A Novel Foveated-JND Profile Based on an Adaptive Foveated Weighting Model

Hongkui Wang[1], Li Yu[1], Shengwei Wang[1], Guangjing Xia[2], Haibing Yin[3]

[1] *School of Electronic Information and Communications Huazhong University of Science and Technology, Wuhan, China*

[2]*College of Information Engineering China Jiliang University, Hangzhou, China*

[3] *School of Communication Engineering Hangzhou Dianzi University, Hangzhou, China*

hkwang@hust.edu.cn, hustlyu@hust.edu.cn

***Abstract*—The aftereffect of visual attention (VA) and foveated masking (FM) are two important HVS characteristics not fully exploited in conventional foveated JND (FJND) models. This paper presents a fixation point estimation method to build adaptive foveated weighting model, by which an improved FJND profile is developed with VA and FM effects considered. The proposed FJND profile overcomes two limitations of the conventional FJND profiles. On one hand, conventional FJND profiles can't get accurate positions of fixation points. This paper proposes a fixation prediction algorithm based on the distribution of cones to identify fixation points and regions. On the other hand, conventional FJND profiles did not incorporate the VA effect. By fully exploiting the VA and FM effects, this work proposes an adaptive foveated weighting model, which is developed as a function of the fixation intensity and retinal eccentricity. Here, the fixation intensity is estimated from the saliency map which is modelled by the Gaussian Mixed Model. With the proposed weight model, a new FJND profile is developed. Experimental results show that the proposed profile tolerates more distortion at the same perceptual image quality compared with other JND profiles.**

***Index Terms*— just noticeable distortion, foveated masking effect, visual attention effect.**

I. INTRODUCTION

Just noticeable distortion (JND), as the name implies, is the minimum distortion visible to the human visual system (HVS). Since the JND profile reflects the perception characteristics of HVS directly, the JND profile has widely been used in perception-based image/video processes [1]-[2]. In 1992, Ahumada *et al*. [3] proposed the first DCT-based JND profile. The JND threshold of each DCT component is estimated by the spatial contrast sensitivity function (CSF). Then, Watson proposed an improved DCT-based JND profile called the DCTune which incorporates the luminance adaptation (LA) effect and the contrast masking (CM) effect [4]. For the sake of operating efficiency, Chou *et al*. [5] proposed the first pixel-based JND profile with LA and CM effects.

After that, many researchers have paid more attention on the estimation of LA and CM effects. In terms of LA effect, Bae *et al* [5] revealed that the LA effect of HVS depends not only on background luminance but also on frequency. As for the CM effect, the edge pixel density metric [6] and directional high-pass filtering based texture complexity metric [1] are adopted to estimate the intensity of the CM effect.

In addition, other important characteristics in the HVS, such as foveated masking (FM) effect, temporal masking (TM) effect are incorporated into JND profiles for procuring the more accurate JND threshold [1]. It is efficient to improve the accuracy of JND profiles by considering these effects. It is worth noting that the FM effect plays an important role in estimating JND thresholds. Psychophysical experiments show that the visual acuity decreases with increased retinal eccentricity. Wang *et al*. [7] deduced a foveated weighting model from a contrast threshold formula. According to this model, Chen *et al*. [1] obtained a classical FJND profile. Bae *et al*. [5] proposed a new FJND profile considering the aftereffect of FM and TM effects. Nevertheless, the accuracy of the FJND profile is limited because the positions of fixation points are estimated inaccurately. Furthermore, the visual attention (VA) effect is not incorporated into the FJND profile completely.

To overcome aforementioned limitations, we propose an improved FJND profile which integrates VA and FM effects. Specifically, Considering the distribution of cones, we divide the image into fixation regions and non-fixation regions using a prediction algorithm. The position of the fixation point in each fixation region is obtained accurately. In order to estimate the fixation intensity of each fixation process, the saliency map of image is modelled with the Gaussian Mixed Model. The fixation intensity of each fixation process can be represented by the normalized increase of the saliency in each fixation region. The adaptive foveated weighting model is formulated as the function of the fixation intensity and the eccentricity. Finally, the weighting model is incorporated into our FJND profile.

II. FOVEATION REGIONS AND FIXATION POINTS PREDICTION

The accuracy of FJND profiles is limited by fixation points and fixation regions prediction. Usually, the fixation point can be obtained from attention models. In this section, an elaborate attention model is reviewed at first. Then, a prediction algorithm is proposed based on the distribution of cones and the attention model.

A. Visual attention model

Empirical results indicate that eye movement is directed by visual information of interest. Specifically, eyes move to find the point including interests and concerns during the saccade process and focus on the interesting area during the fixation

978-1-5386-4458-4/18/$31.00 ©2018 IEEE

process [10]. The cognitive process of the brain is performed by visual information obtained in eye fixation, which means the noise is much easier to be perceived in the fixation region. The fixation region is firstly selected by HVS with bottom-up and top-down scan before further processing [9].

In this work, we adopt the visual attention model proposed in [8] (i.e., Judd's model). It is derived from both bottom-up, space-based contrast stimuli (e.g., texture, intensity, orientation) and top-down, object-based features (e.g., face, people, text). In addition, when humans take pictures, they naturally frame an object of interest near the center of the image. The center prior is also one of major features in this attention model.

In typical FJND profiles, the position of the fixation point is estimated by a bottom-up attention model [9] (i.e., Itti's model). The performance comparison of two attention models is shown in Fig.1. It is obvious that Judd's model is better aligned with the characteristics of HVS. Fig.1-(b) indicates that viewers fixate on the face. Actually, semantically meaningful regions (e.g., face, people, text) attract the attention of viewers and the distortion in these regions will be more visible than others.

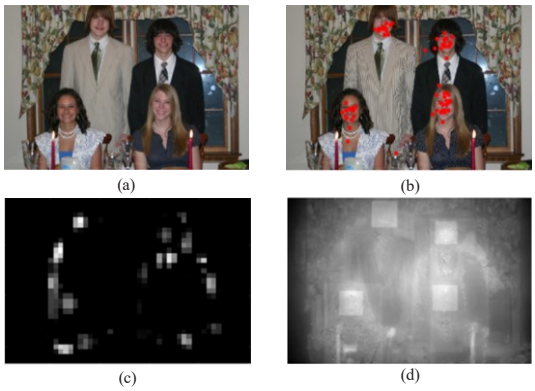


Fig.1 The performance comparison of two attention models: (a) Original image; (b) gaze point locations; (c) Itti' saliency map; (d) Judd's saliency map.

B. The Fixation region prediction algorithm

Fixation points and regions can be predicted from the saliency map of the image. In order to obtain predictive fixation points and regions, we count the measured fixation point number and the maximal salient value in each fixation region. The measured fixation point number can be obtained from Judd's eye tracking database [8].

According to the study of HVS foveation behavior, the visual acuity decreases with increased eccentricity [1]. This is because the density of photoreceptor cells drops with the increased retinal eccectricity [1]. In all photoreceptor cells, Cones,or cone cells play the major role in color vision and function best in relatively bright light. The visual acuity is determined by the distribution of cones to a great extent. In structure, the fovea is surrounded by the parafovea belt and the perifovea outer region. The density of cones in perifovea region is lower than that of the parafovea region and the visual acuity of perifovea is below the optimum. So, the fixation region in image plain means the projection area of the parafovea and fovea in the retina as shown in Fig.2-(a).

$$d = v \times \tan\theta_k \,/\, pw\,(pixel) \tag{1}$$

where d is the radius of the fixation region. v is the viewing distance which is equal to 1.75 times screen height (h). θ_k is the retinal eccentricity which is equal to 0.5 times visual angle. The image resolution is W×H. pw is unit pixel width and equal to h/H. The visual angle of the parafovea region is 8 degrees approximately. The relationship between the measured fixation point number ($FN(k)$) and the maximal salient value ($S(x_k,y_k)$) in k-th region is shown in Fig.2-(b). The k-th region which is determined as a fixation region must have at least one fixation point. As show in Fig.2-(b), the maximal salient value of fixation region should be bigger than a fixed salient threshold (i.e., 0.8). According to the fixed salient threshold and the size of fixation region, the image is divided into fixation regions and non-fixation regions. The position of the predictive fixation point in k-th fixation region is the (x_k,y_k). As shown in Fig.3, our method predicts fixation points and fixation regions more accurately.

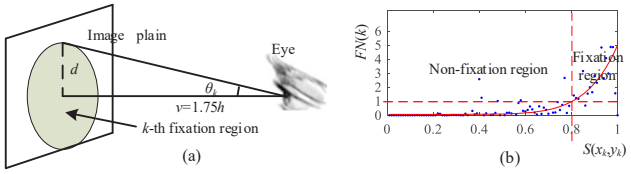


Fig.2 The fixation region: (a)View geometry; (b) $FN(k)$ versus $S(x_k,y_k)$.

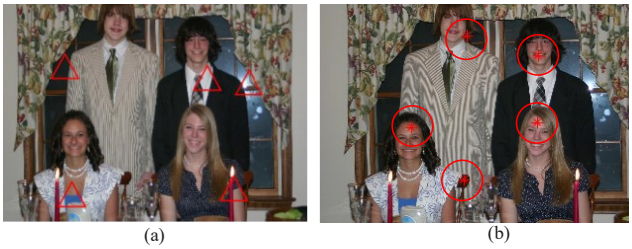


Fig.3 Fixations prediction: (a) typical prediction; (b) our prediction algorithm.

III. The Proposed FJND Profile with VA Effect

The visual acuity increases with the increased fixation intensity in the fixation region. In order to compute the fixation intensity in each fixation process, the saliency map is modelled with the Gaussian Mixed Model at first. The fixation intensity of each fixation process is represented by the normalized increase of the saliency in each fixation region. Then, the adaptive foveated weighting model is defined as the function of the fixation intensity and the retinal eccentricity. At last, the adaptive model is incorporated into our FJND profile.

A. The fixation intensity estimation of each fixation process

In the process of Judd's model training, the fixation map (i.e., distribution of fixation points) is selected as the ground truth. The fixation intensity can be estimated from the saliency map.

Judd's attention model is constructed based on the center prior, space-based and object-based features. Theoretically, the saliency map of the image can be divided into the basic saliency map and the residual saliency map. The basic saliency map is determined by the center prior feature and the residual saliency map is determined by space-based and object-based features. The Gauss Mixture Model (GMM) refers to the linear

combination of multiple Gauss distribution functions. GMM can fit any type of distribution in theory. Therefore, we use the GMM to fit the saliency map ($S(z)$):

$$S(z) = \sum_{z=(x,y)}^{N} \Psi_k(z \mid \mu_k, \Sigma_k) \tag{2}$$

$$\Psi_k(z \mid \mu_k, \Sigma_k) = \frac{\kappa_k}{\sqrt{2\pi \cdot \mid \Sigma_k \mid}} \cdot \exp\left(-\frac{1}{2}(z-\mu_k)^T \Sigma_k^{-1}(z-\mu_k)\right) \tag{3}$$

where z is the pixel index. $\Psi_k(z|\mu_k,\Sigma_k)$ is the k-th component and κ_k is the mixture coefficient. μ_k is the coordinates of the k-th fixation point (i.e., $\mu_k=(x_k,y_k)$). $|\Sigma_k|$ is the determinant of Σ_k and equal to $E\{(z-\mu_k)\cdot(z-\mu_k)^T\}$. Without loss of generality, N is set to $K+1$ and K is the number of fixation points in the image. $\Psi_0(z|\mu_0, \Sigma_0)$ represents the basic saliency map and $\Psi_k(z|\mu_k,\Sigma_k)$ represents the k-th residual saliency map ($k>0$). The fixation region usually contains the meaningful image semantics. The fitting saliency map in k-th fixation region can be expressed as follows:

$$S_k(z) = \Psi_0(z \mid \mu_0, \Sigma_0) + \Psi_k(z \mid \mu_k, \Sigma_k) \quad and \quad \Sigma_k = \begin{bmatrix} \sigma_{xk}^2 & 0 \\ 0 & \sigma_{yk}^2 \end{bmatrix} \tag{4}$$

R_k is the k-th fixation region. σ_{xk} and σ_{yk} can be estimated from the original saliency map. The basic saliency map and the k-th residual saliency map are shown in Fig.4. The increase of saliency in k-th fixation region is defined as the integral of the $\Psi_k(z|\mu_k,\Sigma_k)$. Considering the effect of the center prior, the increase of saliency is expressed by formula (5). The fixation intensity of k-th fixation process ($NS(k)$) is expressed as the normalized saliency increase of this region.

$$DS(k) = \Psi_0(z_k) \cdot \iint_{z \in R_k} \Psi_k(z \mid \mu_k, \Sigma_k) dz = 2\pi \cdot \Psi_0(z_k) \cdot \Psi_k(z_k) \sigma_{xk} \sigma_{yk} \tag{5}$$

$$NS(k) = DS(k) \Big/ \max_{k=1}^{K}(DS(k)) \tag{6}$$

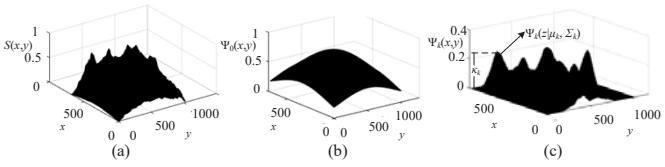


Fig.4 Saliency Modelled: (a) Original saliency map ;(b) Basic saliency map ($\Psi_0(z|\mu_0, \Sigma_0)$); (c) Residual saliency map ($\Psi_k(z|\mu_k, \Sigma_k)$).

B. The proposed foveated weighting model

The spatial foveated weighting model is obtained by normalizing the cutoff frequency model. In real-world digital images, the maximum perceived frequency is also limited by the display cutoff frequency. Thus, the model is expressed as

$$W_f(v,\theta_k) = \frac{\min\left(f_c(v,\theta_k), f_d(v)\right)}{f_c(v,0)} = \min\left(\frac{e_2}{\theta_k + e_2}, \frac{\alpha \pi v / pw}{360 \ln(1/CT_0)}\right) \tag{7}$$

where, e_2, α, CT_0, are the model paraments which are set to 2.3,0.106, and 1/64 [1]. θ_k is the eccentricity in k-th fixation region. Considering the VA effect, the foveated weighting model is defined as the product of the fixation intensity and the traditional foveated weight. Due to the visual acuity of each pixel is affected by multiple fixations, the final foveated weight of z-th pixel is expressed as follows:

$$W_f'(z) = \max_{k=1}^{K}\left(NS(k) \cdot W_f(v,\theta_k,z)\right) \tag{8}$$

As shown in Fig.5, compared with the traditional foveated weighting model, the proposed model considering the VA effect is more consistent with the HVS.

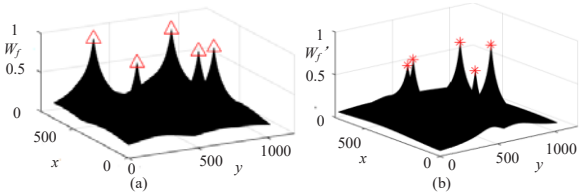


Fig.5 Foveated weighting model comparisons: (a) traditional foveated weighting model; (b) proposed foveated weighting model.

C. The FJND profile with the adaptive weighting model

In this paper, we focus on the pixel-based foveated-JND profile. The proposed JND profile is established based on a multiplicative model as the product of two modulation factors:

$$J(z) = J_s(z) \cdot F_{FM}(z) \tag{9}$$

where $J_s(z)$ is the spatial pixel-JND threshold which is estimated with the NAMM model [12]. $J_s(z)$ is defined as the function of LA and CM factors:

$$J_s(z) = LA(z) + CM(z) - C^{gr} \times \min\left(LA(z), CM(z)\right) \tag{10}$$

$$LA(z) = \begin{cases} 17 \times \left(1 - \sqrt{g(z)/127}\right), & if\ g(z) \le 127 \\ 3 \times \left(g(z) - 127\right)/128 + 3, & else \end{cases} \tag{11}$$

$$CM(z) = \left(0.01 \cdot g(z) + 11.5\right) \times \left(0{,}01 \cdot G(z) - 1\right) - 12 \tag{12}$$

where $g(z)$ is the background luminance of pixel z, i.e., the mean luminance of a small region, $G(z)$ is the maximum edge height of its neighborhood [12]. C^{gr} is the gain reduction parameter due to the overlapping between LA and CM effects. C^{gr} is set to 0.3 (the same as in [12]).

According to Chen's psychophysical experiments [1], the visibility threshold is not only eccentricity dependent, but also background-luminance dependent. the FM factor is shown as

$$F_{FM}\left(W_f'(z), g(z)\right) = \left(1 + \left(1 - W_f'(z)\right)^{\gamma}\right)^{f(g(z))} \tag{13}$$

$$f\left(g(z)\right) = 0.5 + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left(\log_2(g(z)+1) - \mu\right)^2\right) \Big/ \left(2\sigma^2\right) \tag{14}$$

here, $f(g(z))$ is a function of the average intensity value of the block. The parameters are set γ=1, μ = 7 and σ =0.8 [1].

IV. Experimental Results

In order to evaluate the effectiveness of proposed foveated JND profile, we tested three developed pixel-JND profiles: Chen's [1], Wu's [12] and our profiles. Chen's profile is the classical foveated pixel-JND profile. Wu's profile is the state-of-the-art spatial pixel-JND profile. Noise is added to each pixel in an image according to [1] such that

$$F'(z) = F(z) + S_r(z) \times J(z) \tag{15}$$

where F is the original image and F' is the JND noise contaminated image. $S_r(z)$ is a bipolar random noise of ±1. The PSNR is used to measure the capability toleration of the

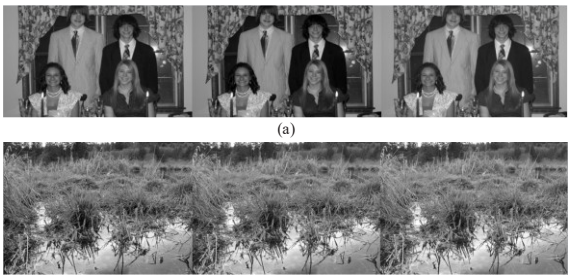
JND profiles. With the same perceptual quality, the lower PSNR is, the more accurate the JND profile is. Eighteen images which contain different high-level image semantics (e.g., face, people, scenery and building) are chosen for testing. The test images are shown in Fig.6.

Fig.6 Test images with different high-level image semantics (I_1,I_2⋯I_{18}).

Table I shows PSNR results of test images with three JND profiles. Specifically, compared with Chen's and Wu's profiles, the proposed FJND profile reduces the average PSNR 1.35dB and 4.32dB, respectively. It also can be found that the average PSNR values of face, people, building, animal and text images are reduced 0.82dB, 0.97dB, 0.40dB, 1.50dB and 0.92dB compared with Chen's profile while the average PSNR value of scenery is reduced 3.50dB. The image contained the meaningful regions attracts the attention of viewers and its JND thresholds are smaller. Compared with Wu's profile, ours and Chen's profiles reduces the average PSNR 4.32 dB and 2.97dB. This indicates that incorporation of VA and FM effects improves the performance of JND profiles. According to the Table I, our JND profile is better in PSNR reduction. Fig.7 shows the images after the JND guided noise injected by three profiles. The noise is hardly noticeable in the resultant images. It is obvious that our profile is more conforming to the characteristics of HVS. The computing time of three JND models is 1.49s, 5.74s and 3.33s on average, respectively. In summary, the JND profiles incorporated VA and FM effects are more accurate with higher complexity. Compared with Chen's model, our model tolerates more distortion at the same perceptual image quality with lower complexity.

TABLE I PSNR COMPARED WITH ORIGINAL IMAGE (DB)
(With the same perceptual quality, the lower PSNR is, the more accurate the JND profile is.)

semantics	images	Wu's	Chen's	ours	ours vs Chen's	ours vs Wu's
face	I_1	35.32	29.87	29.34	-0.82	-6.39
	I_2	32.91	28.64	28.08		
	I_3	36.69	29.70	28.33		
people	I_4	32.26	28.24	27.00	-0.97	-4.87
	I_5	31.28	29.36	28.33		
	I_6	34.15	28.40	27.77		
building	I_7	32.87	27.05	28.12	-0.40	-4.32
	I_8	33.00	29.82	27.14		
	I_9	31.83	29.07	29.46		
scenery	I_{10}	30.50	29.91	25.84	-3.50	-3.62
	I_{11}	27.86	28.77	27.27		
	I_{12}	31.93	31.26	26.33		
animal	I_{13}	31.56	29.59	28.88	-1.50	-1.91
	I_{14}	29.17	28.29	26.80		
	I_{15}	27.81	29.43	27.14		
text	I_{16}	36.63	30.23	29.03	-0.91	-4.84
	I_{17}	31.99	28.83	27.80		
	I_{18}	31.32	29.11	28.60		
Average		32.17	29.20	27.85	-1.35	-4.32



(a)

(b)

Fig.7 Resultant images after noise injection: (a) I_6 image; (b) I_{12} image. Left: by Wu's JND profile, Min: by Chen's JND profile, Right: by our JND profile.

V. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61871437).

VI. CONCLUSION

In this paper, we proposed an adaptive foveated weighting model as well as a novel FJND profile. In view of the distribution of cones, fixation points and regions are identified effectively based on our prediction algorithm. Considering the aftereffect of VA and FM effects, the adaptive weighting model is built and incorporated into the FJND profile. Experimental results show that the proposed profile tolerates more distortion at the same perceptual image quality.

REFERENCES

[1] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol*., vol. 20, no. 6, pp. 806–819, Jun. 2010.

[2] X. Zhang, S. Wang, K. Gu. W. Li and S. Ma, "Just-Noticeable Difference-Based Perceptual Optimization for JPEG Compression" *IEEE Signal Processing Letters*, vol. 24, no.24, pp.96-100, Jan. 2017.

[3] A. J. Ahumada, Jr. and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proc. SPIE, Human Vision, Visual Processing, and Digital Display III*, vol. 1666, pp. 365–374, 1992.

[4] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Sid International Symposium Digest of Technical Papers*, vol. 24, pp. 946–946, 1993.

[5] S. Bae and K. Munchurl, "A DCT-Based Total JND Profile for Spatiotemporal and Foveated Masking Effects", *IEEE Trans. Image Process.*, vol. no. 6, pp. 1196 – 1207, 2017.

[6] W. Wan, J. Wu, X. Xie, G. Shi. "A Novel Just Noticeable Difference Model via Orientation Regularity in DCT Domain", *IEEE Access*, vol. 5, pp.22953-22964, 2017.

[7] W. S. Geisler and J. S. Perry. "A real-time foveated multiresolution system for low-bandwidth video communication". in *Proc. SPIE*, vol.3299, pp. 294–305, 1998.

[8] T. Judd, K. Ehinger, F. Durand, A. Torralba, "Learning to Predict Where Humans Look," in *Proc. IEEE*, vol. 30, no. 2, pp. 2106–2113, 2010.

[9] L. Itti and C. Koch. "A saliency-based search mechanism for overt and covert shifts of visual attention". *Vision Research*, vol.40, pp. 1489–1506, 2000.

[10] T. Oosuga, M. Tanaka, H. Inoue, and Y. Niiyama, "A study on eye fixation time distribution with and without subjective evaluation of food and related pictures," *Sice Conference*, vol.282, no. 5, pp.481-485,2012.

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

[12] J. Wu, L. Li, W. Dong, G. Shi, W. Lin and C.-C. Jay Kuo, "Enhanced Just Noticeable Difference Model for Images with Pattern Complexity," *IEEE Trans. Image Process*., vol. 26, no. 6, pp. 2682–2693, Jun. 2017.