

A Survey of Visual Just Noticeable Difference Estimation

Jinjian WU (✉)^{1,2}, Guangming SHI^{1,2}, Weisi LIN³

¹ School of Electronic Engineering, Xidian University, Xi'an 710071, China

² Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China

³ School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

Abstract

The concept of just noticeable difference (JND), which accounts for the visibility threshold (visual redundancy) of the human visual system (HVS), is useful in perception oriented signal processing systems. In this work, we give a comprehensive review on JND estimation technology. The visual mechanism and its corresponding computational modules are firstly illustrated. These include luminance adaptation, contrast masking, pattern masking, and contrast sensitive function. Next, existing pixel domain and subband domain JND models are presented and analyzed. Finally, the challenges in JND estimation are discussed.

Keywords Just Noticeable Difference, Human Visual System, Luminance Adaptation, Contrast Masking, Pattern Masking, Contrast Sensitive Function

1 Introduction

The human visual system (HVS) plays an extremely important role for outside world understanding. In our daily life, more than 70% information that we received comes from the HVS. As a sophisticated visual signal processing system, the HVS can process the input scene rapidly and effectively. Due to the characters of the optic cells in the retina, the HVS has limited resolution and can only perceive the changes larger than a certain threshold [1]. The just noticeable difference (JND) accounts for such a threshold, which estimates the distinguishability of the HVS [2]. Since

the JND threshold represents the visibility of the HVS, it is useful for perceptual orientated signal processing systems, e.g., perceptual image/video compression [1, 3], visual image/video enhancement [4, 5], perceptual quality assessment [6, 7], watermarking [8–10], information hiding [11–13], and so on.

The distinguishability of the human eye has been firstly investigated by cognitive scientists during the past decades [14–19]. Inspired by the research finds from cognitive science, a large amount of JND estimation models have been proposed. According to the domain that the JND threshold is computed, these existing JND estimation models are classified into two categories: 1) the pixel domain JND estimation models, which directly calculate the JND threshold for each pixel [20–23]; 2) the subband domain JND estimation models, which firstly transfer the image in to the subband domain (e.g., the DCT domain), and then calculate the JND threshold on each subband [24–30].

Pixel domain JND models usually take luminance adaptation and contrast masking into account for JND threshold estimation [20–22]. The human eye presents different sensitivities for different background luminance [31]. Thus, the visibility thresholds for different background luminances are investigated for luminance adaption modeling [32]. Moreover, the visibility of stimulus is decreased with the non-uniform surrounds, as the spatial non-uniformity causes masking effect among stimuli [33]. The contrast masking effect is analyzed for visibility threshold estimation on non-uniform background [34].

By considering the effect from luminance adaptation and contrast masking, Chou and Li proposed a pioneer JND estimation model [20]. Yang et al. adopted the Canny

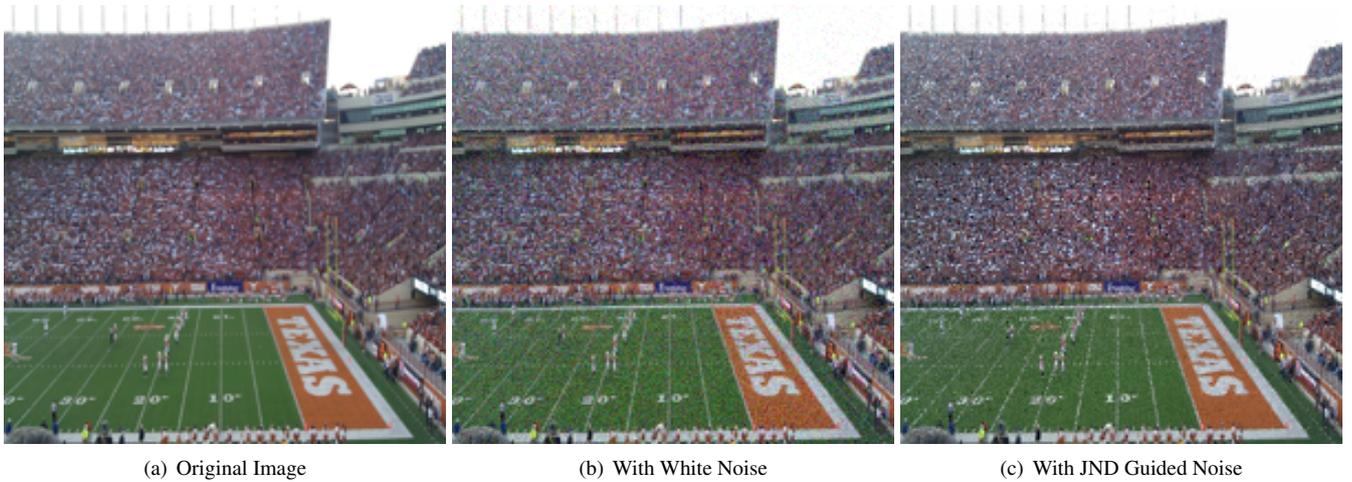


Fig. 1: JND demonstration, where the same scene (with size of 200×200) contaminated by the white noise and the JND guided noise. Though the noise level in the two contaminated images are the same (with $\text{PSNR}=26.84\text{dB}$), their visual qualities are obviously different.

operator to protect the edge region during JND estimation [21]. Moreover, Liu et al. firstly decomposed an image into two parts (edge and texture regions) for separately computation [22]. Recently, Wu et al. suggested to take structural regularity into account for JND estimation [2, 35]. Moreover, Wang et al. extended the JND estimation to screen content images [36]. Since the pixel domain model directly calculates the JND threshold of each pixel, it returns a direct view of the JND mask, which is useful for motion estimation, image enhancement, quality assessment, and so on [6, 37]. However, such kind of JND models can not take the contrast sensitive function (which describes the sensitivity of the HVS for difference spatial frequencies within the transform domain) into account to further improve the performance, and are not convenience for signal compression system (which is subband based) [25].

Subband domain JND estimation models mainly consider the masking effects from contrast sensitive function (CSF), contrast masking, and luminance adaptation [24–26]. It is well known that the HVS has different sensitivities (visibilities) for different frequencies. The visibility thresholds for different subbands (frequencies) are measured through subjective viewing test, and the CSF is built to account for the fundamental/base JND threshold for each subband [24].

By adjusting the base JND with the affection from other factors, e.g., luminance contrast and background luminance, a DCT domain JND estimation model was proposed [24]. Zhang et al. further consideration of luminance contrast in

both inter and intra bands for JND estimation [26]. In [25], a thorough analyses on CSF is made for a more accurate JND estimation. A temporal color JND estimation model was proposed in [27]. Moreover, a generalized DCT-based JND for any size of transform is proposed in [38]. Since the subband domain model calculates the JND threshold on each subband, such model is popular for perceptual image/video compression (which is subband coded) [39]. However, these subband domain JND models isolate each block from its surrounds, as a result, the masking effects (especially for the complicated texture regions with strongly content changes) can not be accurately estimated [2].

The rest of this paper is organized as following. The modules that affect the JND threshold are firstly reviewed in Section 2. Next, the two categories of the existing JND models are analyzed in Section 3. The subjective viewing test for JND model verification is demonstrated in Section 4. Section 5 provides a discussion on JND technology. Finally, the conclusion is drawn in Section 6.

2 Modules that affect the JND threshold

The JND reveals the visibility of the HVS. An example is given in Fig. 1, in which the original image (a) is contaminated by the white noise (as (b) shows) and the JND guided noise (as (c) shows). Though the noise levels in the two contaminated images are the same (with $\text{PSNR}=26.84\text{dB}$), their visual qualities are obviously different. As shown in Fig. 1 (b), it is easy to sense the

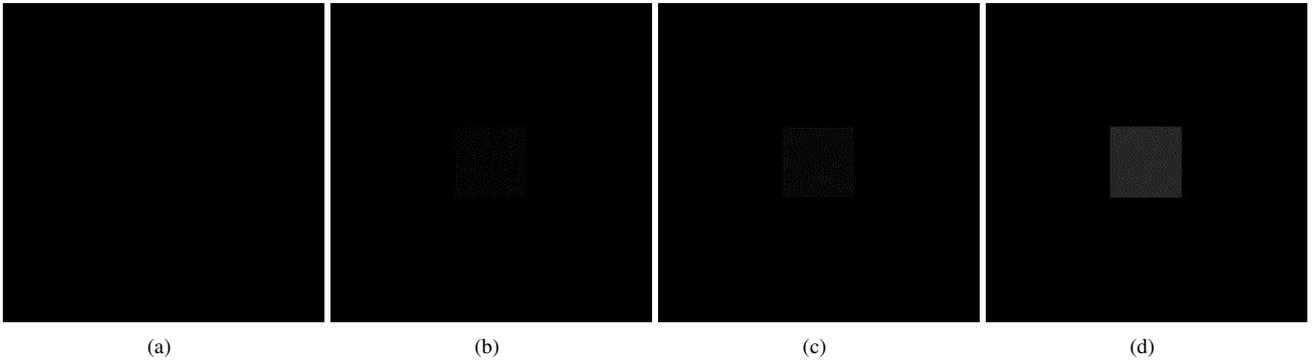


Fig. 2: Visibility threshold for the uniform background (the viewing condition for the subjective viewing test is set according to the ITU-R BT.500-11 standard [40]). From (a) to (d), the gray levels for the central region is 0, 5, 15, and 30, respectively.

disturbances caused by the white noise. With the help of the JND model, much noise is guided into these insensitive regions, while less into the sensitive regions. As a result, the noise in Fig. 1 (c) is almost invisible.

There are many factors which affect the visibility of the HVS. During the past decade, several modules have been investigated for JND modeling, which are luminance adaptation, contrast masking, pattern masking, CSF, and so on. In this section, we will give detailed introductions on these modules.

2.1 Luminance Adaptation

The resolution of the HVS is limited. We can hardly sense the small changes even for uniform background. As shown in Fig. 2, (a) is the original patch with uniform black background (i.e., the luminance level $B = 0$). The luminance value of the central region for such patch is increased with different levels (ΔB), as shown in Fig. 2 (b)-(d). When projecting these images onto a monitor for visibility threshold measurement (the viewing condition follows the ITU-R BT.500-11 standard [40], e.g., subjects are asked to sit in the front of the monitor, and the viewing distance is four times of the image height), it can be seen that: (1) with small increased luminance value at the central region of Fig. 2 (b) (i.e., the change value is $\Delta B = 5$), we cannot sense any change; (2) with an increase $\Delta B = 15$, we can still hardly sense the change even with carefully staring at the central region of Fig. 2 (c); (3) when the increase is as large as 30 ($\Delta B = 30$), most of the subjects can easily sense the change in the central region, as shown in Fig. 2 (d).

Besides, the HVS has different visibility thresholds to different background luminance values [42]. As an example, the visibilities of our eyes are greatly decreased in the night

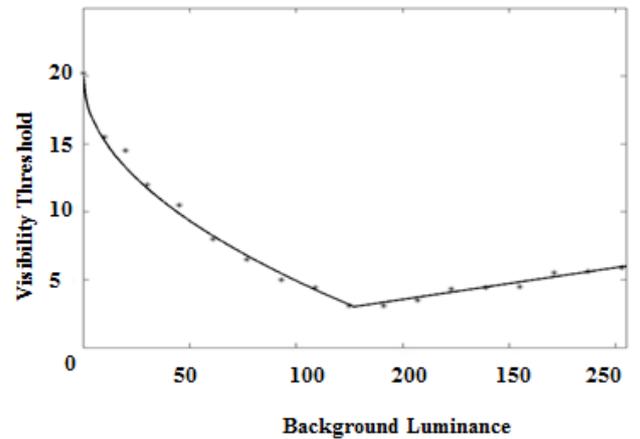


Fig. 3: Luminance adaptation: visibility threshold for each luminance level [20, 23].

with limited light, while much more sensitive for a suitable environment with soft light. The visibility thresholds have been thoroughly investigated during the past decade. Inspired by the Weber's law, the ratio of its visibility changes to the luminance of the stimulus is approximately constant [43] (i.e., the Weber fraction) for a given stimulus. Moreover, results from the subjective viewing test show that the HVS is less sensitive in the dark environment than that with strong light. Thus, for dark environment with low background luminance, the value of Weber fraction decreases with the increases of the background luminance; while for light environment, the value of Weber fraction remains constant for different luminance [20]. Therefore, the luminance adaptation threshold is with quasi-parabola curves. For a given scene with luminance level in $[0, 255]$, its visibility threshold of luminance adaptation can be

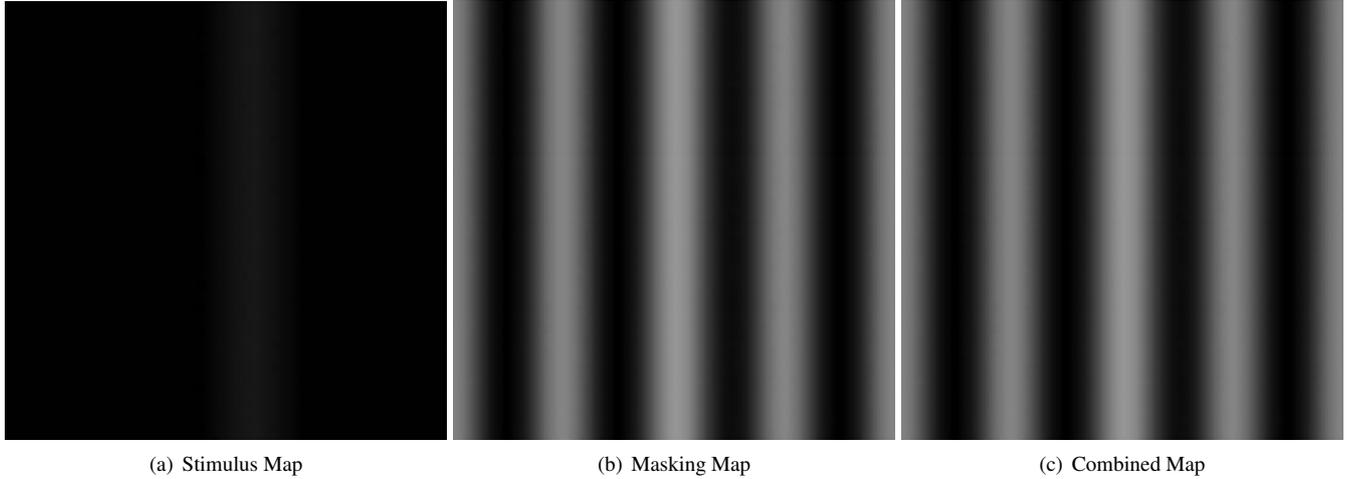


Fig. 4: Contrast masking effect [41]. (a) Stimulus map: a single light bar in a uniform black background region; (b) Masking map: non-uniform region with gratings; (c) Combined map: (a)+(b).

calculated as following,

$$LA(x) = \begin{cases} 17 \times (1 - \sqrt{\frac{B(x)}{127}}) & \text{If } B(x) < 127 \\ \frac{3}{128} \times (B(x) - 127) + 3 & \text{else} \end{cases}, \quad (1)$$

where $LA(x)$ is the visibility threshold of luminance adaptation of pixel x ; $B(x)$ is the background luminance, which is calculated as the mean luminance of a local region. The quasi-parabola curve of the luminance adaptation is shown in Fig. 3.

2.2 Contrast Masking

The visibility threshold for a non-uniform region is obviously higher than that of a uniform region. As shown in Fig. 4 (a), though the single light bar is weak, we can sense it in the uniform black background. When the background becomes non-uniform (as the grating map shown in Fig. 4 (b)), we can hardly sense the single light bar in such background (as shown in Fig. 4 (c), which is a combination of Fig. 4 (a) and (b)). That is because there exist spatial masking effects among stimuli in such a non-uniform region (i.e., between the single light bar and the gratings) [16, 41]. The spatial masking is a comprehensive response caused by many factors. During the past decades, several factors have been investigated to estimate the spatial masking effect. However, the spatial masking effect is too complicated to be modeled with a single theoretical formulation [20]. It is still an open problem to accurately model the spatial masking effect.

As a straightforward and simple factor, luminance contrast is usually chosen to estimate the spatial masking

0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	-1	0
1	3	8	3	1	0	8	3	0	0	0	0	3	8	0	0	3	0	-3	0
0	0	0	0	0	1	3	0	-3	-1	-1	-3	0	3	1	0	8	0	-8	0
-1	-3	-8	-3	-1	0	0	-3	-8	0	0	-8	-3	0	0	0	3	0	-3	0
0	0	0	0	0	0	0	-1	0	0	0	0	-1	0	0	0	1	0	-1	0

Fig. 5: Four directional filters.

effect [14, 20]. Generally, an image region with a high luminance contrast has a higher visibility threshold than the one with low luminance contrast. Moreover, the visibility threshold is increased with the increase of luminance contrast. For contrast masking computation, the luminance contrast (LC) is firstly calculated,

$$LC(x) = \max_{k=1,\dots,4} G_k(x), \quad (2)$$

$$G_k = |\varphi \mathcal{I} * \nabla_k|, \quad (3)$$

where G_k is the gradient value along the k -th direction, φ is a parameter which relates to the directional filters, \mathcal{I} is the input image, and ∇_k is the directional filter. Four frequently-used ∇_k is shown in Fig. 5. According to the numbers in the four directional filters, φ is set as $1/16$.

With the increase of luminance contrast, the visual masking effect is also increased. In order to model the JND threshold caused by the luminance contrast, a subjective viewing test is set to measure the visibility threshold under different luminance contrasts. The output is shown as the hollow square points in Fig. 6. By fitting these points [20, 23], the visibility threshold of the luminance contrast masking (CM) is acquired,

$$CM(x) = 0.115 \cdot LC(x). \quad (4)$$

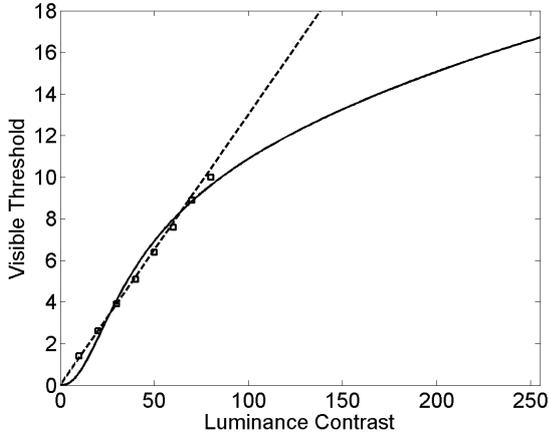


Fig. 6: Contrast masking effect [2]: the visibility threshold is increased with the increase of luminance contrast.

The fitting result from Eq. (4) is shown as the dash line in Fig. 6, which demonstrates that the visibility threshold of the contrast masking is increased with a fixed ratio. As a result, the JND threshold for the contrast region with a high edge height is always overestimated [21]. Research on human perception reported that the response of the HVS to contrast gain is nonlinear (i.e., logarithmic [44]), whose increasing ratio will be decreased with the increase of luminance contrast [14]. To this end, the contrast masking is calculated with a nonlinear transducer in [2],

$$CM(x) = 0.115 \times \frac{\alpha LC(x)^{2.4}}{LC(x)^2 + \beta^2}. \quad (5)$$

where α and β are two parameters which determine the shape of the logarithmic curve. By fitting Eq. (5) with these points (from the subjective contrast masking test), the values for the two parameters are acquired, i.e., $\alpha = 16$ and $\beta = 26$ [2]. The output of Eq. (5) is shown as the solid curve in Fig. 6.

2.3 Pattern Masking

As we have mentioned in the above subsection, the spatial masking effect is a complicated visual phenomena, which is determined by many factors [17, 45–47]. For simplicity, the spatial masking effect is usually calculated as contrast masking effect. However, if we only consider the contrast masking for spatial masking estimation, we may not always acquire the right result. As an example, two representative composite maps are shown in Fig. 7. According to Eq. (5) (or Eq. (4)), image regions with high luminance contrast values have strong spatial masking effect, while image

regions with low luminance contrast values have weak spatial masking effect. Since Fig. 7 (a) has higher edge height than that of Fig. 7 (b), the left one has stronger spatial masking effect than the right one according to the contrast masking function. This result is obviously in conflict with our subjective perception. The interaction among contents in Fig. 7 (b) are much more complex than that in Fig. 7 (a), and it is obvious that the spatial masking effect of Fig. 7 (b) is much stronger than that of Fig. 7 (a). Therefore, we can not simply estimate the spatial masking only with luminance contrast.

In [2, 23], Wu et al. suggested that we should take not only luminance contrast but also structural uncertainty into account for spatial masking estimation. As shown in Fig. 7, though the left map (composed with oblique bars) has higher luminance contrast value, its structure is very regular and is with limited uncertainty. According to the Free-energy principle [48], the HVS can easily predict its structure and can fully understand its visual content. With limited structural uncertainty, the spatial masking effect in Fig. 7 (a) is weak. While for the right image in Fig. 7, it is composed with random noise, which presents low luminance contrast and high structural uncertainty. When we perceiving such map, it is hard to extract its structural rule and can hardly understand its visual content. With high structural uncertainty, Fig. 7 (b) has strong spatial masking effect and high visibility threshold.

In order to calculate the structural uncertainty, an input image is firstly decomposed into two parts, namely, orderly and disorderly portions. Inspired by the Bayesian brain theory [49], the content of the image is actively predicted with an autoregressive model [50, 51],

$$I'(x) = \sum_{x_i \in \mathcal{X}} C_i I(x_i) + \varepsilon, \quad (6)$$

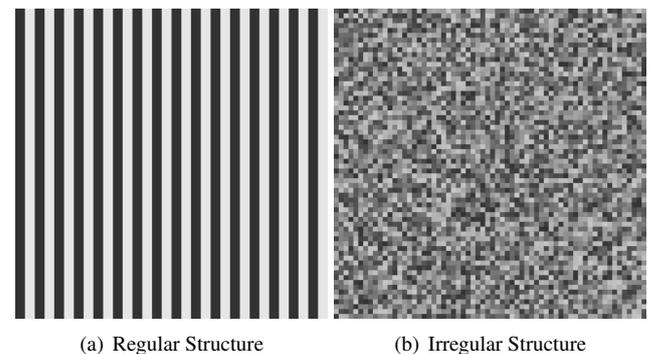


Fig. 7: Two representative composite maps for spatial masking illustration.

where \mathcal{I}' is the orderly portion of image \mathcal{I} , C_i is the normalized weighting coefficient, and ε is the random noise. The disorderly portion of image \mathcal{I} is regarded as the prediction residual $\Delta\mathcal{I} = \mathcal{I} - \mathcal{I}'$.

Then the structural uncertainty is computed with the disorderly portion $\Delta\mathcal{I}$. The local binary pattern (LBP) [52] procedure is adopted to analyze the structural characteristic of the disorderly portion. For pixel x , its LBP value is calculated as the relationship with its symmetric surroundings,

$$\text{LBP}(x) = \sum_{k=1}^N S(\Delta\mathcal{I}(x) - \Delta\mathcal{I}(x_k))2^{k-1}, \quad (7)$$

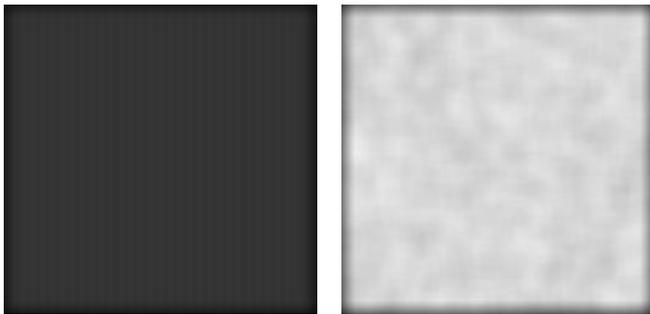
$$S(y_k) = \begin{cases} 1, & y \geq C \\ S(y_{k-1}), & y < C \\ 0, & y \leq -C. \end{cases} \quad (8)$$

where $\Delta\mathcal{I}(x)$ is the intensity value of pixel x , N is the number of pixels in the symmetric surroundings (N is set as 8 in this work), and C is a threshold constant (C is set as 5 in this work).

With Eq. (7), the LBP characteristic for each pixel is acquired. Then, the structural uncertainty of each pixel x is analyzed as the distribution characteristic of pixels in its surrounding \mathcal{X} . In [2], the structural uncertainty (SU) is calculated as the entropy of the LBP features of \mathcal{X} ,

$$\text{SU}(x) = \sum_{i=1}^{2^p} -p_i(x) \log p_i(x), \quad (9)$$

where $p_i(x)$ is the probability at the i -th bin of \mathcal{X} . In this work, the size of \mathcal{X} is set as 21 [2]. Therefore, the probability value of $p_i(x)$ is calculated as the ratio of the i -th bin of the histogram which is mapping from the 21×21 neighborhood centered at x . The structural uncertainty



(a)

(b)

Fig. 8: LBP based structural uncertainty masks for the two composite maps as shown in Fig. 7.

masks for the two representative composite maps (shown in Fig. 7) are calculated with Eq. (9). As shown in Fig. 8, the orderly content (Fig. 7 (a)) has low structural uncertainty value (i.e., the dark mask shown in Fig. 8 (a)), while the disorderly content (Fig. 7 (b)) has high structural uncertainty value (i.e., the bright mask shown Fig. 8 (b)).

2.4 Contrast Sensitive Function

The HVS presents band-pass character for visual signal perception in the spatial frequency domain, which is highly sensitive to signals with modulate changing frequency and is insensitive to those with other frequencies (especially for the high frequency). Thus, the visibility threshold (\mathcal{T}) of the HVS is directly related to the spatial frequency (w),

$$\mathcal{T}(w) = f(w), \quad (10)$$

In the past decades, the sensitivity of the HVS for different spatial frequencies had been thoroughly investigated, and various models (i.e., contrast sensitive function, CSF) have been proposed [53, 54]. In [25], the CSF based visibility threshold is defined as,

$$\mathcal{T}(w) = \frac{\exp(cw)}{a + bw}, \quad (11)$$

where a , b , and c are three constant coefficients ($a=1.33$, $b=0.11$, and $c=0.18$ in [25]), w is the spatial frequency. In the DCT domain, the spatial frequency for the (x,y) subband of each $N \times N$ DCT block is computed as,

$$w_{x,y} = \frac{1}{2N} \sqrt{(x/\theta_h)^2 + (y/\theta_v)^2}, \quad (12)$$

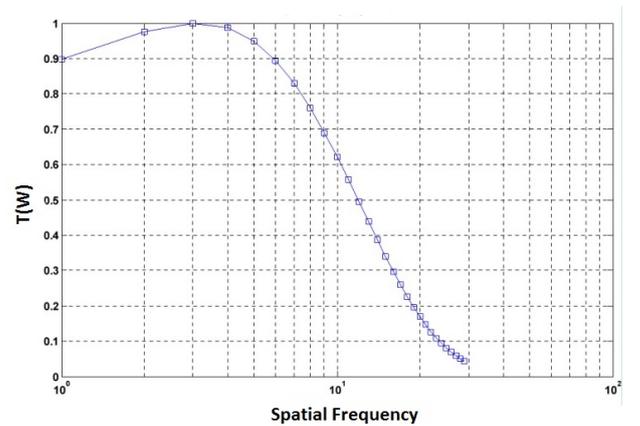


Fig. 9: CSF curve: visibility thresholds for different spatial frequencies.

where θ_h and θ_v are the angles along horizontal and vertical, which can be calculated as,

$$\theta_x = 2 \cdot \arctan \frac{S_x}{2d}, \quad \theta_y = 2 \cdot \arctan \frac{S_y}{2d} \quad (13)$$

where S_x is the width of the display, S_y is the length of the display, and d is the viewing distance. An example of the CSF based visibility thresholds for different frequencies is shown in Fig. 9.

3 Existing JND computation models

According to the domain for the visibility threshold to be computed, the existing JND models are usually classified into two categories, namely, the spatial domain and subband domain JND estimation models. Detailed computational processes for the two categories of models are given in the following.

3.1 Pixel domain JND computation

Pixel domain JND models are popular in motion estimation, image enhancement, quality assessment, and so on. For pixel domain JND computation, the luminance adaptation, contrast masking, and structural uncertainty modules are usually considered. During the past decades, many pixel domain JND models have been proposed [2, 20–22]. We will give a brief review about these models.

In [20], a pioneer pixel domain JND model was proposed. By considering the effects from both luminance adaptation and contrast masking, the JND threshold for each pixel is calculated as [20]:

$$\text{JND}(x) = \max\{\text{LA}(x), \text{CM}(x)\}, \quad (14)$$

where LA is the luminance adaptation from Eq. (1), and CM is the contrast masking from Eq. (4).

With the above computational model, the JND threshold for the edge region is always overestimated, and that for the texture region is always underestimated. In order to suppress the JND threshold of the edge region with high contrast value, the Canny edge protector (E_p) is adopted in [21],

$$\text{CM}_p(x) = \text{CM}(x) \cdot E_p(x). \quad (15)$$

And then, the JND threshold is calculated as an interacting combination of the luminance adaptation and the contrast masking [21],

$$\text{JND}(x) = \text{LA}(x) + \text{CM}_p(x) - c \cdot \min\{\text{LA}(x), \text{CM}_p(x)\}, \quad (16)$$

where c is a constant parameter which accounts for the overlapping between LA and CM_p , and is set as $c = 0.25$ in this work [21].

Besides, in order to enhance the JND threshold of the texture region, Liu et al. [22] suggested to firstly separate image regions into texture and non-texture ones for JND estimation. However, there is still no rigorous definition for texture, and it is difficult to accurately separate texture regions from non-texture regions for different images.

Recently, Wu et al. suggested that the HVS is insensitive to regions with disorderly texture (e.g., Fig. 7 (b)), rather than all texture regions (actually, the HVS is sensitive to orderly texture, as an example shows in Fig. 7 (a)) [23]. And structural uncertainty is another important factor which determines the JND threshold. Thus, a new JND model is introduced, which takes background luminance, edge contrast, and structural uncertainty into account [2],

$$\text{JND}(x) = f(B(x), \text{LC}(x), \text{SU}(x)), \quad (17)$$

where SU is the degree of structural uncertainty from Eq. (9).

3.2 Subband domain JND computation

The subband domain JND model is popular for signal compression. For the subband domain JND computation, images are firstly transferred into the subband (e.g., DCT) domain. Then, the CSF, luminance adaptation, and contrast masking are considered for each subband. We will give a brief review about the existing subband domain JND models in the following.

With the CSF, the base/fundamental JND threshold in the DCT domain is calculated as [24, 25],

$$\text{JND}_B(n, x, y) = \frac{1}{4\phi_x\phi_y} \cdot \frac{\mathcal{T}(w_{xy})}{0.6 + 0.4 \cdot \cos^2 \varphi_{xy}} \quad (18)$$

where $\mathcal{T}(w_{xy})$ is the CSF based visibility threshold from Eq. (11), ϕ_x and ϕ_y are normalization factors, which can be calculated as:

$$\phi_i = \begin{cases} \sqrt{\frac{1}{N}} & i = 0 \\ \sqrt{\frac{2}{N}} & i > 0 \end{cases} \quad (19)$$

and φ_{xy} is the directional angle, which is related to the frequency of its corresponding DCT subband,

$$\varphi_{x,y} = \arcsin \left(\frac{2 \cdot w_{x0} \cdot w_{0y}}{w_{xy}^2} \right) \quad (20)$$

By adjusting the base JND (Eq. (18)) with the affection from the luminance contrast and the background luminance,

a DCT domain JND estimation model was proposed [24],

$$\text{JND}(n, x, y) = \text{JND}_B(n, x, y) \cdot F_{LA}(n, x, y) \cdot F_{CM}(n, x, y) \quad (21)$$

where F_{LA} and F_{CM} are the adjustment factors from the luminance adaption and the contrast masking, respectively [24]

The adjustment factor F_{LA} is determined by the background luminance,

$$F_{LA}(n, x, y) = \begin{cases} \frac{30-B(n)}{150} + 1 & B(n) \leq 60 \\ 1 & 60 < B(n) < 170 \\ \frac{B(n)-170}{425} + 1 & B(n) \geq 170 \end{cases} \quad (22)$$

where $B(n)$ is the average intensity value of the block n .

The adjustment factor F_{CM} (in Eq. 21) is determined by the contrast masking [25]. Generally, the contrast masking in the texture region is stronger than that in the plane and edge regions, and thus, a larger F_{CM} value is given for the texture region than the other two types of regions (more detail about the adjustment factor F_{CM} setting can be found in [24, 25]).

Moreover, with further consideration of luminance contrast in both inter and intra bands, a more accurate contrast masking procedure was proposed for DCT domain JND estimation [26]. In order to extend the JND threshold into color images, the chroma character was investigated and a temporal color JND estimation model was proposed in [27].

4 Subjective viewing test for JND model verification

In order to verify the effectiveness of these JND models, the subjective viewing test is set to examine the accuracy of the computed visibility threshold. For a given image, its corresponding visibility threshold is firstly calculated with the JND model. Then, Gaussian White Noise is injected into the original image with the guide of the computed JND threshold for quality comparison. For a given image, its corresponding JND guided contaminated image is shaped as [20]:

$$\hat{I}(x) = I(x) + \beta \cdot \text{rand}(x) \cdot \text{JND}(x), \quad (23)$$

where \hat{I} is the contaminated image with JND guided noise, β regulates the energy of JND guide noise, and $\text{rand}(x)$ randomly takes +1 or -1.

Generally, a more accurate JND model outputs higher visibility threshold for a region which is insensitive to the HVS, while lower visibility threshold for a region which is

Table 1: The quantitative scores for quality comparison

Description	Score
Indistinguishable	0
Slightly Better	1
Obviously Better	2
Far Better	3

sensitive to the HVS. From the perspective of Eq. (23), a more accurate JND model will distribute much more noise into the insensitive region and less into the sensitive region. Therefore, if we inject the same level of noise into an original image, a more accurate JND model will return a better quality, and a less accurate one will return a worse quality.

An example of JND model comparison is shown in Fig. 10. With the guidance of Eq. (23), the white noise is injected into the original image (Fig. 10). Five different JND models are chosen for comparison, namely, Chou et al.'s model (Chou95) [20], Yang et al.'s model (Yang05) [21], Zhang et al.'s model (Zhang08) [26], Wei et al.'s model (Wei09) [25], and Wu et al.'s model (Wu13) [2]. Moreover, three of them (i.e., Chou95, Yang05, and Wu13) are pixel domain JND models, and the other two (i.e., Zhang08 and Wei09) are subband domain JND models. With the help of β in Eq. (23), the energies of noise in the five contaminated images are adjusted into a same level (with PSNR=27.72 dB).

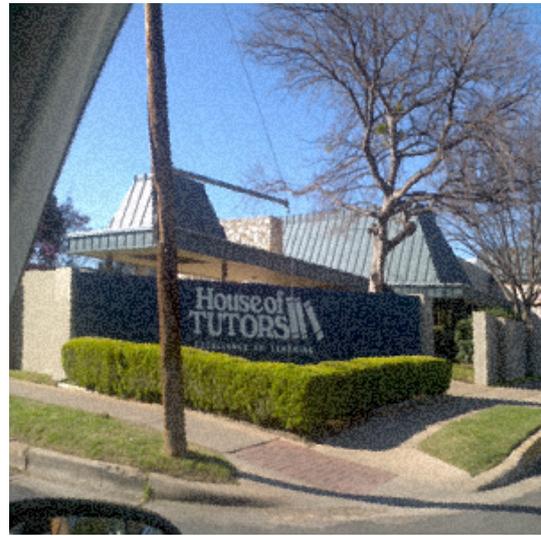
By comparing the outputs of the JND models with the original image, we can see that: 1) the disturbance in the smooth region is visible in Fig. 10 (b) (the output of Chou et al.'s model) and (c) (the output of Yang et al.'s model); 2) the distortion in the structure region with large luminance change (especially for the letter region) is obvious in Fig. 10 (d) (the output of Zhang et al.'s model) and (e) (the output of Wei et al.'s model); 3) and the noise in Fig. 10 (f) is almost invisible. In other words, Fig. 10 (f) has a better quality than the other contaminated images (i.e., Fig. 10 (b)-(e)).

Finally, the accuracy of the computed visibility threshold is verified with subjective viewing test. As shown in Fig. 11, two images are projected onto a screen for quality comparison. The condition setting for the subjective viewing test (e.g., the viewing distance and environment) follows the ITU-R BT.500-11 standard [40]. During each test, subjects are required to evaluate which one (left or right image) is with better quality, and how much better it is. The quantitative quality scores are given in Tab. 1.

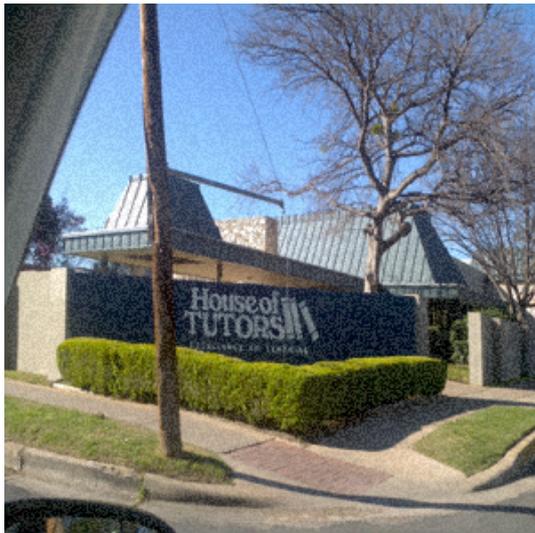
Following the rule mentioned above for subjective



(a) Original Image



(b) Chou95 [20]



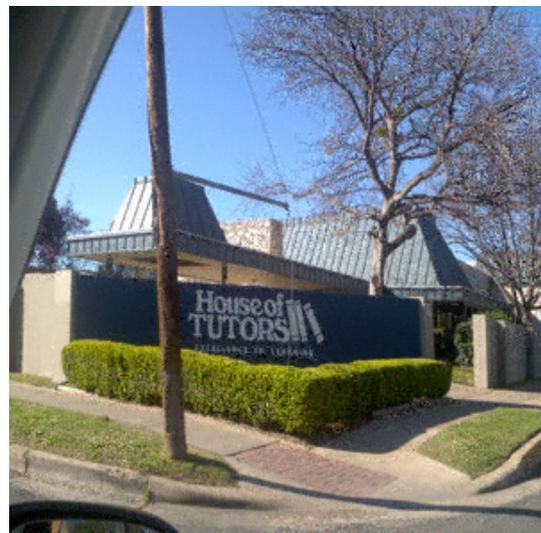
(c) Yang05 [21]



(d) (Zhang08 [26])



(e) Wei09 [25]



(f) Wu13 [2]

Fig. 10: JND model comparison. Under a same level of noise (PSNR=27.72), different JND models (i.e., Chou95, Yang05, Zhang08, Wei09, and Wu13) output contaminated images with different qualities.



Fig. 11: Subjective viewing test for different JND model comparison.

viewing test, the original image (Fig. 10 (a)) and one of the contaminated images (Fig. 10 (b)-(f)) are projected onto a screen for quality assessment, as shown in Fig. 11. Twenty six subjects are invited for the test. The comparison results (mean score) between the original image and the five contaminated (from Fig. 10 (b) to (f)) are 1.13, 1.38, 1.01, 0.76, and 0.34 (a small value means the noise in the contaminated image is less visible), respectively.

5 Further Discussion

The JND threshold reveal the visibility limitation of the HVS, which is useful for perceptual based visual signal processing. As an example, we are living in a multimedia era surrounded by big data, which possesses a large amount of redundancy to be removed [55, 56]. with the help of the JND threshold, the visual redundancy can be estimated, and then removed from the original signal for compression. As shown in Fig. 12, the visual redundancy of the original image (Fig. 12 (a)) is firstly estimated with its corresponding JND threshold of each pixel with Eq. (17). A direct view of the visual redundancy map for the original image is shown in Fig. 12 (b). By removing the visual redundancy for image preprocessing, we can improve the performance of the perceptual coding algorithms [57],

$$\tilde{I}(x) = \begin{cases} I(x) + \text{JND}(x), & \text{if } I(x) - \bar{I}_B < -\text{JND}(x) \\ \bar{I}_B, & \text{if } |I(x) - \bar{I}_B| \leq \text{JND}(x) \\ I(x) - \text{JND}(x), & \text{if } I(x) - \bar{I}_B > \text{JND}(x) \end{cases} \quad (24)$$

where \tilde{I} is the preprocessed image, and \bar{I}_B is the mean value for each coding block. Fig. 12 (c) shows the preprocessed image \tilde{I} with redundancy removing. By comparing Fig. 12 (c) with the original image (Fig. 12 (a)), it is hard to sense any difference. However, we can save about 16% bit rate with Fig. 12 (c) under JPEG compression (with QP=2) [23].

Though the JND technology has achieved success in many perception oriented visual signal processing systems, there still has room for further improvement. As the dominated factor which determines the JND threshold, estimation of the visual masking effect is a complicated procedure, which is caused by both physical and psychological phenomena. In the existing JND models, the luminance contrast masking is firstly adopted to simply estimate the visual masking effect [20–22]. In the recent years, the pattern masking effect, which takes both luminance contrast and structural uncertainty into account, is introduced for visual masking estimation [2, 23].

More factors should be taken into account for the complicated visual masking estimation, e.g., memory based visual content prediction [58–60], visual attention [61–63], and so on. It is well known that we usually present bias with the help of memory for visual perception. In other words, the priori knowledge will affect the sensitivity of the HVS to the input visual contents, especially for these objects that are extremely familiar to us (e.g., human face). For visual attention, the sensitivity of the attended region is always higher than the other regions. Thus, visual attention will increase the sensitivity and reduce the JND threshold for the attended region. By considering more factors which affect the visual masking, a more accurate JND model is expected to be deduced.

6 Conclusion

The JND accounts for the visibility limitation of the human visual system, below which the change is invisible to the human eye. The JND threshold reveals the visual redundancy, and thus is useful for perception oriented visual signal processing, e.g., perceptual signal compression, image/video enhancement, information hiding, and so on.

In this work, we have highlighted the importance and the challenges in designing a JND model which performs consistently with the human perception. Though JND estimation is a challenge task, researches have made a lot of efforts on JND modeling, and we have surveyed these modules (e.g., luminance adaptation, contrast masking,

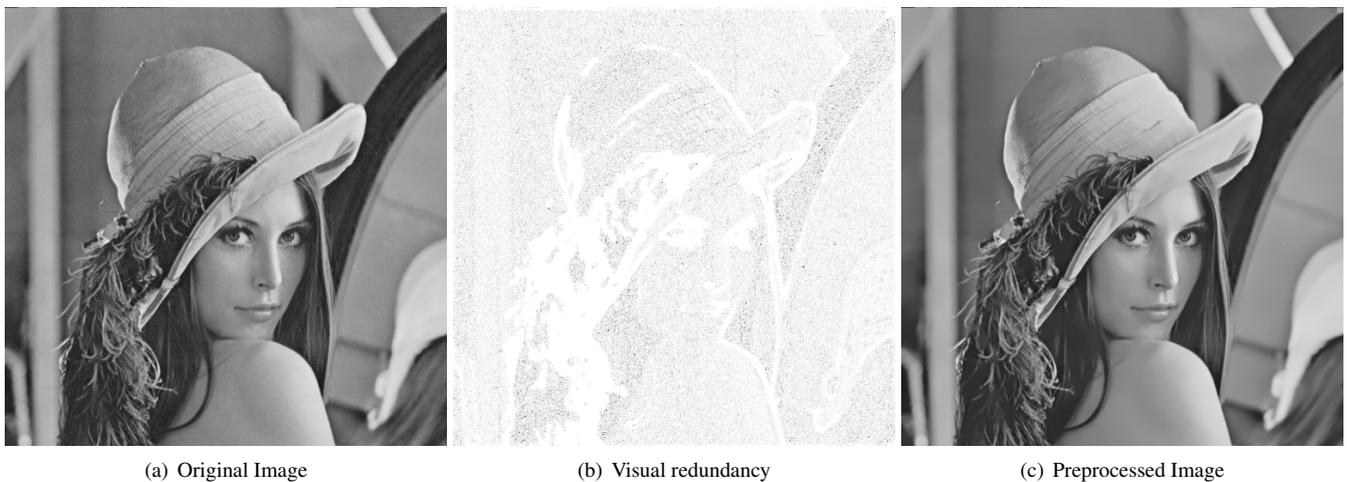


Fig. 12: Visual redundancy removal with JND threshold for image compression.

pattern masking, CSF) which effect the visibility threshold and these existing JND models (pixel domain and subband domain models). Finally, the applications of JND have been demonstrated and the remaining challenges in visual masking are discussed.

7 Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61401325), the Research Fund for the Doctoral Program of Higher Education (No. 20130203130001), and the Young Talent fund of University Association for Science and Technology in Shaanxi (No. 20150110).

References

- Jayant N, Johnston J, Safranek R. Signal compression based on models of human perception. *Proceedings of the IEEE*, 1993, 81(10): 1385–1422
- Wu J, Lin W, Shi G, Wang X, Li F. Pattern masking estimation in image with structural uncertainty. *IEEE Transactions on Image Processing*, 2013, 22(12): 4892–4904
- Yang X, Lin W, Lu Z, Lin X, Rahardja S, Ong E, Yao S. Rate control for videophone using local perceptual cues. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(4): 496–507
- Ji T L, Sundareshan M K, Roehrig H. Adaptive image contrast enhancement based on human visual properties. *IEEE Transactions on medical imaging*, 1994, 13(4): 573–586
- Xuan DONG J W. A pixel-based outlier-free motion estimation algorithm for scalable video quality enhancement. *Frontiers of Computer Science*, 2015, 9(5): 729
- Lin W, Kuo C J. Perceptual visual quality metrics: A survey. *J. Visual Communication and Image Representation*, 2011, 22(4): 297–312
- CUI L. Swvfs: a saliency weighted visual feature similarity metric for image quality assessment. *Frontiers of Computer Science*, 2014, 8(1): 145
- Li W, Yang C, Li C, Yang Q. Jnd model study in image watermarking. In: Jin D, Lin S, eds, *Advances in Multimedia, Software Engineering and Computing Vol.2*, number 129 in *Advances in Intelligent and Soft Computing*, 535–543. Springer Berlin Heidelberg, 2011
- Chou C H, Liu K C. A perceptually tuned watermarking scheme for color images. *IEEE Transactions on Image Processing*, 2010, 19(11): 2966–2982
- Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Transactions on Information Forensics and Security*, 2016, 11(11): 2594–2608
- Cheng Q, Huang T S. An additive approach to transform-domain information hiding and optimum detection structure. *Trans. Multi.*, 2001, 3(3): 273–284
- Fu Z, Ren K, Shu J, Sun X, Huang F. Enabling personalized search over encrypted outsourced data with efficiency improvement. *IEEE Transactions on Parallel and Distributed Systems*, 2016, 27(9): 2546–2559
- Xia Z, Wang X, Sun X, Liu Q, Xiong N. Steganalysis of lsb matching using differences between nonadjacent pixels. *Multimedia Tools and Applications*, 2016, 75(4): 1947–1962
- Legge G E, Foley J M. Contrast masking in human vision. *Journal of the Optical Society of America*, 1980, 70(12): 1458–1471
- Daly S J. Visible differences predictor: an algorithm for the assessment of image fidelity. *Proceedings of SPIE*, 1992, 1666(1): 2–15
- Foley J M. Human luminance pattern-vision mechanisms: masking

- experiments require a new model. *Journal of the Optical Society of America A*, 1994, 11(6): 1710–1719
17. Kovács G, Vogels R, Orban G A. Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences*, 1995, 92(12): 5587–5591
 18. Watson A B, Solomon J A. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 1997, 14(9): 2379–2391
 19. Daly S J. Engineering observations from spatiovelocity and spatiotemporal visual models. *Proceedings of SPIE*, 1998, 3299(1): 180–191
 20. Chou C H, Li Y C. A perceptually tuned subband image coder based on the measure of just-noticeable distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology*, 1995, 5(6): 467–476
 21. Yang X K, Ling W S, Lu Z K, Ong E P, Yao S S. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 2005, 20(7): 662–680
 22. Liu A, Lin W, Paul M, Deng C, Zhang F. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(11): 1648–1652
 23. Wu J, Shi G, Lin W, Liu A, Qi F. Just noticeable difference estimation for images with free-energy principle. *IEEE Transactions on Multimedia*, 2013, 15(7): 1705–1710
 24. Jia Y, Lin W, Kassim A. Estimating just-noticeable distortion for video. *IEEE Transactions on Circuits and Systems for Video Technology*, 2006, 16(7): 820–829
 25. Wei Z, Ngan K. Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 2009, 19(3): 337–346
 26. Zhang X, Lin W, Xue P. Just-noticeable difference estimation with pixels in images. *Journal Visual Communication and Image Representation*, 2008, 19(1): 30–41
 27. Chen H, Hu R, Hu J, Wang Z. Temporal color just noticeable distortion model and its application for video coding. In: 2010 IEEE International Conference on Multimedia and Expo (ICME). July 2010, 713–718
 28. Ma L, Ngan K N, Zhang F, Li S. Adaptive block-size transform based just-noticeable difference model for images/videos. *Signal Processing: Image Communication*, 2011, 26(3): 162–174
 29. Bae S H, Kim M. A novel dct-based jnd model for luminance adaptation effect in dct frequency. *IEEE Signal Processing Letters*, 2013, 20(9): 893–896
 30. Bae S H, Kim M. A DCT-based total JND profile for spatio-temporal and foveated masking effects. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016
 31. Rovamo J, Mustonen J, Näsänen R. Modelling contrast sensitivity as a function of retinal illuminance and grating area. *Vision Research*, 1994, 34(10): 1301–1314
 32. Safranek R J, Johnston J D. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression. In: *International Conference on Acoustics, Speech, and Signal Processing*. May 1989, 1945–1948
 33. Moon P, Eberle Spencer D. The visual effect on non-uniform surrounds. *Journal of the Optical Society of America*, 1945, 35: 233
 34. Netravali A N, Prasada B. Adaptive quantization of picture signals using spatial masking. *Proceedings of the IEEE*, 1977, 65: 536–548
 35. Wu J, Shi G, Lin W, Kuo C C J. Enhanced just noticeable difference model with visual regularity consideration. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. March 2016, 1581–1585
 36. Wang S, Ma L, Fang Y, Lin W, Ma S, Gao W. Just noticeable difference estimation for screen content images. *IEEE Transactions on Image Processing*, 2016, 25(8): 3838–3851
 37. Pan Z, Lei J, Zhang Y, Sun X, Kwong S. Fast motion estimation based on content property for low-complexity H.265/HEVC encoder. *IEEE Transactions on Broadcasting*, 2016, 62(3): 675–684
 38. Bae S H, Kim M. A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images. *IEEE transactions on image processing*, 2014, 23(8): 3227–3240
 39. Pan Z, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding. *IEEE Transactions on Broadcasting*, 2015, 61(2): 166–176
 40. Method for the subjective assessment of the quality of television pictures. ITU, Document ITU-R BT.500-11, 2002. Geneva, Switzerland
 41. Wu H R, Reibman A R, Lin W, Pereira F, Hemami S S. Perceptual visual signal compression and transmission. *Proceedings of the IEEE*, 2013, 101(9): 2025–2043
 42. Jarsky T, Cembrowski M, Logan S M, Kath W L, Riecke H, Demb J B, Singer J H. A synaptic mechanism for retinal adaptation to luminance and contrast. *The Journal of Neuroscience*, 2011, 31(30): 11003–11015
 43. Netravali A N, Haskell B G. *Digital Pictures: Representation, Compression and Standards*, 2nd edition. New York: Plenum Press, January 1995
 44. Jourlin M, M.Carré, Breugnot J, M.Bouabdellah. Logarithmic image processing: Additive contrast, multiplicative contrast, and associated metrics. *Advances in Imaging and Electron Physics*, 2012, 171: 357–406
 45. Foley J M, Boynton G M. New model of human luminance pattern vision mechanisms: analysis of the effects of pattern orientation, spatial phase, and temporal frequency. *Proceedings of SPIE*, 1994, 2054(1): 32–42
 46. Truchard A M, Ohzawa I, Freeman R D. Contrast gain control in the visual cortex: monocular versus binocular mechanisms. *Journal of Neuroscience*, 2000, 20(8): 3017–3032
 47. Zhou Z, Wang Y, Wu Q J, Yang C N, Sun X. Effective and efficient global context verification for image copy detection. *IEEE Transactions on Information Forensics and Security*, 2016
 48. Friston K. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 2010, 11(2): 127–138
 49. Knill D C, Pouget R. The bayesian brain: the role of uncertainty in neural coding and computation. *Trends In Neuroscience*, 2004, 27: 712–719
 50. Zhang X, Wu X. Image interpolation by adaptive 2-D autoregressive

- modeling and Soft-Decision estimation. *IEEE Transactions on Image Processing*, 2008, 17(6): 887–896
51. Wu J, Lin W, Shi G, Liu A. Perceptual quality metric with internal generative mechanism. *IEEE Transactions on Image Processing*, 2013, 22(1): 43–54
 52. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971–987
 53. Nill N B. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Transactions on Communications*, 1985, 33(3): 551–557
 54. Ngan K N, Leong K S, Singh H. Adaptive cosine transform coding of images in perceptual domain. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1989, 37(11): 1743–1750
 55. Fu Z, Sun X, Liu Q, Zhou L, Shu J. Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing. *IEEE Transactions on Communications*, 2015, E98-B(1): 190–200
 56. Ren Y, Shen J, Wang J, Han J, Lee S. Mutual verifiable provable data auditing in public cloud storage. *Journal of Internet Technology*, 2015, 16(2): 317–323
 57. Yang X, Lin W, Lu Z, Ong E, Yao S. Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(6): 742–752
 58. Downing P E. Interactions between visual working memory and selective attention. *Psychological Science*, 2000, 11(6): 467–473
 59. Bar M. The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 2009, 364(1521): 1235–1243
 60. Chaumon M, Kveraga K, Barrett L F, Bar M. Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex (New York, N.Y.: 1991)*, 2014, 24(11): 2899–2907
 61. Koch C, Ullman S. Shifts in selection in visual attention: Toward the underlying neural circuitry. *Human Neurobiology*, 1985, 4(4): 219–227
 62. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, 20(11): 1254–1259
 63. Itti L, Koch C. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2001, 2(3): 194–203