Visual Masking Estimation Based On Structural Uncertainty

Jinjian Wu
School of Electronic Engineering
Xidian University
Xi'an, Shannxi, China
jinjian.wu@mail.xidian.edu.cn

Weisi Lin
School of Computer Engineering
Nanyang Technological University
Nanyang 639798, Singapore
wslin@ntu.edu.sg

Guangming Shi
School of Electronic Engineering
Xidian University
Xi'an, Shannxi, China
gmshi@xidian.edu.cn

Abstract—A model of visual masking, which reveals the visible threshold of human perception, is useful in perceptual based image/video processing. The existing visual masking formulation, which mainly considers luminance contrast, cannot accurately estimate the visible threshold. Recent researches indicate that human perception is highly adaptive to extract orderly structures and is insensitive to disorderly structures. Therefore, we suggest that the structural characteristic is another determining factor for visual masking, and deduce a novel visual masking function based on structural uncertainty. Experimental results demonstrate that the proposed model is more consistent with human perception than the existing visual masking model.

Index Terms—Visual Masking, 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. 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 processing, such as compression, watermarking, quality assessment, and so on.

As the HVS is highly sensitive to luminance change of the input stimuli, researchers always estimate visual masking based on luminance contrast [1] for simplicity. In [2], a psychophysical experiment is designed to investigate the relationship between luminance edge height and the visibility threshold. Moreover, according to the recording data from [2], a well accepted contrast masking function is deduced in [3]. However, the contrast masking function only takes the luminance contrast into account, which always overestimates the masking effect of the edge region with orderly structures and underestimates the texture region with complex structures [4].

Recent research on visual perception indicates that the HVS is highly adaptive to extract orderly structures and tries to avoid disorderly structures for image perception and understanding [5]. In other words, the HVS is less sensitive to uncertain regions (e.g., some texture regions) which possess disorderly structures [6], and the visual masking effects in

these uncertain regions are strong [7]. Therefore, visual masking is related to not only luminance contrast but also structural uncertainty, and we suggest to take both factors into account for visible threshold estimation.

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 [8]. By considering these features, a famous local binary pattern (LBP) algorithm is introduced in [8] 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 [9]. Meanwhile, the Bayesian brain theory [10] further indicates that the HVS possesses an internal generative mechanism, 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 advocate to consider the unpredictable information as the uncertainty. And the structural uncertainty is computed based on the LBP values of the unpredictable information.

Finally, combining structural uncertainty with luminance contrast, we deduce a novel computational function for visual masking. Furthermore, to demonstrate the effectiveness of the proposed pattern masking, we extend the proposed visual masking function to 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 [3]) estimation. Experimental results from subjective viewing tests confirm that the proposed visual masking function correlate better with the HVS than the existing contrast masking function.

The organization of this paper is as follows: in Section II, structural uncertainty is estimated to deduce a novel visual masking function. Experimental results of the proposed visual masking function are presented in Section III. Finally, conclusions are drawn in Section IV.

II. STRUCTURAL UNCERTAINTY BASED VISUAL MASKING

In this section, we firstly analyze structural uncertainty with the LBP algorithm. And then, by considering both luminance
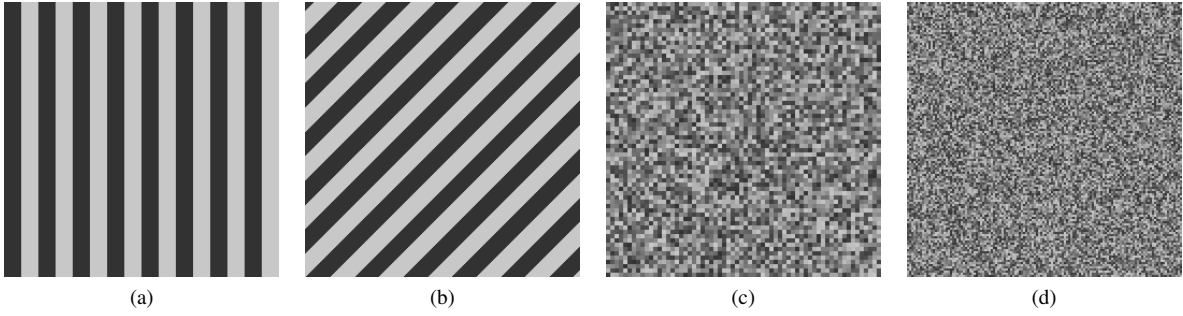
Fig. 1: Structural uncertainty illustration. (a) and (b) Orderly structures. (c) and (d) Disorderly structures.

contrast and structural uncertainty, a novel visual masking function is deduced.

A. Structural Uncertainty Computation

The HVS is an efficient and effective visual signal processing system, which helps us to understand the outside colorful world. Rather than literally translates the input scene, the HVS possesses an internal generative mechanism, which actively predicts the visual content for perception and understanding [5]. Fig. 1 shows four concept images with different structures. We can fully understand Fig. 1 (a) and (b), since their structures are orderly and can be easily predicted. However, Fig. 1 (c) and (d) are with much more uncertain information, which represent disorderly structures. And therefore, it is difficult to predict their visual contents.

Inspired by the internal generative mechanism in the HVS, a Bayesian brain theory is introduced to mimic the active prediction [10]. 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 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 active prediction of the HVS [9],

$$x' = \sum_{x_i \in \mathcal{X}} \mathcal{C}_i x_i + \varepsilon, \qquad (1)$$

where x' is the predicted value of pixel x, $\mathcal{C}_i = I(x; x_i)/\sum_k I(x; x_k)$ is the normalized coefficient, and ε is white noise. With (1), the visual contents of an input scene is 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 uncertainty U, namely, $U = \mathcal{M} - \mathcal{M}'$.

And then, the LBP algorithm is employed to compute structural uncertainty on the uncertainty portion U. By considering the spatial distribution of structural information, Ojala et al. [8] analyzed the spatial relationship among pixels, and introduced a LBP algorithm. In the LBP algorithm, the pattern number of

a pixel x_c is computed by comparing its value g_c with those g_i in its surroundings [8],

$$\text{LBP}(x_c) = \sum_{i=1}^{p} s(g_i - g_c) 2^p, \qquad (2)$$

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

According to the pattern number of the uncertainty portion U from (2), the probability distribution of U is calculated within 2^p bins (where p is usually set as 8 [8]). And then, with the probability distribution of U, the quantity of structural uncertainty (H_U) is computed with the Shannon Entropy equation.

B. Visual Masking Function

The HVS is highly sensitive to both luminance difference and structural information, which represent the visual information of an input scene. Therefore, visual masking is determined by both luminance contrast and structural uncertainty. For a uniform region with no luminance change, the visual masking effect is weak and its corresponding visible threshold is low. When it comes to an edge region with orderly luminance change (e.g., Fig. 1 (a) and (b)), its visible threshold will become higher with the increasing of the luminance edge height [1]. Furthermore, for a image region with fixed luminance edge height, the more structural uncertainty it possesses, the higher visible threshold it has [7]. Therefore, we estimate visual masking by taking both luminance contrast and structural uncertainty into account,

$$\text{V}_t(x) = f_e(x) f_s(x), \qquad (4)$$

where $\text{V}_t(x)$ is the visible threshold of pixel x, $f_e(x)$ is the effect from luminance edge height, and $f_s(x)$ is the effect from structural uncertainty.

The existing contrast masking function always calculates the visible threshold with a fixed increasing ratio to luminance contrast [3], which always overestimates the visible threshold for regions with high luminance contrast [4]. That is because the increasing ratio will be decreased with the increase of luminance contrast [1]. To this end, a nonlinear transducer for luminance contrast is introduced [1], and the contrast masking

is computed as follows,

$$f_e(x) = 0.115 \times \frac{\alpha E(x)^{2.4}}{E(x)^2 + \beta^2}, \tag{5}$$

where α is a constant of proportion, β determines the positively accelerating and compressive regions of the nonlinearity, and $E(x)$ is the luminance edge height of pixel x [3]. By fitting (5) with subjective visible thresholds (which are acquired from a subjective experiment [11]), we set $\alpha = 16$ and $\beta = 26$.

Meanwhile, we suggest that there also exists a nonlinear transducer for structural uncertainty when measuring the pattern masking effect, and $f_s(x)$ is computed as follows,

$$f_s(x) = 1 + \frac{k_1 H_U(x)^{k_2}}{H_U(x)^2 + k_3^2}, \tag{6}$$

where k_1, k_2 and k_3 are fixed parameters which determine the shape of the nonlinear transducer, and $H_U(x)$ is the structural uncertainty of pixel x. In order to determine the three fixed parameters, a subjective viewing test is designed. By fitting (6) with the acquired subjective viewing data, we set $k_1 = 2.67$, $k_2 = 3.22$ and $k_3 = 1.19$.

With (4), (5), and (6), the visual masking effect can be calculated. As can be seen, when the structural uncertainty is zero (for orderly region with no uncertainty), $f_s(x) = 1$ and $V_t(x) = f_e(x)$. Therefore, the contrast masking function is a special case of the proposed visual masking function.

III. EXPERIMENTAL RESULT AND DISCUSSION

In this section, we firstly make a comparison between the proposed visual masking function and the existing contrast masking function to demonstrate the effect of structural uncertainty on visual threshold. And then, by replacing contrast masking with the proposed visual masking function (i.e., (4)) for JND estimation, a novel pixel domain JND model is introduced to further demonstrate the effectiveness of the proposed visual 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 [4], Liu et al.'s model [12], and Wu et al.'s model [6]) on a set of images.

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 with the noise shaping function [3], [6] as shown in follows,

$$\hat{\mathcal{M}}(x) = \mathcal{M}(x) + \mathcal{E} \, \mathrm{rand}(x) \, F(x), \tag{7}$$

where $\hat{\mathcal{M}}$ is the white noise contaminated image, $\mathcal{M}$ is the original image, $\mathcal{E}$ regulates the energy of the white noise, which makes the same noise energy for different visual masking functions F, and $\mathrm{rand}(x)$ randomly takes $+1$ or -1. Fig. 2 shows two white noise contaminated images with contrast masking and the proposed visual masking models, respectively. Thanks to the parameter $\mathcal{E}$, the energies of the two contaminated images are the same (with MSE = 100).

TABLE I: Scores for Subjective Viewing Test

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

The contrast masking function is mainly based on luminance contrast for visible threshold computation. As a result, the edge regions, which always possesses high luminance edge height, acquires high visible threshold under the contrast masking function. As shown in Fig. 2 (a), the white words on the black board, the steel bars, and the edge of the brick wall are severely distorted with too much noise. Meanwhile, these disorderly regions, such as the trees, the grass, and the surface of the brick in Fig. 2 (a), 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 visible thresholds for images than the contrast masking function. Though these edge regions possess large luminance edge heights, they represent much order structures (with little structural uncertainty). Therefore, the visible thresholds of these edge regions are not so high and less noise is injected into these regions with the proposed visual masking function, as shown in Fig. 2 (b). Meanwhile, these disorderly regions (i.e., the trees, the grass, and the surface of the brick) have high visible thresholds, and much noise is injected into these disorderly regions.

Though the two images (i.e., Fig. 2 (a) and (b)) have the same level of noise energy, the noise in Fig. 2 (b) generates less perceptual quality degradation than that in Fig. 2 (a). Therefore, the proposed visual masking function is more consistent with the HVS than the contrast masking function.

In order to further demonstrate the effectiveness of the proposed visual masking function, we replace contrast masking in the JND model (e.g., [4]) with the proposed visual masking, and introduced a novel JND model. And then, we compare the proposed JND model with three latest JND models, namely, Yang et al.'s model [4], Liu et al.'s model [12], and Wu et al.'s model [6], through a subjective viewing test experiment. Following the ITU-R BT.500-11 standard [13], two contaminated images are juxtaposed (randomly on the left or right) on a 17-in screen (two noise-injected images with the guidance of the proposed JND model and other comparison JND model, respectively), and 38 subjects are invited to evaluate which one is better and how much better it is (follows the evaluation rule as shown in Table I).

Table II shows the comparison results between the proposed JND model and three latest JND models (i.e., [4], [12] and [6]). By comparing with Yang et al.'s [4] and Liu et al.'s [12] 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 several images (i.e., the Lena, Pepper, and Caps images, which mainly represent orderly structures; that is because the proposed visual masking is much similar with contrast masking under little uncertainty,

(a)

(b)

Fig. 2: Visual masking VS. Contrast masking on Cemetry image (with size 482×627). (a) and (b) are contaminated images with contrast masking and pattern masking guide noise (under a same noise level MSE = 100), respectively.

TABLE II: 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
Indian	1.750	1.228	0.611	1.248	1.111	0.836
Lena	0.639	0.931	0.028	0.878	0.750	0.819
Barbara	1.778	0.760	0.694	0.980	1.042	1.261
Peppers	-0.056	0.860	0.250	0.604	0.292	0.844
Tank	0.306	0.980	1.556	1.081	0	1.009
Airplane	0.306	1.117	1.389	1.225	0.139	1.023
Huts	0.778	0.637	0.472	0.845	0.417	0.974
Boats	0.611	0.803	0.528	1.108	0.222	1.316
Stream	0.444	1.054	0.083	0.841	0.764	0.997
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
Ocean	0.806	0.786	1.139	1.046	0.389	0.836
Average	0.670	–	0.559	–	0.503	–

as mentioned in Section II-B). When comparing with Wu et al.'s model, the proposed JND model performs better on most of these images except Tank and Plane images (Wu et al.'s model [9] also considers the structural character for spatial masking estimation). In addition, the average scores (Our vs. Yang is 0.670, Our vs. Liu is 0.559, and Our vs. Wu is 0.503) on all of the images are all positive, which indicate that the proposed JND model outperforms the three latest JND models.

IV. CONCLUSION

In this paper, a novel visual masking function is introduced. The existing contrast masking function always overestimates the edge region and underestimates the texture regions. According to the recent researches on visual perception, we have advocated that the edge region is much orderly and the HVS can easily predict its structural character, while the texture region always possesses uncertainty which impedes the prediction of structural information. And therefore, we have

suggested that structural uncertainty is another determining factor on visual masking. By considering both luminance contrast and structural uncertainty, we have deduced a novel visual masking function for visible threshold estimation. Experimental results have demonstrated that the proposed visual masking function is highly consistent with the HVS and outperforms the existing contrast masking function.

REFERENCES

[1] 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.

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

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] J. Wu, W. Lin, G. Shi, and A. Liu, "A perceptual quality metric with internal generative mechanism," IEEE Transactions on Image Processing, 2012.

[10] 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.

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

[12] 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.

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