Pattern Masking Estimation in Image with Structural Uncertainty

Jinjian Wu, Weisi Lin, *Senior Member, IEEE*, Guangming Shi, *Senior Member, IEEE*, Xiaotian Wang and Fu Li

***Abstract*—A model of visual masking, which reveals the visibility of stimuli in the human visual system (HVS), is useful in perceptual based image/video processing. The existing visual masking function mainly takes luminance contrast into account, which always overestimates the visibility threshold of the edge region and underestimates that of the texture region. Recent research on visual perception indicates that the HVS is sensitive to orderly regions which possess regular structures, and insensitive to disorderly regions which possess uncertain structures. Therefore, structural uncertainty is another determining factor on visual masking. In this paper, we introduce a novel pattern masking function based on both luminance contrast and structural uncertainty. By mimicking the internal generative mechanism of the HVS, a prediction model is firstly employed to separate out the unpredictable uncertainty from an input image. And then, an improved local binary pattern is introduced to compute the structural uncertainty. Finally, combining luminance contrast with structural uncertainty, the pattern masking function is deduced. Experimental result demonstrates that the proposed pattern masking function outperforms the existing visual masking function. Furthermore, we extend the pattern masking function to just noticeable difference (JND) estimation and introduce a novel pixel domain JND model. Subjective viewing test confirms that the proposed JND model is more consistent with the HVS than the existing JND models.**

***Index Terms*—Pattern Masking, Internal Generative Mechanism, Structural Uncertainty, Local Binary Pattern, Human Perception, Just Noticeable Difference**

I. Introduction

The last two decades have witnessed the tremendous growth of digital image/video processing techniques, by which signals are processed, transmitted, stored, and reconstructed for various applications. Since the human eye is the ultimate reviewer of digital signals, researchers hope to improve the processing techniques by considering the characters of the human visual system (HVS). Visual masking [1], which reveals the visibility of stimuli in the HVS, is useful in perceptual based image/video compression [2], scene enhancement [3], quality assessment [4], and so on.

Visual masking is caused by interaction or interference among stimuli [1], [5]. It is a complicated visual perceptual mechanism, which describes a broad range of phenomena [6]. Here we mainly focus on pattern masking, which appears at static images. Pattern masking refers to that one pattern will mask another one. As shown in Fig. 1, the detection and identification of the text (target) will be strongly impaired when it is followed by the hatching background (pattern) [7]. In addition, the pattern masking effect on a uniform background is very weak, on which a spatial target is most easily seen; however, when the background becomes much complex with spatial patterns, the pattern masking effect will be much stronger [8], which will inhibit the target detection obviously. Therefore, pattern masking relies on the visual content of an input scene [9].

As the HVS is highly sensitive to the luminance change of an input scene, researchers always intend to estimate the pattern masking effect based on luminance contrast [5], [10] for simplicity. In [11], a psychophysical experiment is designed to investigate the relationship between luminance edge height and the visibility threshold. Moreover, according to the recording data from [11], a well-accepted contrast masking function is deduced in [12]. However, the contrast masking function only takes the luminance edge height into account, which always overestimates the visibility thresholds of edge regions and underestimates that of texture regions [13]. Some other image features, such as the spatial frequency [14], orientation [8], contours [15], and shapes [16], are further investigated, and experimental results indicate that the visibility threshold is much higher when the content becomes more complex. Therefore, we should not only take luminance contrast into account for visual masking estimation. In [17] and [18], image blocks are firstly classified into three types (i.e., plain, edge, and texture), and then three different weights are set for the three types on visual masking estimation. With the help of a big weight, the texture region is highlighted on the computation of visibility threshold. However, the HVS is highly sensitive to the orderly texture regions [19], and these orderly regions will be overestimated with [17] and [18].

Since the HVS is highly adapted to extract structural in-

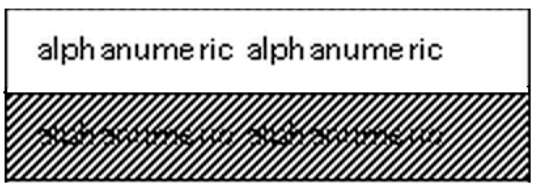


Fig. 1: Example of pattern masking effect.

Jinjian Wu, Guangming Shi, Xiaotian Wang and Fu Li are with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, School of Electronic Engineering, Xidian University. E-mail: jinjian.wu@mail.xidian.edu.cn; gmshi@xidian.edu.cn.

Weisi Lin (corresponding author) is with the School of Computer Engineering, Nanyang Technological University, Nanyang 639798, Singapore. E-mail: wslin@ntu.edu.sg.

This work was supported by the Major State Basic Research Development Program of China (973 Program, No. 2013CB329402), NSF of China (No. 61070138, 61072104, 61100155, and 61227004), Fundamental Research Funds for the Central Universities of China (No.K5051399020, K5051202034, and K50513100005).

formation for image understanding [20], we try to estimate pattern masking by taking image structures into account. Empirical studies show that disorderly texture regions contain more uncertain information and present more disorderly structures than edge regions [21]. Furthermore, recent research on brain theory [22] indicates that the HVS actively predicts the orderly contents of the input visual information and tries to avoid some uncertainties for image perception. As a result, the HVS is less sensitive to disorderly regions (e.g., some texture regions) which possess uncertain structures [21], and the pattern masking effects in these regions are strong [23]. Therefore, pattern masking is related to not only luminance contrast but also structural uncertainty, and we suggest to take both factors into account to create a novel pattern masking function.

However, the computation of structural uncertainty is still an open problem. Images represent various structures due to variations in orientation, scale, frequency, and other visual appearance [24]. By considering these features, a famous local binary pattern (LBP) algorithm is introduced in [25] to analyze the structural information. But, structural information is unequal to structural uncertainty, because the HVS can understand most of the orderly structural information and only the residual represents structural uncertainty [19]. Meanwhile, the Bayesian brain theory [26] further indicates that the HVS possesses an internal generative mechanism (IGM), within which the content of the input scene is actively predicted and some unpredictable information (i.e., residual of the prediction) is avoided for understanding. Therefore, we suggest to consider the unpredictable information as the uncertainty. By mimicking the active prediction in the IGM, an autoregressive model [19] is employed to separate the disorderly uncertainty from an input scene. And then, an improved LBP algorithm is introduced to compute the structural uncertainty on the disorderly uncertainty.

Finally, combining structural uncertainty with luminance contrast, a computational function for pattern masking is deduced. Furthermore, to demonstrate the effectiveness of the proposed pattern masking, we extend the proposed pattern masking to estimate the just noticeable difference (JND, which accounts for such a visibility threshold and below which the change cannot be detected by the majority (e.g., 75%) of viewers [12]). Since most of the existing JND models estimate the spatial masking effect based on contrast masking, they always underestimate the JND thresholds for these places with uncertain structures (e.g., the texture regions) [13], [21]. Therefore, we replace contrast masking with pattern masking and introduce a novel JND model. With the help of the proposed pattern masking function, the JND thresholds for the sensitive and insensitive regions can be accurately computed. Experimental results from subjective viewing tests confirm that the proposed JND model correlates better with the HVS than the existing JND models [1].

The organization of this paper is as follows: the structural uncertainty of an image is analyzed and estimated in Section II. And then, in Section III, by taking both luminance contrast and structural uncertainty into account, a novel pattern masking function is deduced. Experimental results of the proposed pattern masking function is presented in Section IV. Finally, conclusions are drawn in Section V.

[1]The source code is avaliable at http://web.xidian.edu.cn/wjj/en/index.html

II. Structural Uncertainty

In this section, visual character is firstly considered for image structure analysis, and an improved LBP algorithm is introduced for structural information computation. Then, by mimicking the active prediction of the HVS, the uncertain information is separated from an input image. Finally, the structural character of the uncertain information is analyzed with the improved LBP algorithm to acquire structural uncertainty.

A. Structural Information Analysis

Image structures convey the primary visual information of a scene, and the HVS is highly adapted to extract them for image perception and understanding [19], [20]. Therefore, structural information is usually measured for quality assessment [20], texture classification [25], image denoising [27] and deblurring [28]. For simplicity, image structures are analyzed with some statistical values [20], such as variance and covariance, which are effective to represent the luminance change but not good enough to represent the spatial distribution of structural information [29]. To this end, Ojala et al. [25] analyzed the spatial relationship among pixels, and introduced a classic LBP algorithm, within which a joint difference distribution is adopted to represent the structural characteristic. For a given pixel x_c, the structural characteristic (T) is always analyzed with its circularly symmetric neighborhood x_i ($i = 1, 2, ..., p$) [25], [29],

$$\mathrm{T}(g_c) = t(g_1 - g_c, g_2 - g_c, \cdots, g_p - g_c), \tag{1}$$

where $t(\cdot)$ represents the joint difference distribution, g_c corresponds to the gray value of the central pixel x_c, g_i corresponds to the gray value of pixel x_i in the local neighborhood, and the neighborhood size p is always set as 8 (by considering the accuracy and the computational complexity) [25], [29].

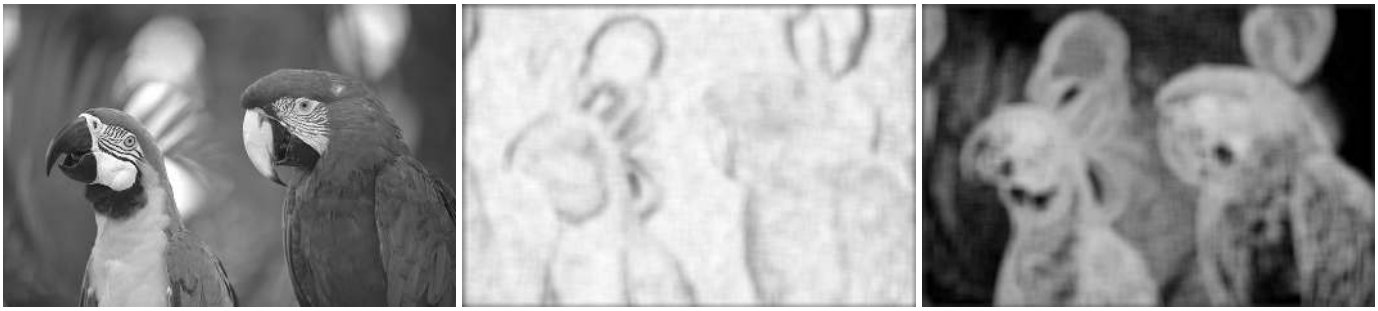
In order to be invariant against gray-scale shift, the signs of differences are adopted to replace the exact difference values in T [24], [25],

$$\mathrm{T}(g_c) \approx t\left(s(g_1 - g_c), s(g_2 - g_c), \cdots, s(g_p - g_c)\right), \tag{2}$$

where the sign is defined as

$$s(g_i - g_c) = \begin{cases} 1, & g_i - g_c \geq 0 \\ 0, & g_i - g_c < 0. \end{cases} \tag{3}$$

However, (3) is too sensitive to the gray value change. According to the subjective experiment on luminance adaptation [30], the HVS cannot sense some small change on gray value. For example, as shown in Fig. 2 (a), if some background regions are almost uniform to our eye, their structural measures should be to zero. However, with (3), any tiny change (which is too small to be sensed by the HVS) on gray value will be counted for structural information computation. As a result,



(a) (b) (c)

Fig. 2: LBP based structural information. (a) The original image. (b) Structural information based on the original LBP (which is acquired from (7) with the original sign equation (3)). (c) Structural information based on the improved LBP (which is acquired from (7) with the improved sign equation (5)).

the computed values of structural information of the uniform background regions based on LBP are overestimated, as shown in Fig. 2 (b), and this is not consistent with human perception.

In order to perform consistently with the human perception, the luminance adaptation effect should be taken into account when computing the structural characteristic. To this end, we adopt the luminance adaptation threshold to improve (3). Subjective perception experiments demonstrate that the HVS is insensitive to dark/bright background, and is highly sensitive to moderate luminance (e.g., for digital images with 256 gray level, the HVS is insensitive to the background around 0 or 255, and sensitive to that around 127). According to the data from a subjective viewing test [12], the luminance adaptation threshold (LA) is computed as follows [13], [21],

$$\mathrm{LA}(x_c) = \begin{cases} 17 \times (1 - \sqrt{\frac{B(x_c)}{127}}), & \text{If } B(x_c) \leq 127 \\ \frac{3}{128} \times (B(x_c) - 127) + 3, & \text{else,} \end{cases} \tag{4}$$

where $B(x_c)$ is the background luminance of pixel x_c, i.e., the mean luminance of an image region (e.g., a 3×3 neighborhood).

According to the definition of luminance adaptation, if $|g_i - g_c| < \mathrm{LA}(x_c)$, the HVS cannot sense the difference between the two pixels; therefore, the corresponding spatial structure to the HVS is uniform, and we suggest the sign of $g_i - g_c$ to be consistent with the sign of the prior one. In this paper, we calculate the sign of the first neighboring point with (3) (i.e., $s'(g_1 - g_c) = s(g_1 - g_c)$), and calculate the rest (i.e., $i = 2, ..., p$) as follows,

$$s'(g_i - g_c) = \begin{cases} 1, & g_i - g_c \geq \mathrm{LA}(x_c) \\ s'(g_{i-1} - g_c), & |g_i - g_c| < \mathrm{LA}(x_c) \\ 0, & g_i - g_c \leq -\mathrm{LA}(x_c). \end{cases} \tag{5}$$

And then, by assigning a binomial factor 2^p for each sign s' from (5), the local binary pattern (LBP), which characterizes the spatial structure for pixel x_c, is deduced [24], [25],

$$\mathrm{LBP}(x_c) = \sum_{i=1}^{p} s'(g_i - g_c) 2^{i-1}. \tag{6}$$

Finally, the structural information is calculated based on the LBP values. For a pixel x_c, a 2^p bins histogram is acquired by mapping the LBP values of its neighboring local region $\mathcal{X}$ [31] (e.g., a 21×21 surrounding region). And the structural information of x_c is represented by the Shannon entropy of $\mathcal{X}$ [31], which is calculated as follows,

$$H(x_c) = \sum_{b=1}^{2^p} -p_b(x_c) \log p_b(x_c), \tag{7}$$

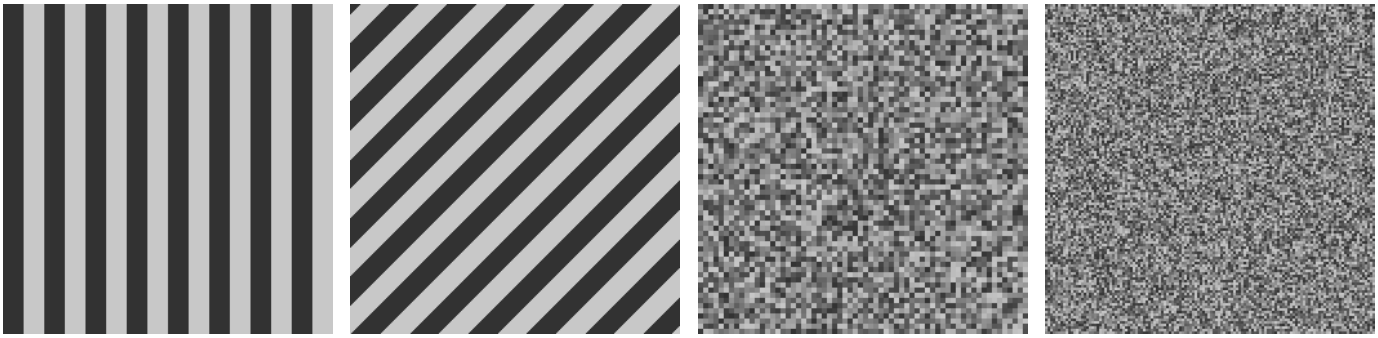
where $p_b(x_c)$ is the probability at bin b of $\mathcal{X}$. With the help of luminance adaptation, the structural information of the uniform background region is approximate to zero, as shown in Fig. 2 (c), which is much more consistent with the HVS than that in Fig. 2 (b).

B. Human Perception and Structural Uncertainty

However, structural information is unequal to structural uncertainty for human perception. The HVS is an efficient and effective visual signal processing system, which helps us to understand the colorful outside world [32]. Rather than literally translates the input scene, the HVS actively predicts the visual content for perception [22]. Fig. 3 shows four concept images with different structures. We can fully understand Fig. 3 (a) and (b), since their structures are orderly and can be easily predicted. However, Fig. 3 (c) and (d) possess much more uncertain information, which represent disorderly structures. And therefore, it is difficult to understand their visual contents. In order to effectively estimate the uncertain information, we should further analyze the characteristic of the HVS on image processing.

Recent research on human perception indicates that the HVS possesses an internal generative mechanism (IGM) for visual signal processing [26], [22]. Furthermore, the IGM theory suggests that the brain will adjust its configuration, e.g., it changes the way of sampling or the way of encoding, to actively predict the visual information for input scene perception and understanding [33]. Therefore, the IGM performs as an active prediction system, and a Bayesian brain theory is introduced to mimic the performance of the IGM [26].

The key of the Bayesian brain theory is a Bayesian probabilistic model that optimizes an input scene by minimizing the prediction error. For example, in image domain, by considering



(a) (b) (c) (d)

Fig. 3: Concept images for structural uncertainty illustration. (a) and (b) Images with orderly structures. (c) and (d) Images with disorderly structures. The structural uncertainty values of (a)-(d) are 0, 0, 1.69, and 3.19, respectively.

(a) (b)

Fig. 4: Structural uncertainty analysis. (a) Original image. (b) Uncertainty mask (H_U), in which we have mapped the values into [0, 255] for a better view, and light regions represent high uncertainty.

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

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

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

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

Fig. 5: Edge filters for four directions.

the relationships among pixels, a pixel x is predicted with its surrounding $\mathcal{X}$ by maximizing the conditional probability $p(x/\mathcal{X})$ for error minimization. With further analysis on the relationships between the central pixel x and surrounding pixels x_i in $\mathcal{X}$, the mutual information $I(x;x_i)$ is adopted as the autoregressive coefficient, and an autoregressive model is created to mimic the IGM for active prediction [19],

$$g' = \sum_{g_i \in \mathcal{X}} \mathcal{C}_i \, g_i + \varepsilon, \tag{8}$$

where g' is the predicted value of pixel x, $\mathcal{C}_i = \frac{I(x;x_i)}{\sum_k I(x;x_k)}$ being the normalized coefficient, and ε is white noise.

With (8), the visual contents of an input scene are actively predicted. And the residual information (i.e., prediction error) between the original image ($\mathcal{M}$) and its corresponding predicted image $\mathcal{M}'$ is regarded as the uncertainty U, namely, $U = \mathcal{M} - \mathcal{M}'$. Then, we analyze the structural information of the uncertainty portion U with (6), and acquire the structural uncertainty H_U of $\mathcal{M}$ according to (7). An example of structural uncertainty is shown in Fig. 4, where the disorderly regions (such as the trees) are with larger structural uncertainty values than that of the orderly regions (such as the sky).

III. Pattern Masking

In this section, we firstly deduce the pattern masking function by considering both luminance contrast and structural uncertainty. And then, in order to determine the parameters in the pattern masking function, a subjective viewing test is designed.

A. The Pattern Masking Function

The HVS is highly sensitive to both luminance change and structural information. Therefore, pattern masking is determined by both luminance contrast and structural uncertainty. For a uniform region with no luminance change, the pattern masking effect is weak and its corresponding visibility threshold is low. When it comes to an edge region with orderly luminance change (such as Fig. 3 (a) and (b)), its visibility threshold will become higher with the increase of the luminance edge height [5]. Furthermore, for an image region with fixed luminance edge height, the more structural uncertainty it possesses, the higher visibility threshold it has [23]. Therefore, we suggest to take both luminance contrast and structural uncertainty into account for pattern masking estimation,

$$\text{PM}(x_c) = f(E(x_c), H_U(x_c)), \tag{9}$$

where $\text{PM}(x_c)$ is the visibility threshold of pixel x_c caused by pattern masking, $H_U(x_c)$ is the structural uncertainty of x_c. $E(x_c)$ is the luminance edge height, which is usually computed as follows [12], [13], [21],

$$E(x_c) = \max_{k=1,\ldots,4} Grad_k(x_c), \tag{10}$$

$$Grad_k = |\varphi \mathcal{M} * \nabla_k|, \tag{11}$$

where ∇_k are four directional filters, as shown in Fig. 5, φ=1/16, and symbol $*$ denotes the convolution operation.

Since research about luminance contrast has been done and the contrast masking effect is investigated throughly, we firstly

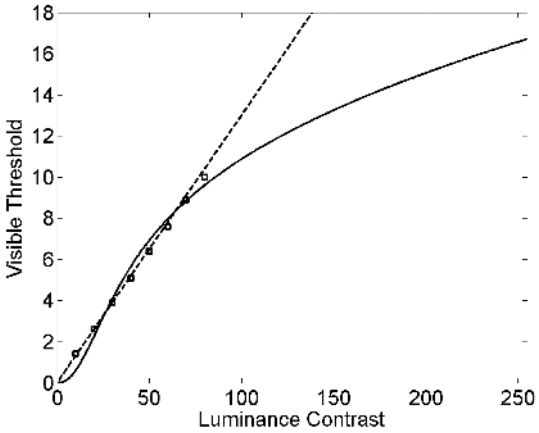


Fig. 6: Fitting the nonlinear transducer of luminance contrast $f_1(x_c)$ with subjective visibility thresholds. The dash line represents the fixed increasing ratio, and the solidline curve represents the fitting results of (13).

analyze the effect from luminance contrast E in (9), namely, the contrast masking effect ($f_1(E)$). And then, by considering the interaction of luminance contrast and structural uncertainty on pattern masking ($f_2(E, H_U)$), (9) can be divided as follows,

$$\text{PM}(x_c) = f_1\left(E(x_c)\right)\, f_2(E(x_c), H_U(x_c)). \tag{12}$$

The existing contrast masking function always calculates the visibility threshold with a fixed increasing ratio to luminance contrast [12], as shown of the dash line in Fig. 6. As a result, the computed threshold for large luminance contrast region is too high. In other words, the visibility threshold for regions with high luminance contrast are overestimated [13]. Perceptual research indicates that the human eye's response to changes in light intensity is nonlinear (e.g., logarithmic [34]), and the increasing ratio should be decreased with the increase of luminance contrast [5]. To this end, a nonlinear transducer for luminance contrast is introduced [5], [14], [8], and the contrast masking is computed as follows,

$$f_1(x_c) = 0.115 \times \frac{\alpha E(x_c)^{2.4}}{E(x_c)^2 + \beta^2}, \tag{13}$$

where α is a constant of proportion and β determines the positively accelerating and compressive regions of the nonlinearity. By fitting (13) with subjective visibility thresholds (which are acquired from a subjective experiment [30]), as shown in Fig. 6, we set $\alpha = 16$ and $\beta = 26$.

Meanwhile, we found that the human eye's response to structural uncertainty is also nonlinear, and there exists a nonlinear transducer ($\mathcal{N}$) for structural uncertainty when measuring the pattern masking effect.

$$\mathcal{N}(H_U(x_c)) = \frac{k_1\, H_U(x_c)^{k_2}}{H_U(x_c)^2 + k_3^2}, \tag{14}$$

where k_1, k_2, and k_3 are fixed parameters which determine the shape of the nonlinear transducer $\mathcal{N}$.

Furthermore, there exists interaction between luminance contrast and structural uncertainty ($\mathcal{I}$). From subjective viewing tests, we have found that under low luminance contrast and high structural uncertainty, the sensitivity of the human eye is low and the visible threshold is high. In this condition, a large value of $f_2(E, H_U)$ is needed to highlight the visual masking effect. While under low structural uncertainty, even though the luminance contrast is high, the sensitivity of the human eye is high and the visible threshold is low. In this condition, a small value of $f_2(E, H_U)$ is needed to restrain the visual masking effect. To this end, we suggest to compute the interaction term $f_2(E, H_U)$ as follows,

$$\mathcal{I}(E(x_c), H_U(x_c)) = (1 + k_4 \exp(-\frac{f_1(x_c)}{k_5}))\,\mathcal{N}(H_U(x_c)), \tag{15}$$

$$f_2(x_c) = 1 + \mathcal{I}(E(x_c), H_U(x_c)), \tag{16}$$

where k_4 is a constant of proportion and k_5 is a decay factor for the interaction between luminance contrast and structural uncertainty. All of the five parameter (i.e., k_i) will be set in the next subsection through a subjective viewing test.

With (12), (13), and (16), the pattern masking effect can be calculated. As can be seen, when the structural uncertainty $H_U(x_c) = 0$ (for orderly region with no uncertainty), $f_2(x_c) = 1$ and $\text{PM}(x_c) = f_1(x_c)$. Therefore, the contrast masking function is a special case of the proposed pattern masking function when structural uncertainty is equal to zero.

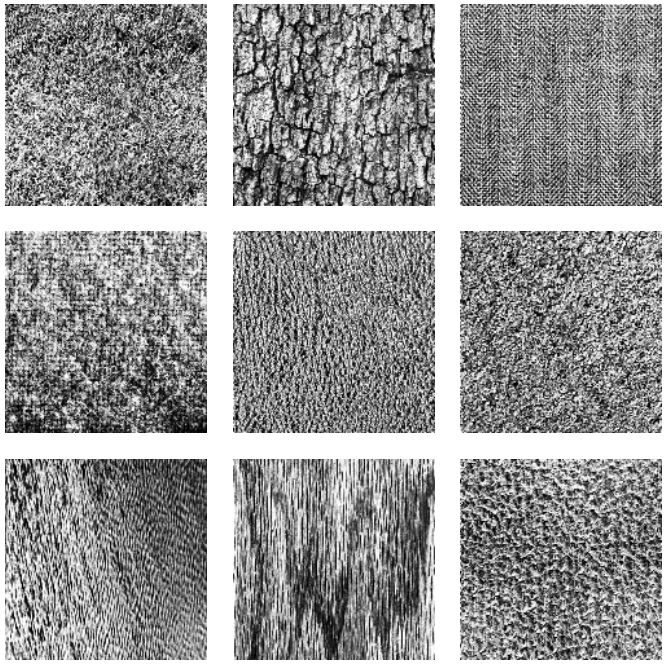


Fig. 7: Images with different structures [35] for parameter determination in the pattern masking function.

B. Determination of Parameters

There are several parameters that need to be determined for the pattern masking function. In this paper, we tune the parameters with a subjective experiment on 9 texture images (with size 512×512) from USC-SIPI database [35]. As shown in Fig. 7, the 9 images possess different kinds of structures, and from the upper left to lower right we name them as $\mathcal{M}_1$ to $\mathcal{M}_9$.

TABLE I: The subjective visibility thresholds (the mean values and the standard deviations) of the 9 texture images.

Image	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
H_U	3.06	2.97	3.16
V	31.5 ± 3.5	28.6 ± 2.0	33.6 ± 4.4
Image	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$
H_U	3.12	3.14	3.18
V	30.0 ± 3.3	31.0 ± 3.1	31.3 ± 2.6
Image	$\mathcal{M}_7$	$\mathcal{M}_8$	$\mathcal{M}_9$
H_U	2.70	2.61	3.02
V	28.0 ± 4.3	25.0 ± 1.2	28.2 ± 4.0

With these images, a subjective viewing test experiment is firstly designed to acquire the subjective visibility thresholds. And then, by fitting the pattern masking equation with these subjective visibility thresholds, the parameters are determined.

In the subjective viewing test experiment, testing images are juxtaposed on a 17-in monitor for visibility threshold measurement. And we set the viewing condition (e.g., the viewing distance, the light condition of the environment, and so on) based on the ITU-R BT.500-11 standard [36]. Twenty viewers (their eye sight is either normal or has been corrected, and ten of them are experts in image processing and the others are naive) are invited in this test, and the viewing distance is four times of the image height. For each image $\mathcal{M}$, the Gaussian white noise is injected with the guidance of the following equation [13], [21],

$$\mathcal{M}_n(x_c) = \mathcal{M}(x_c) + V\,\mathrm{rand}(x_c), \tag{17}$$

where $\mathcal{M}_n$ is the white noised contaminated image, V regulars the energy of the noise, and $\mathrm{rand}(x_c)$ randomly takes $+1$ or -1. Viewers are asked to adjust the value of V (which begins from 1, and Viewers can increase or decrease it) until they can sense the noise, and this value is recorded as the subjective visibility threshold of the current image. The visibility thresholds determined by the subjective tests of the 9 images (as shown in Fig. 7) are listed in Table I, and the structural uncertainty of these images are also listed.

Then, we fit the pattern masking equation (9) with these subjective visibility thresholds based on the lease squares to determine the parameters,

$$\arg \min_{k_1,\cdots,k_5} \sum_{i=1}^{9} [\mathrm{PM}_i - V_i]^2, \tag{18}$$

where PM_i is the average value of computed visibility thresholds on the ith image with (9), and V_i is the visibility threshold determined by the subjective tests of the ith image, as shown in Table I. As a result, the parameters in the pattern masking equation are set as: $k_1 = 2.67$, $k_2 = 3.22$, $k_3 = 1.19$, $k_4 = 2.03$, and $k_5 = 0.19$. As shown in Fig. 8, the visibility threshold (pattern masking) increases with luminance contrast and structural uncertainty.

IV. Experimental Result and Discussion

In this section, we firstly make a comparison between the proposed pattern masking function and the existing contrast masking function [12] to demonstrate the effectiveness of structural uncertainty on visibility threshold estimation. And then, we extend the pattern masking function to JND estimation and introduce a novel pixel domain JND model to further demonstrate the effectiveness of the proposed pattern masking function. Finally, a subjective viewing test is designed to make a comprehensive comparison between the novel JND model and three latest pixel domain JND models (i.e., Yang et al.'s model [13], Liu et al.'s model [37], and Wu et al.s model [21]) on a set of images from two public databases [35], [38]. For color images, the proposed model is performed on all color channels.

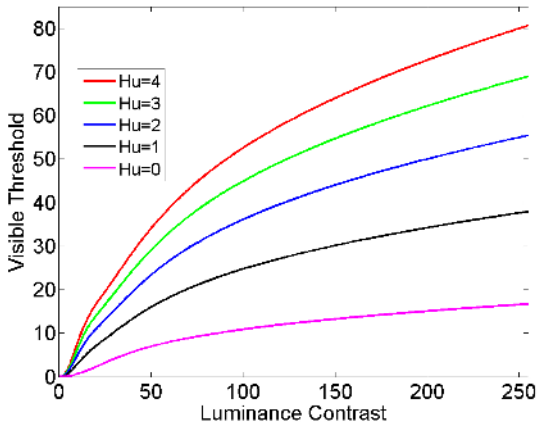


Fig. 8: Pattern masking by considering luminance contrast and structural uncertainty (the output of equation (12)).

A. Pattern Masking VS. Contrast Masking

An effective visual masking function should be able to accurately indicate the sensitivity of the HVS to different image regions. In order to demonstrate the effectiveness of a visual masking function, the sensitive testing experiment is always adopted [13], [37]. For a test image, the white noise is injected with the guidance of the visual masking model (F), which is shaped as follows [12], [21],

$$\hat{\mathcal{M}}(x_c) = \mathcal{M}(x_c) + \mathcal{E}\,\mathrm{rand}(x_c)\,F(x_c), \tag{19}$$

where $\hat{\mathcal{M}}$ is the white noise contaminated image, $\mathcal{E}$ regulates the energy of the white noise, which makes the same noise energy for different visual masking models (F), and $\mathrm{rand}(x_c)$ randomly takes $+1$ or -1.

By taking F as contrast masking [12], [13] (or pattern masking) in (19), we inject white noise into the Cemetry image with the guidance of the contrast masking function (or the pattern masking function), as shown in Fig. 9. With the parameter $\mathcal{E}$, the energies of the two noise-contaminated images (i.e., Fig. 9 (b) and (c)) are adjusted to be the same (with MSE $= 100$). Therefore, we can make a fair comparison between existing contrast masking function and the proposed pattern masking function.

The contrast masking function [12], [13] is mainly based on luminance contrast for visibility threshold computation. As a result, an edge region leads to high visibility threshold under the contrast masking function. As shown in Fig. 9 (d), the

(a)

(b)

(c)

(d)

(e)

Fig. 9: Pattern masking VS. contrast masking on Cemetry image (only a part of the image is cut due to the resolution limitation of the scene). (a) The original image. (b) and (c) Contaminated images with contrast masking and pattern masking guide noise (under the guidance of (19) with a same noise level MSE $= 100$), respectively. (d) and (e) The noise mask of (b) and (c) (i.e., the value of $\mathcal{E}\cdot F(x_c)$ in (19)), respectively. And light regions of (d) and (e) represent high masking effect.

white words on the black board, the steel bars, and the edge of the brick wall are highlighted. However, the HVS is sensitive to the edge region [13], and therefore, the contrast masking function overestimates the visibility threshold of the edge region. As shown in Fig. 9 (b), the noise in these edge regions is easily perceived and generates obvious quality degradation. Meanwhile, the disorderly regions, such as the trees, the grass, and the surface of the brick in Fig. 9 (b), which do not have high luminance contrast but are insensitive to the HVS, are underestimated with the contrast masking function.

The proposed pattern masking function, which takes both luminance contrast and structural uncertainty into account, returns more accurate visibility thresholds for images than the contrast masking function. Though these edge regions (i.e., the white words on the black board, the steel bars, and the edge of the brick wall) possess large luminance edge heights, they represent much orderly structures (with little structural uncertainty). Therefore, the HVS is sensitive to them and their visibility thresholds are not so high, as shown in Fig. 9 (e). Furthermore, the disorderly regions (i.e., the trees, the grass, and the surface of the brick) possess much structural uncertainty, the HVS cannot fully understand their detail and is insensitive to them. As a result, these disorderly regions have high visibility thresholds. As shown in Fig. 9 (e), the output of the proposed pattern masking function highlights these disorderly regions and suppresses orderly edge regions. With the guidance of the proposed pattern masking function, much more noise can be injected into the insensitive regions and less into the sensitive regions. Therefore, the noise in Fig. 9 (c) generates less perceptual quality degradation than that in Fig. 9 (b), though the two images have the same level of noise energy. In summary, the proposed pattern masking function is more consistent with the HVS than that of the contrast masking function.

B. Pattern Masking based JND Model

In order to further demonstrate the effectiveness of the proposed pattern masking function, we adopt pattern masking to improve the JND estimation. In general, luminance adaptation and contrast masking are taken into account for pixel domain JND estimation [12], [13], [37]. In this paper, by replacing contrast masking with pattern masking, we introduce a novel JND estimation model,

$$\mathrm{JND}(x_c) = \mathrm{LA}(x_c) + \mathrm{PM}(x_c) - C^{gr} \times \min\{\mathrm{LA}(x_c), \mathrm{PM}(x_c)\}, \tag{20}$$

where C^{gr} is the gain reduction parameter due to the overlapping between luminance adaptation LA and pattern masking PM, and is set as $C^{gr} = 0.3$ (the same as in [13]).

And then, we compare the proposed JND model with three latest JND models, namely, Yang et al.'s model [13], Liu et al.'s model [37], and Wu et al.'s model [21] (we have not compare with the classic Chou and Li's model [12], because [13], [37] are two improved models of [12]). By setting F in (19) as a JND model, white noise is injected into an image with the guidance of the corresponding JND model. Fig. 10 shows the visibility threshold maps of the four different JND models on Cemetry image, and Fig. 11 shows their corresponding noise-contaminated images.

Considering that the HVS is sensitive to the edge region and the contrast masking function always overestimates the visibility threshold of the edge regions, Yang et al.'s suggested to protect the edge region for JND estimation [13]. And therefore, the canny edge detection is adopted to protect the primary edge regions. As shown in Fig. 10 (a), the visibility thresholds of the primary edge regions (such as the white words on the black board and the steel bars) are suppressed. However, the visibility thresholds of the other edge regions are pop-out (such as the trunk of the trees, the edge of the brick wall, and the other words regions). In summary, with the protection of the canny edge detection in [13], the secondly edge regions (which always with not so high edge heights) are highlighted and these disorderly regions are still underestimated. As a result, we can still easily perceive the noise in Fig. 11 (a) (such as some word regions and the edge of the brick wall).

Since the contrast masking function overestimates edge regions and underestimates texture regions, Liu et al.'s [37] suggested to separately estimate the contrast masking of the two kinds of regions (i.e., edge regions and texture regions). Therefore, a texture classification algorithm is firstly employed to separate the two kinds of regions. And then, a bigger weight is multiplied to the contrast masking function of the texture regions and a smaller one for the edge regions. The JND threshold map of Liu et al.'s model is shown in Fig. 10 (b). As can be seen, the steel bars regions are correctly separated into the edge regions and are effectively protected. However, some sensitive regions, such as the white words on the black board, are separated into the texture regions and are highlighted. Meanwhile, some disorderly regions, such as the grass and the surface of the brick, are still underestimated. In summary, Liu et al.'s model tries to improve the JND estimation by protecting edge regions and highlighting texture regions. But the texture regions cannot always hide much noise (such as the white words on the black board in Fig. 10 (b)), and actually, only these texture regions with uncertain structures can hide much noise. As a result, too much noise is injected into the words regions, where the HVS is highly sensitive, as shown in Fig. 11 (b).

In [21], the spatial structural character is considered for JND estimation. By computing the structural regularity based on the self-similarity of image structure, an ad hoc spatial masking function is introduced for JND estimation. This JND model effectively protects the orderly regions (e.g., the edge regions), while overestimates the disorderly texture regions. As shown in Fig. 10 (c), the visibility thresholds are quite low in the orderly regions (such as the steel bars and the words regions), and are very high in the disorderly regions (such as the surface of the brick and the grass). However, the edge region, which can hide more noise than the smooth region, is overprotected. In addition, too much noise is injected into the disorderly regions. As shown in Fig. 11 (c), we can obviously sense the spots in these disorderly regions, which degrade the perceptual quality of the image.

In the proposed JND model, the sensitive regions acquire

(a) (b) (c) (d)

Fig. 10: The visibility threshold maps (i.e., the value of $\mathcal{E}\cdot F(x_c)$ in (19)) of the four different JND models on Cemetry image [38]. Form (a) to (d), they are the outputs of Yang et al.'s [13] model, Liu et al.'s model, Wu et al.'s model, and the proposed JND model, respectively. And light regions represent high masking effect.

low visibility thresholds and the insensitive regions has high visibility thresholds. As shown in Fig. 10 (d), by considering the effect of luminance adaptation (caused by the background luminance) and the effect of pattern masking (caused by luminance contrast and structural uncertainty), the noise is appropriately deployed into image regions with different contents: very little noise is injected into the uniform regions, much into the orderly edge regions, and most into these regions with both luminance change and structural uncertainty. Though the level of noise energy in the four contaminated images (i.e., Fig. 11 (a)-(d)) is the same, Fig. 11 (d) represents a better perceptual quality than the other three images. Therefore, the proposed JND model outperforms the latest three pixel domain JND models (i.e., [13], [37], [21]).

C. Subjective Viewing Test

For a more comprehensive comparison between the proposed pattern masking function (the proposed JND model) and the existing contrast masking function (the three latest JND models), a subjective viewing test experiment is designed. The setting of the viewing condition is the same as mentioned in Subsection III-B, which follows the ITU-R BT.500-11 standard [36]. By considering the availability and efficiency, eight images from [35] (which are oft-used in JND comparison experiments) and eight representative images from [38] (which are oft-used in quality assessment experiments, and four of them are mainly composed with orderly texture and the other four are composed with disorderly texture) are chosen, as shown in Fig. 12. In each test, two noise-contaminated images about a same scene are juxtaposed on the screen (two noise-injected images with the guidance of our model and other comparison model, respectively. And they are randomly juxtaposed on the left or right). Then 38 subjects (their eye sight is either normal or has been corrected, and fifteen of them are experts in image processing and the other twenty three are naive) were invited to evaluate which one is better and how

(a) (b) (c) (d)

Fig. 11: Comparison among four different JND models on Cemetry image [38]. Form (a) to (d), they are noise-contaminated images with the guidance of Yang et al.'s [13] model, Liu et al.'s model, Wu et al.'s model, and the proposed JND model, respectively. Under the guidance of (19), the level of the noise energy of the four images are the same (MSE $= 150$).

TABLE II: Scores for Subjective Viewing Test

Score	0	1	2	3
Description	Same quality	Slightly better	Better	Much better

much better it is (following the evaluation rule as shown in Tab II; and if the left one is better than the right one, then give a positive score; otherwise, give a negative score).

Table III shows the subjective scores of comparison results between the proposed pattern masking function and the existing contrast masking function, in which the positive (negative) score means the proposed pattern masking function performs better (worse) than the existing contrast masking function. Since the proposed pattern masking function takes structural uncertainty into account, it performs better (has positive score) on most of images than the existing contrast masking function, especially on these images with much disorderly regions (e.g., the Tank image with a large region of grass and the Ocean image with a big unsmooth sea surface). Meanwhile, there are

TABLE III: Subjective viewing test result (pattern masking vs. contrast masking).

Image	Mean	Std	Image	Mean	Std
Indian	1.389	1.554	Ocean	2.194	0.710
Lena	0.056	1.241	Caps	0.028	1.362
Barbara	1.750	1.204	Plane	1.861	0.867
Peppers	-0.028	0.609	Paint	1.222	0.959
Tank	2.000	0.926	Rapids	0.083	1.402
Airplane	0.722	1.162	House	0.250	0.874
Huts	0.861	1.073	Beacon	1.583	1.156
Boats	0.778	1.245	Stream	0.278	1.059
Average	0.939	–	–	–	–

several scores which are approximate to zeros, which means that the proposed pattern masking function performs equally to the existing contrast masking function on their corresponding images (i.e., Lena, Pepper, and Caps). With further analysis, we found a common feature among these images, namely, most regions in the three images are uniform/orderly and there

Fig. 12: Images for subjective viewing test. The first and second rows of images are from [35] which are oft-used in JND comparison experiments, and the other two rows of images are from [38] which are oft-used for quality assessment

are little uncertain structures. Under this special condition, the structural uncertainty is approximated to zero, and the pattern masking function is similar with the contrast masking function (as analyzed in Subsection III-A). Moreover, the average score (0.939) in Table III further confirms that the proposed pattern masking function performs better than the existing contrast masking function.

Table IV shows the comparison results between the proposed JND model and three latest JND models (i.e., [13], [37] and [21]). By comparing with Yang et al.'s [13] and Liu et al.'s [37] models (both of them are based on contrast masking), the proposed JND model performs better on almost all of these images, and performs equally on one image (i.e., the Caps, which mainly represents orderly structures; and the reason is much similar to the description in the above paragraph). Wu et al.'s model [19] also considers the structural character for spatial masking estimation. However, this ad hoc JND model cannot accurately measure the effect caused by the structural uncertainty, and overestimates the visibility threshold of the region with uncertainty. When compared with Wu et al.'s model, the proposed JND model performs better on most of these images except the Plane image. That is because the plane image is mainly composed with a large disorderly grass region and a very orderly object region, and both of the two JND models inject almost all of the noise into the grass region

TABLE IV: Subjective viewing test results (the proposed JND model vs. three latest JND models, respectively).

Image	Our vs. Yang		Our vs. Liu		Our vs. Wu	
	Mean	Std	Mean	Std	Mean	Std
Ocean	0.806	0.786	1.139	1.046	0.389	0.836
Caps	0.028	1.383	-0.056	1.264	0.486	1.513
Plane	0.778	1.017	1.306	1.037	0.042	1.040
Paint	0.694	0.822	0.250	1.131	0.514	0.950
Rapids	0.778	1.290	0.583	1.402	0.528	1.137
House	0.694	1.009	0.083	0.967	0.486	1.086
Beacon	0.278	0.944	0.028	0.910	0.875	0.971
Stream	0.444	1.054	0.083	0.841	0.764	0.997
Average	0.563	–	0.427	–	0.510	–

which is highly insensitive to the HVS. In addition, the average scores (Our vs. Yang is 0.563, Our vs. Liu is 0.427, and Our vs. Wu is 0.510) on all of the images are all positive, which indicate that the proposed JND model outperforms the three latest JND models.

V. Conclusion

In this paper, we have introduced a pattern masking function by considering both luminance contrast and structural uncertainty. Visual masking, which effectively represents the visual redundancy, is useful in image/video compression, scene enhancement, quality assessment, watermarking, etc. The existing contrast masking function only takes luminance contrast into account for visual masking estimation, which overestimates the edge region and underestimates the texture region. According to the recent research on visual perception, we suggested that the edge region is much orderly and the HVS can easily predict its structural character, while the texture region possesses uncertainty which impedes the prediction of structural information by the HVS. In other words, structural uncertainty effects the sensitivity of the HVS for visual perception. Therefore, we advocated that structural uncertainty is another determining factor on visual masking.

By mimicking the internal generative mechanism (IGM) on scene understanding, we employed an autoregressive based prediction model to separate the uncertain information from an input scene. And then an improved local binary pattern (LBP) scheme has been introduced for structural uncertainty estimation. Finally, a pattern masking function has been deduced based on luminance contrast and structural uncertainty. Experimental results demonstrated that the proposed pattern masking function outperforms the contrast masking function. Furthermore, we extended the proposed pattern masking function and introduced a novel pixel domain JND estimation model. Subjective viewing test further confirmed that the proposed JND model is more consistent with the HVS than the existing JND models.

References

[1] S. L. Macknik and M. S. Livingstone, "Neuronal correlates of visibility and invisibility in the primate visual system," *Nature Neuroscience*, vol. 1, no. 2, pp. 144–149, Jun. 1998.

[2] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 806–819, 2010.

[3] T. Ritschel, K. Smith, M. Ihrke, T. Grosch, K. Myszkowski, and H.-P. Seidel, "3D unsharp masking for scene coherent enhancement," *ACM Transactions on Graphics*, vol. 27, no. 3, p. 90:190:8, Aug. 2008.

[4] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, p. 011006, 2010.

[5] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.

[6] G. Kovcs, R. Vogels, and G. A. Orban, "Cortical correlate of pattern backward masking," *Proceedings of the National Academy of Sciences*, vol. 92, no. 12, pp. 5587–5591, Jun. 1995.

[7] "Masking by patterns," http://colorusage.arc.nasa.gov/masking.php, [Online; accessed 2013].

[8] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2379–2391, 1997.

[9] J. M. Foley and G. M. Boynton, "New model of human luminance pattern vision mechanisms: analysis of the effects of pattern orientation, spatial phase, and temporal frequency," *Proceedings of SPIE*, vol. 2054, no. 1, pp. 32–42, Mar. 1994.

[10] E. Peli and J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models For Target Detection And Recognition*. WORLD SCIENTIFIC, May 1995, pp. 245–283.

[11] A. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proceeding of the IEEE*, vol. 65, no. 4, pp. 536–548, 1977.

[12] 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.

[13] X. K. Yang, W. S. Ling, 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.

[14] J. M. Foley, "Human luminance pattern-vision mechanisms: masking experiments require a new model," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[15] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *IEEE Transactions on Image Processing*, vol. 12, no. 7, p. 729739, Jul. 2003.

[16] N. Petkov and M. A. Westenberg, "Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition," *Biological cybernetics*, vol. 88, no. 3, pp. 236–246, Mar. 2003.

[17] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *Journal Visual Communication and Image Representation*, vol. 19, no. 1, pp. 30–41, Jan. 2008.

[18] Z. Wei and K. Ngan, "Spatio-Temporal just noticeable distortion profile for grey scale Image/Video in DCT domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337–346, 2009.

[19] J. Wu, W. Lin, G. Shi, and A. Liu, "A perceptual quality metric with internal generative mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

[20] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[21] J. Wu, F. Qi, and G. Shi, "Self-similarity based structural regularity for just noticeable difference estimation," *Journal of Visual Communication and Image Representation*, vol. 23, no. 6, pp. 845–852, 2012.

[22] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, Feb. 2010.

[23] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *Proceedings of SPIE*, vol. 1666, no. 1, pp. 2–15, Aug. 1992.

[24] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikinen, "Texture discrimination with multidimensional distributions of signed gray-level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, 2001.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[26] D. C. Knill and R. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *Trends In Neuroscience*, vol. 27, pp. 712–719, 2004.

[27] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 60–65.

[28] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and Super-Resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[29] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[30] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*. New York: Plenum, 1988.

[31] J. Wu, F. Qi, G. Shi, and Y. Lu, "Non-local spatial redundancy reduction for bottom-up saliency estimation," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 1158–1166, 2012.

[32] P. Jacob and M. Jeannerod, *Ways of Seeing: The Scope and Limits of Visual Cognition*. Oxford University Press, USA, 2003.

[33] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology, Paris*, vol. 100, no. 1-3, pp. 70–87, Sep. 2006.

[34] M. Jourlin, M.Carr, J. Breugnot, and M.Bouabdellah, "Logarithmic image processing: Additive contrast, multiplicative contrast, and associated metrics," *Advances in Imaging and Electron Physics*, vol. 171, pp. 357–406, 2012.

[35] "The USC-SIPI image database," http://sipi.usc.edu/database/database.php, [Online; accessed 2011].

[36] "Method for the subjective assessment of the quality of television pictures," ITU, Document ITU-R BT.500-11, 2002, geneva, Switzerland.

[37] 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.

[38] H. R. Sheikh, Z. Wang, L. K. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," http://live.ece.utexas.edu/research/quality.