A Novel Edge-pattern-based Just Noticeable Difference Model for Screen Content Images

Xueqin Liu, Xuanni Zhan
Shenzhen University (SZU)
Shenzhen, Guangdong, P.R. China
xueqin.l@foxmail.com

Miaohui Wang*
Shenzhen University (SZU)
Shenzhen, Guangdong, P.R. China
mhwang@szu.edu.cn

***Abstract*—Just noticeable difference (JND) reflecting visual redundancy of human visual system (HVS) has been widely adopted in image/video processing. In this work, we introduce a novel JND model in pixel domain for screen content (SC) images. Due to the obvious different impacts on HVS between pictorial region and computer-generated texture, we decompose a SC image into screen content set (SCS) and non-screen content set (non-SCS) for JND modeling. For SCS, edge masking effect is mainly considered by using a parametric model with an adaptive edge representation. For non-SCS, both the edge masking effect and pattern masking effect are taken into account to improve distortion tolerance by injecting more noise to pattern complexity region without perceptual quality degradation. Furthermore, visual saliency is employed to adjust pattern masking effect in the view of HVS. Experimental results show that compared with state-of-the-arts, the proposed JND model can tolerate more distortion, and provide better perceptual quality at the same noise level.**

***Index Terms*—just noticeable difference; edge masking effect; pattern masking effect; visual attention effect.**

I. Introduction

Based on the investigation of human visual system (HVS), there is an accepted view that the perceptual ability of human eyes to the changes of pixel values is limited and nonlinear, which means HVS hardly detects the distortion when pixel values change below a certain threshold. So the concept of just perceptible distortion (JND) is proposed to describe the visibility threshold. JND estimation has been given extensive attention by researchers at home and abroad due to the characteristic of reflecting visual redundancy. The performance of many image/video processing technologies based on visual perception can be improved by effectively incorporating a JND model.

Numerous JND models were proposed in both pixel domain and transform domain. Compared with transform domain, the JND threshold could be given intuitively in pixel domain. Chou *et al*. [1] proposed a typical JND model by considering the luminance and contrast masking effects. Yang *et al*. [2] used the canny operator separating edge pixels from texture pixels to avoid overestimating the threshold of edge pixels and introduced the nonlinear additivity model for masking (NAMM). Wang *et al*. [3] first developed a JND model based on a parametric edge model for screen content (SC) images. Wu *et al*. [4] proposed the pattern masking effect to adjust the JND map by considering the diversity of pixel orientations in a local region. Hadizadeh *et al*. [5] and Zeng *et al*. [6] added visual attention effect into the JND map for better performance.

Most of the proposed JND models are used for processing non-screen content (non-SC) images, and there are a few for SC images. With the explosive growth of Internet technologies, SC images have played an important role in modern multimedia applications, such as *distance education*, *screen sharing*, and *real-time communication*. SC images contain more types of image elements, which indicates SC images not only consist of pictorial region but also add computer-generated graphics, icons and text. Shape edges, thin lines, and extensive smooth areas are the obvious features of SC images. These characteristics imply that an accurate model to analyze edge properties is necessary. In [3], a parametric edge JND model suitable for SC images is developed. The overall idea of this model is to describe the edge as three independent parts including edge luminance, edge contrast and edge structure for final visibility threshold. However, the model underestimates the visibility threshold of pictorial region, because the texture and smooth pixels in the pictorial region share the same threshold.

Due to the distinct differences between pictorial region and computer-generated texture, it is meaningful to divide a SC image into screen content set (SCS) and non-screen content set (non-SCS) for JND evaluation. According to the development of cognitive science, HVS is adept in understanding the content with a simple pattern. In other words, the visual masking effect of a regular pattern is weaker than an irregular pattern [4]. Motivated by this, we investigate the combined effect of edge masking effect and pattern masking effect in non-SCS. Furthermore, human eyes usually pay more attention to salient region, so we employ the visual saliency to adjust the pattern masking effect for more in line with human visual characteristics. Finally, we develop an edge-pattern-based JND

This work was supported in part by the National Natural Science Foundation of China (61701310 and 61902251), in part by Natural Science Foundation of Guangdong 2019A1515010961, and in part by Free Exploration Project for Basic Research of Shenzhen City JCYJ20180305124209486 (Corresponding author: Miaohui Wang).

Xueqin Liu and Miaohui Wang are with Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen Institute of Artificial Intelligence and Robotics for Society, and Shenzhen University, Shenzhen, P.R. China.

Xueqin Liu and Xuanni Zhan contribute equally to this work.

model considering the unique features of different content types for SC images in this work.

The rest of this paper is organized as follows. The details of our proposed JND profile are shown in Section II. Experimental results are given and discussed in Section III. Finally, Section IV concludes the work.

II. The Proposed Method

The visual masking effects of pictorial content and computer-generated SC are different due to their distinct components. We divide a SC image into non-SCS and SCS for precise operation. In SCS, edge masking effect is taken into account by a parametric edge model. Besides the edge masking effect, the pattern masking effect adjusted by a visual saliency map is introduced for non-SCS. Finally, we introduce the edge-pattern-based JND model which can effectively protect edge information and tolerate more noise in pattern complexity content.

A. Edge Masking Effect

Compared with non-SC images, SC images contain prominent edge features. It is desirable to use an accurate model to depict the edge. A step edge x_0 in a SC image can be described by a smoothed unit step edge in one-dimensional representation [7],

$$s(x;b,c,w,x_0)=b+\frac{c}{2}\left(1+erf\left(\frac{x-x_0}{w\sqrt{2}}\right)\right), \quad (1)$$

where b, c, and w denote the edge basis, the edge contrast and the standard deviation of the Gaussian function respectively, and $erf(\cdot)$ is the error function. With the larger value of w, the edge structure is smoother. Thus the parameter w is used to represent the edge width. To get edge points, the image is filtered by convolving the $s(x;b,c,w,x_0)$ with the derivative of the Gaussian filter $g(x;\sigma_d)$,

$$d(x;c,w,\sigma_d)=\frac{c}{\sqrt{2\pi(w^2+\sigma_d^2)}}\exp\left(\frac{-(x-x_0)^2}{2(w^2+\sigma_d^2)}\right). \quad (2)$$

And then the edge points are obtained by finding the local maxima in the filtered output. Parameters w, c, b and edge position x_0 are estimated by sampling the value at $x=0$, a and $-a$.

Such a form of $s(x;b,c,w,x_0)$ is adaptive, so edge luminance, contrast and structure can be reconstructed by using this model with modified parameters [3]. As we all know, the sensitivity of human eyes varies with the background luminance, and the visual threshold increases as the luminance increases. Therefore, the luminance masking effect should be considered [1],

$$T_l(p)=\begin{cases}\alpha_1\cdot\left(1-\sqrt{\frac{\overline{I(p)}}{127}}\right)+\beta & \text{if } \overline{I(p)}\le 127\\ \alpha_2\cdot\left(\overline{I(p)}-127\right)+\beta & \text{otherwise,}\end{cases} \quad (3)$$

where $\alpha_1=17$, $\alpha_2=3/128$ and $\beta=3$. It is worth noting that $\overline{I(p)}$ is an average intensity value calculated within a 5×5 window [2], but it is calculated as $\overline{I(p)}=b+c/2$ when p is an edge pixel. And the luminance adaption of the edge pixel $T_{el}(p)$ in a SC image is also decided by $T_l(p)$. $T_{el}(p)$ is depicted as the difference between $s(p;b,c,w,x0)$ and $s(p;b+T_l(p),c,w,x0)$.

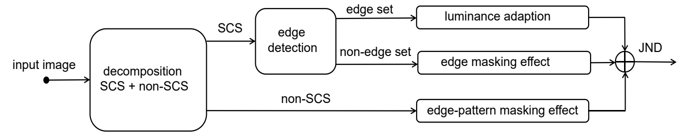


Fig. 1. The diagram of the proposed JND model.

Compared with luminance value, HVS is more sensitive to the contrast luminance. Relatively high contrast is a significant feature of SC images. Therefore, it is essential to explore the edge contrast masking effect. Based on the hypothesis that edge contrast masking effect is only decided by the difference of the normalized contrast, the visibility threshold on edge contrast $T_{ec}(p)$ is calculated as the absolute value of the difference between $s(p;b,c,w,x0)$ and the corresponding value when c is set as the tolerable increased or decreased contrast. In addition, edge structure plays an important role in delivering perceptual information. Edge structural distortion sensitivity is described by the change of edge width Δw. As a result, the structural masking effect $T_s(p)$ of edge set is represented as the absolute value of the difference between $s(p;b,c,w,x0)$ and the corresponding value when $w=w+\Delta w$ with Δw set as around 0.1. More details can be found in [3].

The edge masking effect is calculated by combining the three masking effects with nonlinear additivity model,

$$\begin{cases} T_{lc}(p)=T_{el}(p)+T_{ec}(p)-C_{elc}\cdot\min\{T_{el}(p),T_{ec}(p)\},\\ T_e(p)=T_s(p)+T_{lc}(p)-C_{slc}\cdot\min\{T_{el}(p),T_{ec}(p)\},\end{cases} \quad (4)$$

here C_{elc} and C_{slc} are both suggested as 0.2 [3] to cut down the overlapping among T_{el}, T_{ec} and T_s.

B. Pattern Masking Effect

HVS usually pays less attention to the content with a complex pattern. As analyzed by [4], complex patterns include diverse directions, so a method to calculate the pattern complexity is considering the distribution of orientations. The orientation of each pixel p in the local region R is represented as its gradient direction $\theta(p)$ by using Prewitt kernels. And then, the pattern complexity $C_p(p)$ is obtained by the histogram $H_k(p)$ with the quantitative results of $\theta(p)$. The final pattern masking effect M_p is related to pattern complexity C_p and the luminance contrast C_l, so $M_p=f(C_p)\cdot f(C_l)$.

To get a pattern-based JND model, both the luminance adaptation T_l and the contrast masking M_c from a luminance contrast masking experiment should be considered. Pattern masking effect and contrast masking effect usually exist simultaneously. The stronger one M_s ($M_s=\max\{M_p,M_c\}$) is chosen to express the final visibility threshold. Finally, the pattern-based threshold T_p is described as,

$$T_p(p)=T_l(p)+M_s(p)-C_{ls}\cdot\min\{T_l(p),M_s(p)\}, \quad (5)$$

Fig. 2. Test images from SIQAD. From left to right marked as I1, I2, I3, I4, I5, I6, I7, I8, I9, I10, I11, I12, I13, I14, I15, I16, I17, I18, I19 and I20.

where C_{ls} is suggested as 0.3 [4].

We adopt a visual saliency model based on structural dissimilarity [8] to adjust the pattern masking effect. The input image is divided into a series of patches, and the corresponding saliency value of each patch x is calculated by,

$$S(x) = \left(\sum_{p \in D^x} \left(D^x(p) \cdot W^x(p) \right) \right)^{\theta}, \tag{6}$$

where θ is set as 0.8, W^x is the spatial weighting map which is the distance of patch x to other patches, and D^x is an overall dissimilarity map of patch x. Then the saliency map $S(x)$ is normalized to the range [0,1]. The saliency value of each pixel $s'(p)$ is set the same as the value of the patch x where p is located. The higher the saliency map value is, the more attention HVS pays to it. We set 5 times the mean value of the saliency map as the threshold to separate salient area SA from non-salient area NSA. As a result, we get the pattern-based model adjusted by saliency character as follows,

$$T_{sp}(p) = \begin{cases} T_p(p) \times (1 - s'(p)) & \text{if } p \in SA \\ T_p(p) & \text{otherwise.} \end{cases} \tag{7}$$

C. Edge-pattern-based JND Model

We use a feature-based classification scheme [9] to divide the image into SCS and non-SCS. Firstly, the text region is separated from the SC image by a run-length smearing operation followed by the stripe merging procedure. Secondly, an operator is developed to evaluate the distribution of pixels in the rest of the image. Finally, according to the property of pixel distribution, the pictorial region and the graphic region are distinguished. We regard the pictorial region as non-SCS, and the rest is SCS.

Extensive smooth areas and obvious edges usually occupy the main position in SCS, thus it is necessary to divide SCS into edge set and non-edge set [7]. In non-edge set of SCS, just luminance adaption is added to get the threshold because of HVS more sensitive to smooth areas. And we use the edge masking effect which makes use of edge properties reasonably to handle the edge points of SCS.

It is noticed that the edge masking effect can inject noise into the textual edge accurately and achieve better masking. However, it underestimates the masking effect of non-SCS which usually contains extensive complex texture. In non-SCS, we propose an edge-pattern-based JND profile to effectively protect edge details by guiding more noise to the irregular areas which possess strong visual masking effects. We combine the edge masking effect and the pattern-based model via NAMM for non-SCS,

$$T_{ep}(p) = C_e \cdot T_e(p) + T_{sp}(p) - C_{esp} \cdot \min\{T_e(p), T_{sp}(p)\}, \tag{8}$$

where C_e is used to adjust the edge masking effect for avoiding overestimating edge threshold, C_{esp} is used to reduce the overlapping between T_e and T_{sp}. Through a subjective experiment, C_e and C_{esp} are set as 0.4 and 0.3 respectively.

Let F_e, F_{ne} and F_{ns} represent edge set of SCS, non-edge set of SCS and non-SCS respectively. The final edge-pattern-based JND map of the total SC image is given in (9), and the model's diagram is shown in Fig. 1.

$$T_{JND}(p) = \begin{cases} T_e(p) & \text{if } p \in F_e \\ T_l(p) & \text{if } p \in F_{ne} \\ T_{ep}(p) & \text{if } p \in F_{ns}. \end{cases} \tag{9}$$

III. Experimental Results

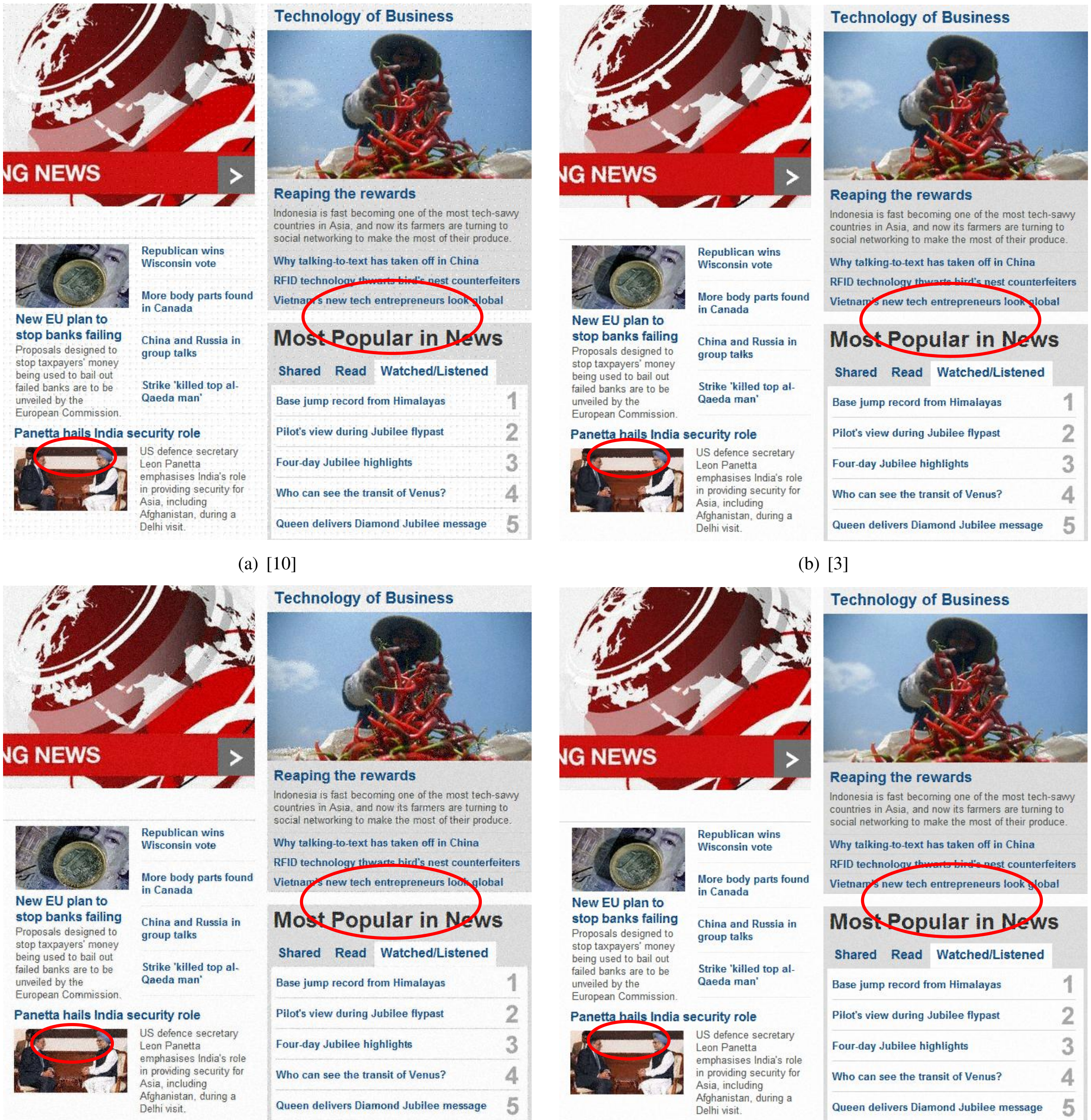
In order to prove the superiority of our proposed JND model, JND noise should be injected into SC images for performance comparison in pixel domain,

$$I'(p) = I + \alpha \times T_{JND}(p), \tag{10}$$

where I is the original image while $I'(p)$ is the contaminated image, α takes +1 or -1 randomly. The performance of the proposed JND model will be compared with three JND models [10], [3] and [4]. There are 20 test SC images in the experiment form the screen content quality assessment database SIQAD [11], which is widely used in SC images processing [12], and the original images are shown in Fig. 2.

A. Distortion Toleration Capability Comparison

The mean square error (MSE) is used to represent the energy of the JND signal and measure the error tolerance ability. Good performance for a better model should show higher MSE values with the same perceptual quality. Table I shows the MSE values of the proposed model and the other three JND profiles. It can be seen that among these models, the proposed profile has the highest JND energy. Compared with [3] and [4], the average MSE value of the proposed model increase 4.68 and 24.27 respectively. It indicates that our model achieves an effective combination of the edge masking effect and pattern masking effect.



(a) [10] (b) [3]

(c) [4] (d) proposed

Fig. 3. Noise-contaminated image I1 comparison among four JND models.

B. Perceptual Quality Comparison

For an intuitive comparison, we take I1 as an example. The JND noise generated by the four models is added to the image with MSE set as 85 randomly, as shown in Fig. 3. The visual quality of Fig. 3(a) is significantly reduced. In the smaller red circle of Fig. 3(b), the JND threshold of the edge in non-SCS is overestimated. The model [3] is weak in processing the edge of SCS leading to obvious distortion, and the larger red circle of Fig. 3(c) illustrates this situation. The proposed JND model preserves edge information in SCS availably, and adjusts the edge masking effect to prevent edge visual threshold from too large in non-SCS. And the noise added by the proposed model is less noticed, as shown in Fig. 3(d).

For a more convincing evaluation, a subjective quality assessment is carried out so as to demonstrate our model is closer to human visual characteristics. We use the adjectival categorical judgment method that two noise-injected images are juxtaposed on the screen for comparison. Table II shows the description of subjective feelings and corresponding quantitative scores. Twenty subjects are invited to participate in

TABLE I
MSE COMPARED WITH ORIGINAL IMAGE

Image	[10]	[3]	[4]	proposed
I1	14.34	79.60	47.02	86.62
I2	17.87	74.65	52.88	81.69
I3	16.02	52.79	25.75	81.02
I4	18.33	57.50	43.82	64.79
I5	15.59	58.42	46.26	62.84
I6	13.11	71.63	49.23	89.05
I7	14.98	73.44	35.13	71.37
I8	15.87	68.66	46.11	51.24
I9	15.17	54.09	50.34	58.68
I10	16.80	82.84	44.86	86.73
I11	16.11	62.02	42.66	69.41
I12	14.66	79.17	44.35	66.21
I13	14.37	68.60	51.81	87.41
I14	16.80	84.55	60.01	75.60
I15	19.17	102.93	64.75	87.71
I16	12.77	50.47	80.94	64.98
I17	11.10	55.11	42.42	62.07
I18	12.16	93.04	48.56	85.67
I19	9.74	81.09	69.17	87.40
I20	15.17	50.12	62.94	73.89
Average	**15.00**	**70.04**	**50.45**	**74.72**

TABLE II
COMPARATIVE SCORES FOR SUBJECTIVE EVALUATION

Subjective Score	Description
3	Left is much better than right
2	Left is better than right
1	Left is slightly better than right
0	Left has same quality as right
-1	Left is slightly worse than right
-2	Left is worse than right
-3	Left is much worse than right

the subjective test, and they are asked to give their subjective quality scores for each pair of compared contaminated images according to Table II.

The subjective comparison results are the average perceived quality scores of our model minus the average scores of the other models, so positive values indicate the proposed model has better performance, as shown in Table III. It is obvious that most values are positive, which implies our JND profile performance stays ahead of the other three JND models. The study further demonstrates the proposed model is more consistent with HVS.

IV. Conclsion

In this work, we have proposed a novel JND model for SC images. The contributions are that the SC image is decomposed into SCS and non-SCS for further processing because of their different visual effects, and the combination masking effect of the edge masking effect and pattern masking effect is investigated. Meanwhile, visual saliency is added for fitting in with HVS. And experimental results demonstrate that the new JND model can tolerate more distortion at the same perceptual quality.

TABLE III
SUBJECTIVE QUALITY EVALUATION RESULTS

Image	[10]	[3]	[4]
I1	1.10	0.55	0.80
I2	0.90	0.50	0.70
I3	1.20	0.30	1.00
I4	0.80	0.10	0.50
I5	0.70	0.05	0.40
I6	1.00	0.20	0.30
I7	0.90	0.10	0.40
I8	1.10	-0.30	0.55
I9	1.00	0.35	0.20
I10	1.30	0.10	0.40
I11	1.00	0.30	0.20
I12	0.90	-0.35	0.50
I13	1.20	0.00	0.90
I14	1.20	-0.40	0.10
I15	1.60	0.40	0.30
I16	0.90	0.50	-0.30
I17	0.60	-0.35	0.60
I18	0.75	0.05	0.10
I19	0.70	0.25	0
I20	0.70	0.20	-0.15
Average	**0.98**	**0.13**	**0.38**

References

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

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

[3] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3838–3851, 2016.

[4] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, and C.-C. J. Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2682–2693, 2017.

[5] H. Hadizadeh, A. Rajati, and I. V. Bajić, "Saliency-guided just noticeable distortion estimation using the normalized laplacian pyramid," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1218–1222, 2017.

[6] Z. Zeng, H. Zeng, J. Chen, J. Zhu, Y. Zhang, and K.-K. Ma, "Visual attention guided pixel-wise just noticeable difference model," *IEEE Access*, vol. 7, pp. 132 111–132 119, 2019.

[7] W. Zhang and W.-K. Cham, "Single-image refocusing and defocusing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 873–882, 2011.

[8] Y. Li and X. Mou, "Saliency detection based on structural dissimilarity induced by image quality assessment model," *Journal of Electronic Imaging*, vol. 28, no. 2, p. 023025, 2019.

[9] K.-C. Fan, C.-H. Liu, and Y.-K. Wang, "Segmentation and classification of mixed text/graphics/image documents," *Pattern Recognition Letters*, vol. 15, no. 12, pp. 1201–1209, 1994.

[10] X. Zhang, W. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Processing*, vol. 85, no. 4, pp. 795–808, 2005.

[11] H. Yang, Y. Fang, Y. Yuan, and W. Lin, "Subjective quality evaluation of compressed digital compound images," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 105–114, 2015.

[12] J. Lin, M. Wang, and W. Xie, "A lightweight quality assessment of screen content images using directional derivative filters," in *IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*, 2018, pp. 292–296.