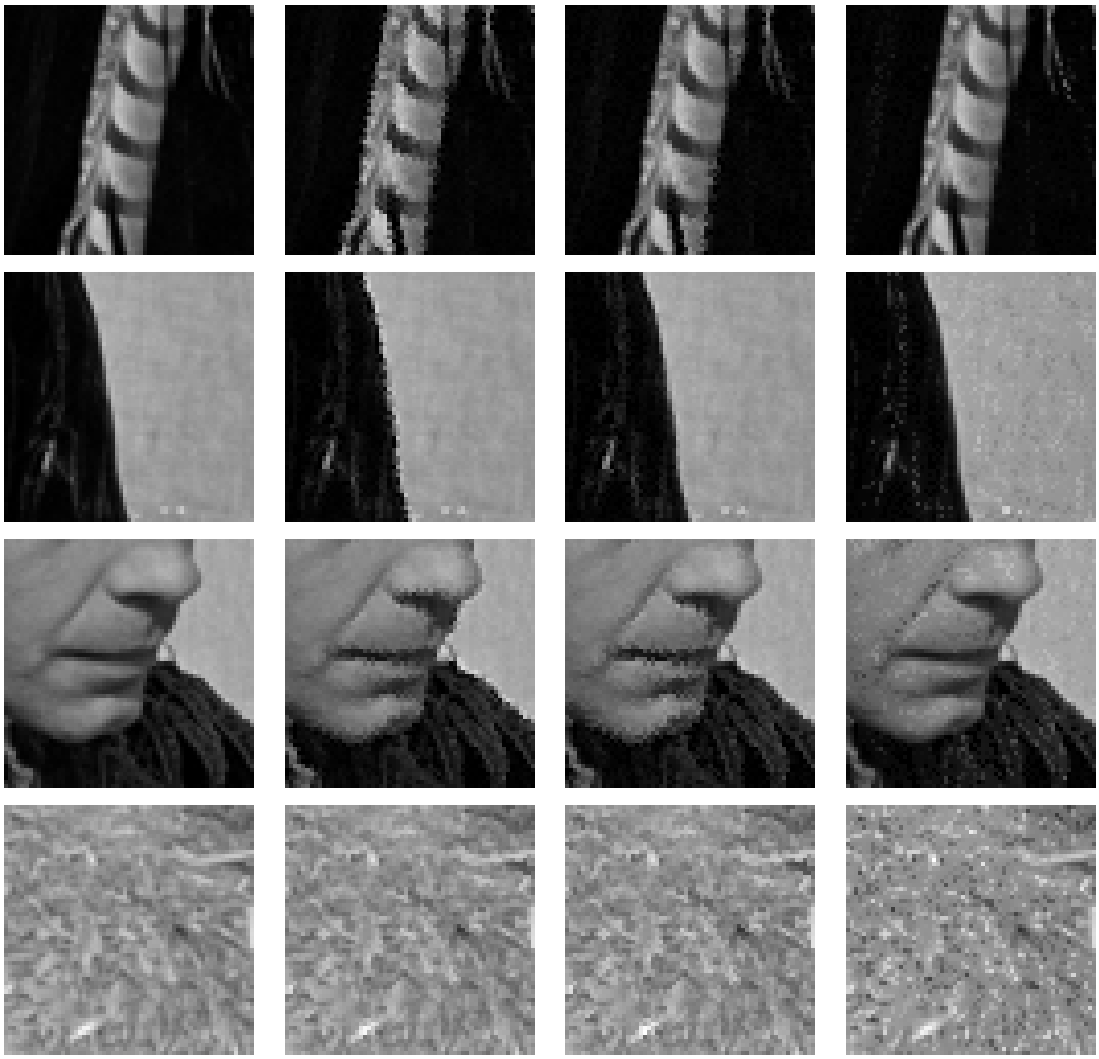


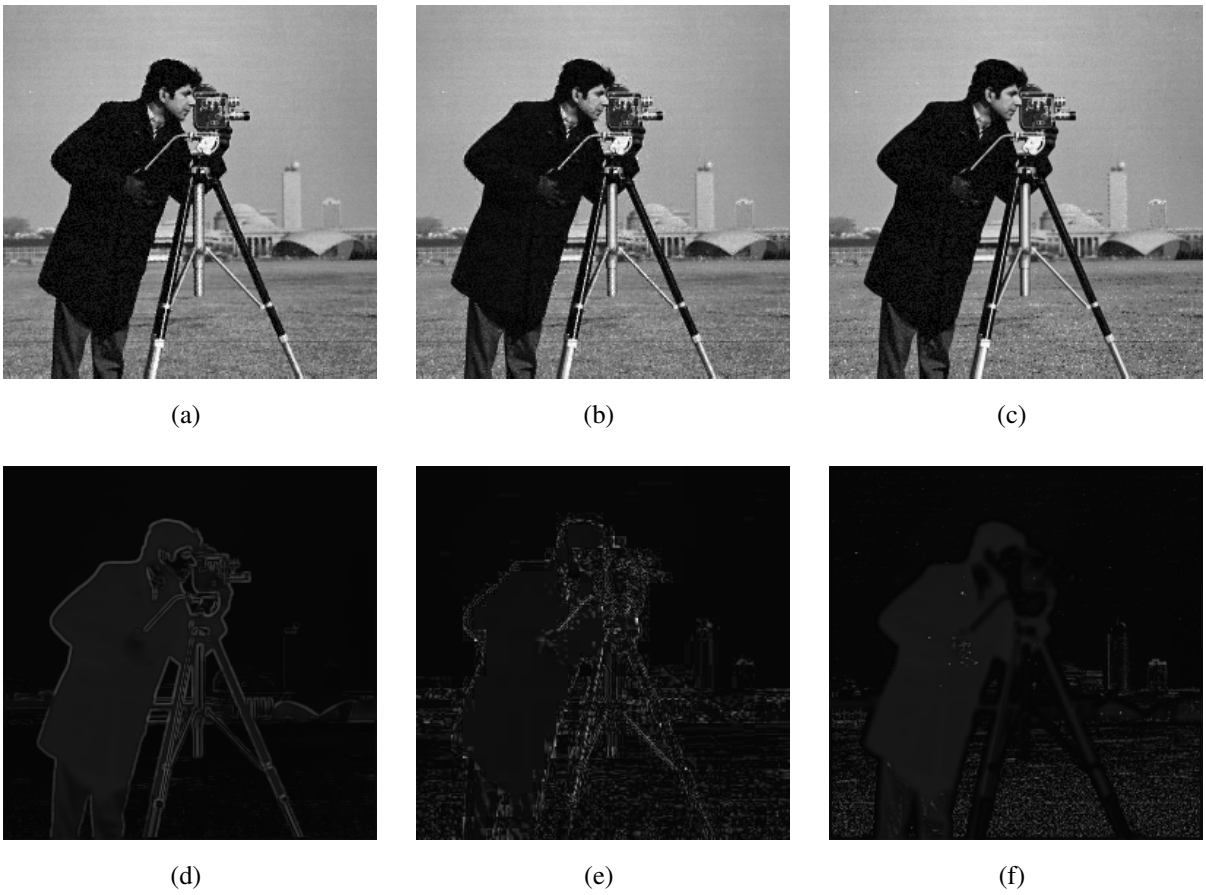
Figure 5: Detailed analysis on some patches (60×60) of the Indian image. From top to bottom, we denote the patches as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively. From left to right, columns are original patches, results of Chou et al.'s model, results of Yang et al.'s model, and results of the proposed model, respectively.

noise mask maps are shown in Fig. 6. Because Chou et al.'s model is a local information based method, most of the noise is uniformly distributed at edge regions and the black coat region, as shown in Fig. 6 (d). The edge regions are evidently distorted by too much noise in Fig. 6 (a). Since the transform domain JND method is operated on individual blocks, as Fig. 6 (e) shows, the noise mask map of Zhang et al.'s model appears strong blocking artifact. That is because pixels in edge regions and their neighborhoods are cut into a same grid for computing the JND threshold, which are overestimated and are severely distorted with strong blur, as Fig. 6 (b) shows. The proposed JND estimation model is content adaptive, it injects most of the noise into texture regions, especially the irregular grass region, and effectively protects the edge and plan regions, as Fig. 6 (f) shows. The contaminated image of the proposed model is shown in Fig. 6 (c), it has a better perceptual vision than the other two.

Three patches extracted from the Cameraman image are shown in Fig. 7 for detailed comparisons. The contaminated patches are extracted from their corresponding noised images. In patch $\mathcal{E}$, the result of Chou et al.'s model is severely distorted at the edge region. The neighbor of the edge region and the face in patch E are grossly distorted with Zhang et al.'s model. However, the output of the proposed model for patch $\mathcal{E}$ appears almost intact. The effects of these algorithms' outputs for patch $\mathcal{F}$ are similar to $\mathcal{E}$, such as there is strong blur in Chou et al.'s and Zhang et al.'s results. The HVS is insensitive to the glass region $\mathcal{G}$, which has irregular texture. The proposed model injects more noise into this region than the other two models. With the novel spatial masking function, the proposed JND estimation model provides a better perceptual quality than Chou et al.'s and Zhang et al.'s models.

3.3. Subjective quality evaluation

In order to provide a comprehensive comparison among the JND estimation of these algorithms, we perform subjective quality evaluation. In our experiments, 10 images are chosen as test images. Due to the limitation of the resolution of our screen, original images are resized to 256×256 first to perform the JND guided contamination. Then, the original image and the contaminated images (with the same level of noise) are shown on the screen for subjective evaluation. During the experiments, the position of the original image is constant (locates at the top left of the



(a) (b) (c)

(d) (e) (f)

Figure 6: Cameraman images, in the size of 256×256, are contaminated by overall JND noises with a same MSE of 53.15. The top row shows contaminated images and the bottom row the JND noise masks. (a) and (d) Chou et al.'s model. (b) and (e) Zhang et al.'s model. (c) and (f) The proposed model.

screen), while the positions of the contaminated images are randomly located at the other three places. Fifteen subjects are invited into our evaluation. Each subject is requested to assign a score to the three images being compared. The one which provides the best perceptual quality and is most similar to the original image gains the highest score 3, the secondary one gains 2, and the poorest gains 1. Images with similar perceptual quality gain the same score.

The subjective evaluation results on the contaminated images with JND threshold due to spatial masking are shown in Table 1. For most images, the proposed spatial masking procedure gains the highest scores. The average score of the proposed procedure is far better than Chou et al.'s and

Figure 7: Detailed analysis on some patches (60×60) of the Cameraman image. From top to bottom, we name the patches as $\mathcal{E}$, $\mathcal{F}$, and $\mathcal{G}$, respectively. From left to right, columns are original patches, results of Chou et al.'s model, results of Zhang et al.'s model, and results of the proposed model, respectively.

Table 1: Subjective quality evaluation on images contaminated by spatial masking guided noises. For each image, noises added by the three models are with a same energy.

Image	Chou	Yang	Proposed
Indian	1.8000	1.2667	2.8667
Barboon	1.0667	2.5333	2.6667
Babara	1.1333	2.1333	2.4000
Lena	1.8000	1.8000	2.2000
Parrot	2.5333	1.8667	2.2000
Pepper	1.6000	1.8667	2.4667
Cameraman	2.0000	1.6667	2.6667
Boat	1.6000	2.5333	2.2000
House	1.1333	2.0000	2.6000
Port	2.0000	2.2000	2.3333
Average	1.6667	1.9867	2.4600

Yang et al.'s models. Table 2 shows the subjective evaluation results on the contaminated images with overall JND guided noise. As suppressed by the luminance adaptation to some extent when combining for the overall JND threshold, the spatial masking effect is not as obvious as in Table 1. However, the proposed model still outperforms Chou et al.' and Zhang et al.'s models, which further confirms the effectiveness of the proposed spatial masking procedure.

4. Conclusions

In this paper, a novel image domain JND estimation method is proposed. From empirical studies, we have found that places with regular structures can hide little noise, whereas irregular places can hide more. So we analyze the effect of structural regularity on spatial masking and introduce a novel spatial masking function. Considering the correlations among pixels, we estimate structural regularity with a non-local self-similarity based procedure. The spatial masking effect is estimated via weighting luminance differences with similarity coefficients. Then, combining spatial mask-

Table 2: Subjective quality evaluation on images contaminated by the overall JND guided noises. For each image, noises added by the three models are with a same energy.

Image	Chou	Zhang	Proposed
Indian	2.4667	1.6667	2.7333
Barboon	2.0000	2.4667	2.6667
Babara	2.3333	1.6667	2.2000
Lena	2.2000	1.8667	2.3333
Parrot	2.3333	1.9333	1.9333
Pepper	1.6000	2.0667	2.4667
Cameraman	2.0667	2.3333	2.4000
Boat	2.2000	1.9333	2.5333
House	1.9333	2.5333	2.6000
Port	2.2000	2.0000	2.4667
Average	2.1333	2.0467	2.4333

ing and luminance adaptation with a nonlinear procedure, the overall JND threshold is deduced. Since the proposed spatial masking function is content adaptive, no explicit approach is required to discriminate plain, edge, and texture regions. Experimental results show that the proposed algorithm injects less noise into the HVS sensitive regions, while much more into insensitive regions. The proposed model has an aggressive JND estimation on gray images without causing visible distortion.

Though the proposed JND estimation model offers a much accurate JND threshold, some aspects remain for improvement. Visual masking is so complicated that there is no complete analysis about it, thus further investigation on the spatial masking function is needed. The proposed model merely considers gray and still images, we plan to extend it to color and dynamic images.

[1] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models of human perception, Proc. IEEE 81 (10) (1993) 1385–1422.

[2] Z. Chen, C. Guillemot, Perceptually-Friendly H.264/AVC video coding based on foveated Just-Noticeable-Distortion model, IEEE Trans. Circuits Syst. Video Technol. 20 (6) (2010) 806–819.

[3] I. Hontsch, L. Karam, Adaptive image coding with perceptual distortion control, IEEE Trans. Image Process. 11 (3) (2002) 213–222.

[4] T. Colleu, S. Pateux, L. Morin, C. Labit, A polygon soup representation for multiview coding, J. Visual Communication and Image Representation 21 (5–6) (2010) 561–576.

[5] W. Lin, C. J. Kuo, Perceptual visual quality metrics: A survey, J. Visual Communication and Image Representation 22 (4) (2011) 297–312.

[6] Y. Niu, J. Liu, S. Krishnan, Q. Zhang, Combined just noticeable difference model guided image watermarking, in: IEEE Int'l Conf. Multimedia and Expo (ICME), 2010, pp. 1679–1684.

[7] A. Watson, G. Yang, J. Solomon, J. Villasenor, Visibility of wavelet quantization noise, IEEE Trans. Image Process. 6 (8) (1997) 1164–1175.

[8] X. Zhang, W. Lin, P. Xue, Just-noticeable difference estimation with pixels in images, J. Visual Communication and Image Representation 19 (1) (2008) 30–41.

[9] Z. Wei, K. Ngan, Spatio-Temporal just noticeable distortion profile for grey scale Image/Video in DCT domain, IEEE Trans. Circuits Syst. Video Technol. 19 (3) (2009) 337–346.

[10] C.-H. Chou, Y.-C. Li, A perceptually tuned subband image coder based on the measure of just-noticeable distortion profile, IEEE Trans. Circuits Syst. Video Technol. 5 (6) (1995) 467–476.

[11] Y.-J. Chiu, T. Berger, A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancement, IEEE Trans. Circuits Syst. Video Technol. 9 (3) (1999) 438–450.

[12] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, S. S. Yao, Just noticeable distortion model and its applications in video coding, Signal Processing: Image Communication 20 (7) (2005) 662–680.

[13] S. L. Macknik, M. S. Livingstone, Neuronal correlates of visibility and invisibility in the primate visual system, Nat Neurosci 1 (2) (1998) 144–149.

[14] M. P. Eckert, A. P. Bradley, Perceptual quality metrics applied to still image compression, Signal Processing 70 (3) (1998) 177–200.

[15] A. Liu, W. Lin, F. Zhang, M. Paul, Enhanced just noticeable difference (JND) estimation with image decomposition, in: 17th IEEE Int'l Conf. Image Processing, ICIP, 2010, pp. 317–320.

[16] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[17] J. Wu, F. Qi, G. Shi, An improved model of pixel adaptive just-noticeable difference estimation, in: IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, ICASSP, 2010, pp. 2454–2457.

[18] Z. Wang, A. Bovik1, L. Lu, Why is image quality assessment so difficult?, in: Proc. IEEE Int'l Conf. Acoust., Speech, and Signal Processing, ICASSP, Vol. 4, 2002, pp. 3313–3316.

[19] A. Netravali, B. Prasada, Adaptive quantization of picture signals using spatial masking, Proc. IEEE 65 (4) (1977) 536–548.

[20] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: IEEE Comp. Soc. Conf. Computer Vision Pattern Recognition, Vol. 2, 2005, pp. 60–65.

[21] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the Nonlocal-Means to Super-Resolution reconstruction, IEEE Trans. Image Process. 18 (1) (2009) 36–51.

[22] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and Super-Resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process. 20 (7) (2011) 1838–1857.

[23] V. Katkovnik, A. Foi, K. Egiazarian, J. Astola, From local kernel to nonlocal Multiple-Model image denoising, Int. J. Computer Vision 86 (2009) 1–32.

[24] V. Dore, M. Cheriet, Robust NL-Means filter with optimal Pixel-Wise smoothing parameter for statistical image denoising, IEEE Trans. Signal Process. 57 (5) (2009) 1703–1716. doi:10.1109/TSP.2008.2011832.