IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2019, accepted August 30, 2019, date of publication September 5, 2019, date of current version September 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939569

Visual Attention Guided Pixel-Wise Just Noticeable Difference Model

ZHIPENG ZENG[1,2], HUANQIANG ZENG[1,2], (Senior Member, IEEE), JING CHEN[1,2], (Member, IEEE), JIANQING ZHU[3], (Member, IEEE), YUN ZHANG[4], (Senior Member, IEEE), AND KAI-KUANG MA[5], (Fellow, IEEE)

[1]School of Information Science and Engineering, Huaqiao University, Xiamen 361021, China
[2]Xiamen Key Laboratory of Mobile Multimedia Communications, Xiamen 361021, China
[3]College of Engineering, Huaqiao University, Quanzhou 362021, China
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[5]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Huanqiang Zeng (zeng0043@hqu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871434, Grant 61802136, and Grant 61602191, in part by the Natural Science Foundation for Outstanding Young Scholars of Fujian Province under Grant 2019J06017, in part by the Natural Science Foundation of Fujian Province under Grant 2017J05103, in part by the Fujian-100 Talented People Program, in part by the High-level Talent Innovation Program of Quanzhou City under Grant 2017G027, in part by the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University under Grant ZQN-YX403 and Grant ZQN-PY418, and in part by the High-Level Talent Project Foundation of Huaqiao University under Grant 14BS201, Grant 14BS204, and Grant 16BS108.

ABSTRACT The *just noticeable difference* (JND) models in pixel domain are generally composed of *luminance adaptation* (LA) and *contrast masking* (CM), which takes *edge masking* (EM) and *texture masking* (TM) into consideration. However, in existing pixel-wise JND models, CM is not evaluated appropriately since they overestimate the masking effect of regular oriented texture regions and neglect the visual attention characteristic of human eyes for the real image. In this work, a novel JND model in pixel domain is proposed, where *orderly texture masking* (OTM) for regular texture areas (also called orderly texture regions) and *disorderly texture masking* (DTM) for complex texture areas (also called disorderly texture regions) are presented based on the orientation complexity. Meanwhile, the visual saliency is set as the weighting factor and is incorporated into CM evaluation to enhance JND thresholds. Experimental results indicate that compared with existing relevant JND profiles, the proposed JND model tolerates more distortion in the same perceptual quality, and brings better visual perception in the same level of the injected JND-noise energy.

INDEX TERMS Just noticeable difference, orientation complexity, visual attention.

I. INTRODUCTION

Images/Videos are commonly explored in various multimedia services and become an indispensable part in people's daily life. To provide a high quality of multimedia experience, there are many researches devoting to the development of image/video processing, image/video coding, and robust transmission technologies. Since human eyes are the ultimate receivers of images/videos in general, how to describe perceptual characteristics of human vision more precisely and efficiently has been drawing lots of attentions from both academic and industrial societies [1]–[4].

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang.

As is known, an important perceptual characteristic of *human visual system* (HVS) is that it presents limited visual sensitivity to the images/videos, only the pixel changes above a certain visibility threshold can be observed by human eyes [1]. To model this perceptual characteristic, the *just noticeable difference* (JND) model has been presented, in which the smallest perceptual visual threshold values of the human eyes for the input image are obtained [5], [6]. Therefore, the JND models are widely applied on variable kinds of perceptual-oriented image/video related tasks, such as perceptual compression [7]–[9], perceptual quality assessment [10], [11], watermarking [12], display [13], to name a few.

Existing JND models can be roughly classified into two categories according to the JND threshold calculating

domain: the pixel-wise JND models (e.g., [1], [14], [15]) and the subband-based (e.g., DCT or wavelet transform) JND models (e.g., [16]–[18]). Compared to the subband-domain JND models, the pixel-wise ones can be calculated directly and avoid the subband transformation, which would be more convenient and cost-effective to estimate the JND thresholds. Base on that, the objective of this work is to design an effective pixel-wise JND model to accurately describe characteristics of HVS on images. Pixel-wise JND models commonly take the *luminance adaptation* (LA) and *contrast masking* (CM) into account. Note that LA reflects the masking effect of the HVS in respect of the luminance of the background, while CM reflects the visibility attenuation of one contrast at the presence of another contrast. Some early JND models, like the one developed by Chou and Li [19] overlooked the interaction between these two masking effects, resulted in a rough JND estimation. Based on Chou and Li [19], Yang *et al.* [20] exploited a nonlinear additivity model to reduce the overlapping effects between LA and CM. Since these two methods overestimated the masking effects in the edge regions, Liu *et al.* [15] decomposed one input image into two images, one is named structural image and the other is the textural image, followed by performing *edge masking* (EM) estimation and *texture masking* (TM) estimation, respectively. Considering that the CM effect is not comprehensively evaluated, Wu *et al.* [21] proposed the disorderly concealment effect based on free-energy principle for JND estimation. Motivated by the observation that the HVS is highly sensitive to the repeated pattern in visual signal, Wu *et al.* [1] introduced the concept of pattern complexity to decide the total masking effects. With image saliency information, Hadizadeh *et al.* [22] developed a saliency-guided JND model by the normalized Laplacian pyramid.

According to the research of cognitive psychology and neuroscience, HVS is usually motivated to fetch the visual regularities for perception and understanding [23], [24]. And in the local receptive field of the image, the visual cortex displays distinct orientation selectivity mechanism for visual content representation and extraction [25], [26], which also indicates that orientation regularity (also called low orientation complexity) plays a significant role in the process of visual perception. Inspired by these, in our JND model, the textural image proposed by Liu *et al.* [15] is further decomposed into two portions according to the regularity of the texture. For regular texture regions, an *orderly texture masking* (OTM) is exploited; for disorderly textural portions, the *disorderly texture masking* (DTM) is used. Furthermore, based on the visual attention mechanism, the higher the visual saliency the higher priority of being processed by HVS, and people's eyes will focus on the saliency areas for a relatively long time. Therefore, the visual saliency regarded as the adjustment factor is incorporated into the proposed CM estimation. Combined with *luminance adaptation* (LA), the proposed JND model is established. Experimental results show that the proposed JND model is well correlated with the HVS perception and outperforms the relevant pixel-wise JND methods.

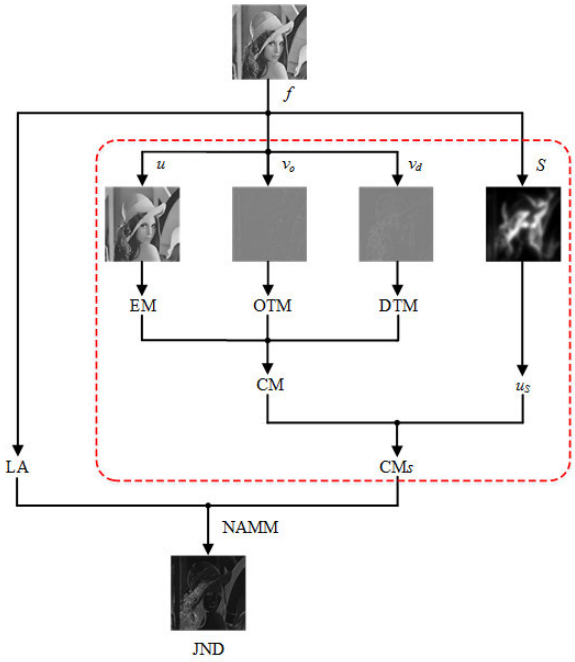


FIGURE 1. The framework of the proposed JND model.

The remaining sections of this paper are arranged as follows. In Section II, the proposed JND model is presented in detail. Section III provides the experimental results and analyses. The conclusion is summarized in Section IV.

II. PROPOSED JND MODEL

As illustrated in Fig. 1, the main body of the proposed pixel-wise JND model consists of three modules, namely *LA*, CM_s and *NAMM*, where *LA* and *NAMM* modules are referred to [14], while CM_s module encircled with the red dashed line is the contribution of this work. In order to predict *contrast masking* (CM) precisely, we estimate *edge masking* (EM), *orderly texture masking* (OTM) and *disorderly texture masking* (DTM) from structural image u, orderly textural image v_o and disorderly textural image v_d, respectively, instead of obtaining CM estimation from the whole image f at first hand. Meanwhile, for the sake of visual attention mechanism of HVS [27], [28], the bottom-up saliency model [29] for non-local spatial redundancy is treated as the weight coefficient to adjust the CM values perceptually.

A. THE CONTRAST MASKING MODEL BASED ON VISUAL ATTENTION

As [30] denotes, contrast masking represents the visibility reduction of one visual component at the presence of another. Based on the visual attention of image/video content, the sensitivity of HVS is diverse in different image areas for CM evaluation. The CM presented by Liu *et al.* [15] is composed of edge masking (EM) for edge regions and texture masking (TM) for textural areas, respectively.

However, TM overestimates the masking effect of homogeneous textural regions. More precise estimation is needed adapting to regular oriented texture region and homogeneous one. Therefore, there are two textural masking estimation in the proposed CM module. One is for regular oriented texture region, named orderly texture masking (OTM); the other is for complex texture region, named disorderly texture masking (DTM). With OTM and DTM, the TM can be estimated more accurately. Moreover, a saliency adjustment factor u_s is introduced concerning about the visual attention of HVS to adjust the CM module.

1) THE RTV MODEL FOR EM MEASUREMENT

It is known that an original image f can be represented by a structural image u (containing large-scale subjects like piecewise smooth and sharp edge) and a textural one v (containing fine-scale details which usually have periodicity and oscillation). That is $f = u + v$. The *relative total variation* (RTV) model is exploited to effectively obtain the structural and textural information of the image [31]. The RTV model is defined as:

$$\arg\min_{u} \sum_{p} \left(u_p - f_p\right)^2 + \lambda \cdot \left(\frac{\mathscr{M}_x(p)}{\mathscr{N}_x(p) + \varepsilon} + \frac{\mathscr{M}_y(p)}{\mathscr{N}_y(p) + \varepsilon}\right) \tag{1}$$

where f and u represent the input image and the output structural image, respectively. p denotes the index for 2-D image pixel. λ is a weighting factor and ε is a small positive value to avoid zero denominator. $\mathscr{M}_x(p)$ and $\mathscr{M}_y(p)$ mean windowed total variations in the x and y directions, which are expressed as:

$$\begin{aligned} \mathscr{M}_x(p) &= \sum_{q \in R(p)} w_{p,q} \cdot \left|(\partial_x u)_q\right| \\ \mathscr{M}_y(p) &= \sum_{q \in R(p)} w_{p,q} \cdot \left|\left(\partial_y u\right)_q\right| \end{aligned} \tag{2}$$

$\mathscr{N}_x(p)$ and $\mathscr{N}_y(p)$ denote the overall spatial variation in the x and y directions, which are defined as:

$$\begin{aligned} \mathscr{N}_x(p) &= \left|\sum_{q \in R(p)} w_{p,q} \cdot (\partial_x u)_q\right| \\ \mathscr{N}_y(p) &= \left|\sum_{q \in R(p)} w_{p,q} \cdot \left(\partial_y u\right)_q\right| \end{aligned} \tag{3}$$

where q belongs to $R(p)$, the rectangular region centered at pixel p. $w_{p,q}$ is a weighting function, which is written as:

$$w_{p,q} \propto \exp\left(-\frac{\left(x_p - x_q\right)^2 + \left(y_p - y_q\right)^2}{2\sigma^2}\right) \tag{4}$$

where σ adjusts the spatial scale of the window, which affects $\mathscr{M}(p)$ and $\mathscr{N}(p)$ directly.

The parameters λ and σ are adjusted to extract the structure image from the original image [31]. The value ranges of λ and σ are set as [0.01, 0.03] and (0, 8], respectively. When λ is larger, the structural image will be fuzzier and the texture details can be retained completely. And the parameter σ plays an opposite role compared to λ. When σ is greater, it can make the structural image keep more fine-scale details and suppress the texture. In this paper, λ and σ are set as 0.01 and 3, respectively, referring to [31]. From Fig. 2a and Fig. 2b, the structural image u and textural image v can be achieved by RTV model.

(a) (b)

FIGURE 2. (a) Structural image u of Lena, (b) Textural image v of Lena.

Therefore, EM estimation for structural image u is calculated as follows,

$$EM^u(x, y) = C_s^u \tag{5}$$

where C_s^u indicates the spatial contrast of u, and C_s denotes the maximum luminance difference within the 5×5 neighborhood of u [19].

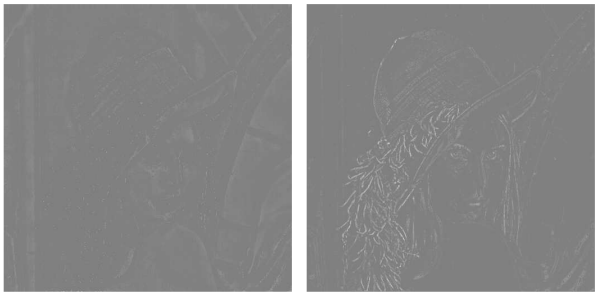
2) THE ORIENTATION COMPLEXITY FOR OTM AND DTM ESTIMATION

Based on the analyses above, the orientation complexity is used to split textural image v obtained by RTV into orderly textural image v_o for OTM estimation and disorderly textural image v_d for DTM estimation, respectivley.

As analysed by [32], the orientation selectivity based pattern can be described as the organization of neighbor pixels. The local perceptive region ψ (3×3) is related to the interactions among the orientation $\theta(x)$ of pixels in ψ. The similarities of pixels preferred orientation is calculated. More specific, if the orientation similarity of region ψ is high, it may be a region with regular orientation. On the contrary, if the similarity of region ψ is low, it may be an irregular orientation region. It has been revealed that dissimilar orientations cause strong masking effect, the higher the dissimilarity, the stronger the masking effect. When orientation difference is larger than a certain threshold, the masking effect is obviously improved.

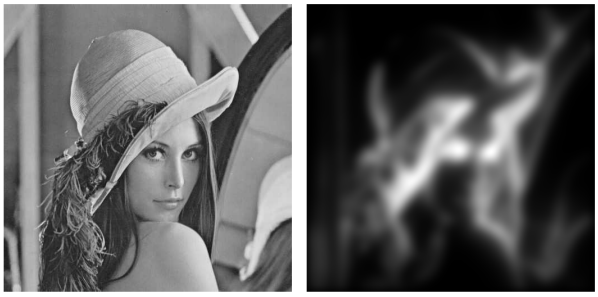
Thus, the complexity $P_C(\mathrm{x})$ of orientation selectivity based pattern of a local region ψ (3×3) is calculated with the histogram $H_m(\mathrm{x})$ of orientations $\hat{\theta}(\mathrm{x})$ by quantifying $\theta(\mathrm{x})$ with the interval $\mathrm{T} = 12^\circ$ [1], [33],

$$P_C(x) = \sum_{m=1}^{M} ||H_m(x)||_0 \tag{6}$$



(a) (b)

FIGURE 3. (a) Orderly textural image v_o of Lena, (b) Disorderly textural image v_d of Lena.



(a) (b)

FIGURE 4. (a) Original Lena image, (b) The saliency map $\hat{S}(x)$ of Lena image.

where $||\cdot||_0$ represents the L_0 norm and M indicates the limit number of $\hat{\theta}(x)$, and the histogram $H_m(x)$ is defined as follows:

$$H_m(x) = \sum_{x \in \psi(x)} \delta(\hat{\theta}(x), m) \tag{7}$$

where $\delta(\cdot)$ represents the pulse function, and for which

$$\delta(\hat{\theta(x)}, m) = \begin{cases} 1, & \text{if } \hat{\theta}(x) = m \\ 0, & \text{if } \hat{\theta}(x) \neq m \end{cases} \tag{8}$$

The results shown in [1] illustrated that the orientation complexity $P_C(x)$ of regular region is low, and for the irregular region, the corresponding complexity $P_C(x)$ is high.

To split the textural image v properly, value "1" is regarded as the threshold of $P_C(x)$ to obtain orderly textural image v_o and disorderly textural image v_d.

$$\begin{cases} v_o, & \text{if } P_C(x) = 1 \\ v_d, & \text{if } P_C(x) \neq 1 \end{cases} \tag{9}$$

When $P_C(x)$ equals to "1", each pixel in the local region ψ (3×3) has similar $\hat{\theta}(x)$, which illustrates that the orderly textural image has the homogeneous pattern complexity. The orderly textural image v_o and the disorderly textural image v_d are shown in Fig. 3a and Fig. 3b.

Hence, OTM and DTM evaluation can be computed as follows:

$$\begin{cases} OTM^{v_o}(x, y) = C_s^{v_o}, \\ DTM^{v_d}(x, y) = C_s^{v_d}, \end{cases} \tag{10}$$

where $C_s^{v_o}$ indicates the spatial contrast for orderly textural image v_o, and $C_s^{v_d}$ denotes the spatial contrast for disorderly textural image v_d.

3) THE SALIENCY ADJUSTMENT FACTOR ESTIMATION

In order to estimate CM values more accurately, visual saliency, which is the perceptual characteristic of HVS, is added to adjust the proposed CM measurement.

The saliency model [29] is adopted to determine the saliency by removing redundant contents instead of measuring the significance.

The visual saliency model is evaluated by:

$$S(x) = \sum_{j=1}^{J} \sum_{k=1}^{K} w_{jk} \hat{H}_{jk}(x) \tag{11}$$

where J and K denote the number of pyramid levels and the number of image channels, respectively. w_{jk} is the normalizing coefficients for each channel and scale, which is set as $w_{jk} = 1/\max_x \hat{H}_{jk}(x)$. As for $\hat{H}_{jk}$, it refers to saliency estimation provided by the redundancy reduction as follows,

$$\hat{H}_{jk} = (1 - \varrho(x))H(x) \tag{12}$$

where $\varrho(x)$ represents the redundancy coefficient of pixel x, and $H(x)$ refers to the entropy of pixel x.

In this paper, $S(x)$ is normalized as $\hat{S}(x) \in [0,1]$ to get the final saliency map [34]. As shown in Fig. 4b, the brighter the region of $\hat{S}(x)$ is, the closer the pixel value of $\hat{S}(x)$ to value "1", and the higher degree of saliency is. Then, a threshold is set as 0.5 to binarize the final saliency map $\hat{S}(x)$ into "saliency" area and "non-saliency" area. Since HVS is more sensitive to changes in the "saliency" area, we use the saliency factor u_S to adjust the CM value adaptively in "saliency" area and "non-saliency" area. The saliency adjustment factor u_S is defined as follows.

$$u_S = \begin{cases} 1 - \hat{S}(x), & \hat{S}(x) \geq 0.5 \\ 1, & \hat{S}(x) < 0.5 \end{cases} \tag{13}$$

4) THE PROPOSED CM MODEL

As aforementioned analyses, the preliminary CM evaluation is calculated as:

$$CM(x, y) = EM^u(x, y) + OTM^{v_o}(x, y) + DTM^{v_d}(x, y) \tag{14}$$

where

$$\begin{cases} EM^u(x, y) = C_s^u \cdot W_e, \\ OTM^{v_o}(x, y) = C_s^{v_o} \cdot W_{v_o}, \\ DTM^{v_d}(x, y) = C_s^{v_d} \cdot W_{v_d} \end{cases} \tag{15}$$

Note that to distinguish the effect of EM, OTM and DTM to the contrast masking, W_e, W_{v_o} and W_{v_d} are regarded as the

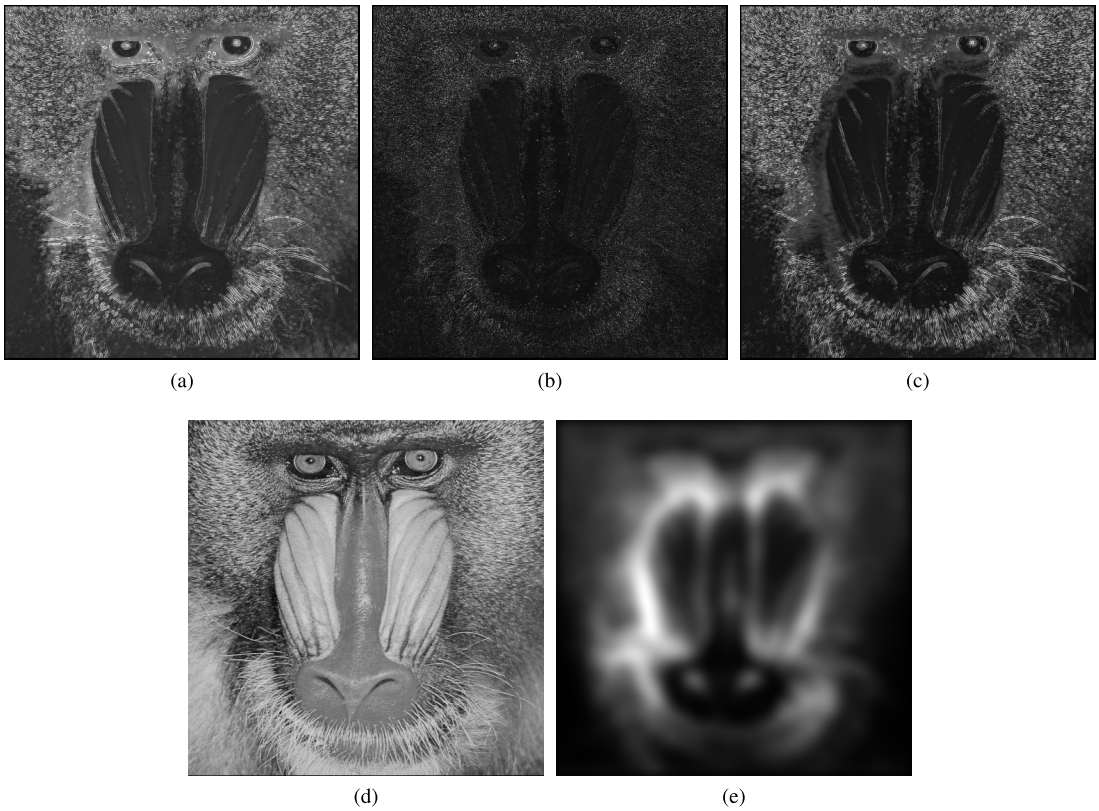


FIGURE 5. (a) JND map of Liu's model [15], (b) JND map of Wu's model [21], (c) JND map of Our proposed model, (d) Original Baboon image, (e) The saliency map of Baboon image.

weight coefficients of the estimation, which are set to 1, 2, and 3, respectively [35].

Combined with the saliency adjustment factor considering the visual attention, the final CM estimation is established as follows:

$$CM_s(x, y) = CM(x, y) \cdot u_S \tag{16}$$

B. LUMINANCE ADAPTATION

It is well known that the human eyes is less sensitive to the distortion of darkness. With the increase luminance, the sensitivity of HVS to image changes may be improved. Therefore, a *luminance adaptation* (LA) model [20] is designed to adapt to the HVS.

$$LA(x, y) = \begin{cases} 17 \times (1 - \sqrt{f(x, y)/127}) + 3, & \text{if } f(x, y) \leq 127 \\ 3 \times (f(x, y) - 127)/128 + 3, & \text{otherwise} \end{cases} \tag{17}$$

where (x, y) is the coordinate in the image.

C. THE PROPOSED JND MODEL

Since LA and CM are usually integrated into the pixel-wise JND model for overall JND estimation via the *nonlinear additivity model for masking* (NAMM) [20], the proposed JND model is established by:

$$JND = LA + CM_s - C_{lc} \times \min\{LA, CM_s\} \tag{18}$$

where C_{lc} is used to settle the overlapping impact between LA and CM_s. As for C_{lc}, it is set as 0.3, same as that in [20].

For illustration purpose, the JND maps for Liu *et al.* [15], Wu *et al.* [21] and the proposed model are displayed in Fig. 5a, Fig. 5b, and Fig. 5c, respectively. From these JND maps, it's obvious that Liu's model overvalued the visual masking in some regions around the baboon's nose with low orientation complexity, and although Wu's model considers the concealment effect in disorderly textural regions, its model still underestimates the visual redundancy in some areas with high orientation complexity, such as the baboon's fur. By contrast, our model shown in Fig. 5c estimates visual redundancy more accurately based on orientation complexity and visual attention of HVS.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETTING

1) TEST IMAGE

In our experiments, twelve commonly-used test images are adopted to comprehensively evaluate the performance of various JND models [1], [15], [36]. These images are of the resolution 512×512 and contain a variety of visual content and spatial complexity, as shown in Fig. 6.

FIGURE 6. The test images. From left to right, [First row: I1, I2, I3, I4], [Second row: I5, I6, I7, I8], [Third row: I9, I10, I11, I12].

2) EVALUATION PROCEDURE

Intuitively, an ideal JND model should be able to tell how to conceal the noise in the image as much as possible with the acceptable image quality. In other words, to inject the noise with the same energy, a better JND model will put more noise into the regions with higher visual redundancy while less noise into the regions with lower perceptual redundancy for achieving the better perceptual quality. According to various pixel-wise JND models, we add the JND noise to each pixel of the test images I for measuring its performance as suggested in [21]:

$$\hat{I}(x, y) = I(x, y) + \eta \cdot \xi \cdot JND(x, y) \tag{19}$$

where (x, y) means the spatial coordinate of the pixel in image, $\hat{I}(x, y)$ denotes the contaminated image by injecting the JND guided noise, the parameter $\eta \in \{-1, +1\}$ is randomly decided to avoid the occurrence of noise change in fixed pattern, and ξ is used to adjust the JND noise injection energy to ensure the noised images contaminated by different JND models at the same level of noise energy.

The contaminated test images resulted from various JND models are compared with the original test images in terms of PSNR and through the subjective quality assessment to evaluate the performances of various JND models. Note that with the same perceived quality (measured by SSIM), the higher the injected-JND-noise energy (measured by PSNR) is, the more reliable the JND model is.

B. PERFORMANCE COMPARISON

1) OBJECTIVE QUALITY COMPARISON

Table 1 shows the objective quality comparison of the proposed JND model and two existing pixel-wise JND models [15], [21] in terms of PSNR. It can be easily seen that the proposed JND model is able to, on average, achieve the lowest PSNR and also the lowest PSNR for all the test images. Compared with Liu *et al.* [15] and Wu *et al.* [21], the additional redundancy yielded by the proposed JND model is 0.86 dB and 1.19 dB, respectively. This study shows the superiority of the proposed JND model, which can tolerate more distortions and exploit the visual redundancy more accurately.

TABLE 1. Objective quality comparison of the proposed model and two pixel-wise JND models in terms of PSNR (dB).

Test Images	Proposed	Liu [15]	Wu [21]
I1	33.29	34.35	34.64
I2	32.32	33.00	33.95
I3	34.92	35.77	36.18
I4	29.70	30.33	30.73
I5	33.96	34.55	35.26
I6	33.68	34.88	35.16
I7	33.97	35.04	34.25
I8	34.25	34.92	34.93
I9	36.15	36.37	37.04
I10	32.73	34.69	34.93
I11	33.83	34.56	34.64
I12	34.65	35.32	36.06
Average	**33.62**	34.48	34.81

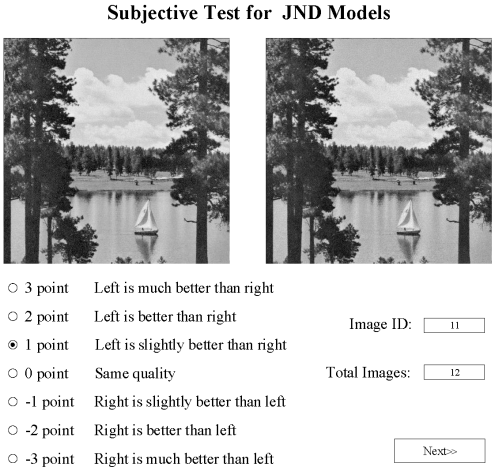


FIGURE 7. A screenshot of the user interface for conducting subjective evaluation.

2) SUBJECTIVE QUALITY COMPARISON

In addition, subjective quality comparison is also performed to demonstrate the effectiveness of our proposed JND model. The subjective quality assessment tests are conducted by exploring the adjectival categorical judgment method and strictly following the ITU-R BT.500-11 standard [37]. The evaluation platform is the desktop PC, which is equipped with a 23-inch LED monitor (with a resolution of 1920×1080), 8 GB RAM, and 64-bit Windows operating system. The evaluation process is conducted indoors, under a normal lighting condition. In each test, two contaminated images by two JND models under comparison presented to the assessor will be judged as one of seven opinion levels, as shown in Fig. 7. These two contaminated images will include the one contaminated by the proposed JND model and the one contaminated by other JND models under comparison, and they will be randomly posed as the left or the right images at the same time. Seven discrete scales from −3 to +3 will

TABLE 2. Subjective quality comparison of the proposed model and two pixel-wise JND models.

Test Images	Proposed vs. Liu [15]		Proposed vs. Wu [21]	
	m	SD	m	SD
I1	-0.625	1.367	-0.750	0.968
I2	-0.688	1.102	0	1.414
I3	-0.125	1.166	0.562	0.998
I4	0.063	0.899	0	0.866
I5	-0.750	1.199	0.188	1.074
I6	0.375	0.927	0.063	0.899
I7	-0.063	0.556	-0.313	0.464
I8	-0.438	0.609	0	0.707
I9	-0.250	1.199	-0.125	0.857
I10	-0.250	1.031	0.125	1.218
I11	0.063	0.747	-0.250	0.829
I12	-0.250	0.750	-0.625	0.992
Average	**-0.245**	0.963	**-0.094**	0.941

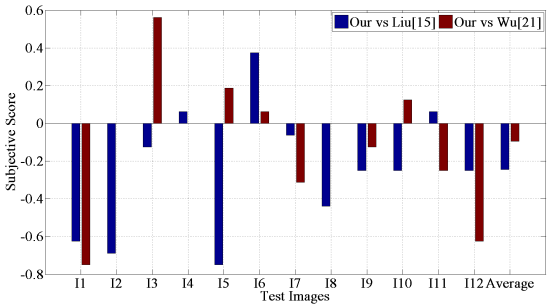


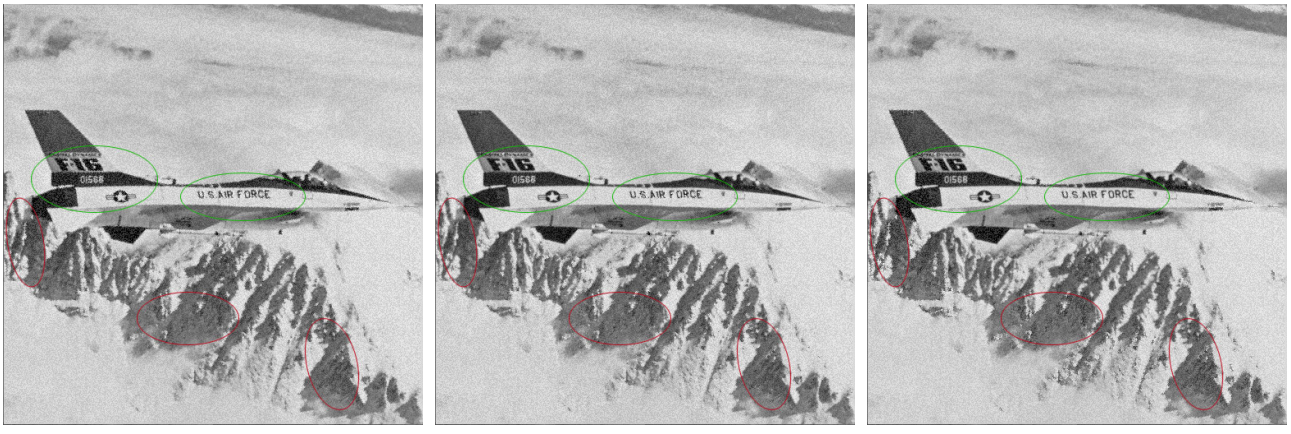
FIGURE 8. The bar graph of the subjective scores of 12 test images and their average scores.

be used to reflect the degree of difference of the subjective quality between the left and right images according to their corresponding definitions in Fig. 7. Twenty assessors were invited to evaluate the subjective quality of all the image pairs, and the assessor is allowed to response after observing the images at least 4 seconds [36].

Table 2 shows the subjective quality comparison of the proposed JND model and two existing pixel-wise JND models [15], [21], where m denotes the mean value of the subjective scores and SD means their standard deviation. Moreover, the bar graph is also shown in Fig. 8 to have a clearer illustration. Note that a larger negative (or positive) subjective score demonstrates that the image processed by our proposed JND model has much better (or worse) perceived quality than that processed by other JND models under comparison. Firstly, we can see from Table 2 that the standard deviations of the subjective scores are quite small (i.e., nearly 1), showing that the subjective evaluation results from twenty assessors are stable and reliable. Then, as shown in Table 2 and Fig. 8, the average mean values of two groups of comparison tests are all negative, i.e., -0.245 and -0.094, respectively, meaning that the contaminated images resulted from the proposed method have overall better subjective quality than that of other JND models [15], [21]. In other words, the proposed JND model consistently outperforms other JND models [15], [21].

Moreover, we further take I9 as an example to show the corresponding contaminated images resulted from different JND models, as displayed in Fig. 9. Note that the same noise energy is injected into the original I9 with different kinds of JND noise. It can be observed that the proposed model achieves better subjective quality.

For the areas encircled with green ellipse to which human eyes are sensitive, the visual effect in Fig. 9 (c) is obviously better than Fig. 9a and Fig. 9b. Tracing it to its cause, for the image in Fig. 9a dealt with Liu *et al.*'s model, although it maintains fairly good edge information, the function of the texture regions to tolerate much distortion is highlighted, actually only the unpredicted texture regions can hide much noise. While for the image in Fig. 9b processed by Wu *et al.*'s model, it emphasizes the masking effect of disorderly regions according to free energy principle. However, for the areas circled by green ellipse, the visual redundancy of the numbers, which are relatively sensitive to HVS and given more attention, are overvalued. And for the regions encircled with red ellipses which are dark and insensitive to human eyes, there seems similar among them. Thus it can be seen that our JND model is superior to Liu's and Wu's models.



(a) (b) (c)

FIGURE 9. The subjective comparison of JND noise-injected images resulted from different JND models (Taking I9 as an example): (a) Liu [15]; (b) Wu [21]; and (c) Proposed.

IV. CONCLUSION

In this work, a novel pixel-wise JND model was proposed based on elaborate image decomposition and the saliency. By means of RTV model and orientation complexity, a real image is split into three portions, namely, structural image, orderly textural image and disorderly textural image for EM, OTM and DTM estimation, respectively. Considering visual attention of HVS, we proposed CM_s for contrast masking estimation combining based on the saliency. From the results of PSNR comparison test and subjective quality comparison, our proposed model is better than the related existing JND models. Furthermore, with the advantage of our model, it will have effective improvement in video coding, image quality evaluation, image watermarking and so on.

REFERENCES

[1] 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 Trans. Image Process.*, vol. 26, no. 6, pp. 2682–2693, Jun. 2017.
[2] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838–3851, May 2016.
[3] L. Yu, H. Su, and C. Jung, "Perceptually optimized enhancement of contrast and color in images," *IEEE Access*, vol. 6, pp. 36132–36142, 2018.
[4] A. Yang, H. Zeng, J. Chen, J. Zhu, and C. Cai, "Perceptual feature guided rate distortion optimization for high efficiency video coding," *Multidimensional Syst. Signal Process.*, vol. 28, no. 4, pp. 1249–1266, 2017.
[5] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1422, Oct. 1993.
[6] W. Wan, J. Wu, X. Xie, and G. Shi, "A novel just noticeable difference model via orientation regularity in DCT domain," *IEEE Access*, vol. 5, pp. 22953–22964, Nov. 2017.
[7] J. Kim, S. H. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1786–1800, Nov. 2015.
[8] Z. Zeng, H. Zeng, J. Chen, J. Hou, and K. K. Ma, "A novel direction-based jnd model for perceptual hevc intra coding," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2017, pp. 186–190.
[9] X. F. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao, "Just-noticeable difference-based perceptual optimization for JPEG compression," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 96–100, Jan. 2017.
[10] Y. Fan, M. Larabi, F. Cheikh, and C. Fernandez-Maloigne, "A survey of stereoscopic 3D just noticeable difference models," *IEEE Access*, vol. 7, pp. 8621–8645, 2019.
[11] G. Nader, K. Wang, F. Hétroy-Wheeler, and F. Dupont, "Just noticeable distortion profile for flat-shaded 3D mesh surfaces," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 11, pp. 2423–2436, Dec. 2015.
[12] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, "Adaptive watermarking and tree structure based image quality estimation," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 311–325, Feb. 2014.
[13] H. Hadizadeh, "Energy-efficient images," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2882–2891, Apr. 2017.
[14] X. H. Zhang, W. S. Lin, and P. Xue, "Just-noticeable distortion estimation for image pixels," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct. 2005, pp. 1–4.
[15] 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 Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.
[16] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337–346, Mar. 2009.
[17] S.-H. Bae and M. Kim, "A novel DCT-based JND model for luminance adaptation effect in DCT frequency," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 893–896, Sep. 2013.
[18] S.-H. Bae and M. Kim, "A DCT-based total JND profile for spatiotemporal and foveated masking effects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1196–1207, Jun. 2017.
[19] 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, Dec. 1995.
[20] 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 Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.
[21] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1705–1710, Nov. 2013.
[22] H. Hadizadeh, A. Rajati, and I. V. Bajić, "Saliency-guided just noticeable distortion estimation using the normalized laplacian pyramid," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1218–1222, Aug. 2017.
[23] N. B. Turk-Browne, J. A. Jungé, and B. J. Scholl, "The automaticity of visual statistical learning," *J. Experim. Psychol. Gen.*, vol. 134, no. 4, pp. 552–564, 2005.
[24] J. R. Saffran and E. D. Thiessen, "Pattern induction by infant language learner," *Develop. Psychol.*, vol. 39, no. 3, pp. 484–494, 2003.
[25] T. D. Albright, "Direction and orientation selectivity of neurons in visual area MT of the macaque," *J. Neurophysiol.*, vol. 52, no. 6, pp. 1106–1130, 1985.
[26] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.
[27] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.
[28] O. Le Meur, P. Le Callet, and D. Barba, "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817, May 2006.
[29] J. Wu, F. Qi, G. Shi, and Y. Lu, "Non-local spatial redundancy reduction for bottom-up saliency estimation," *J. Vis. Commun. Image Represent.*, vol. 23, no. 7, pp. 1158–1166, Oct. 2012.
[30] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.
[31] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, p. 139, 2012.
[32] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
[33] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *J. Physiol.*, vol. 187, no. 2, pp. 437–445, 1966.
[34] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.
[35] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Process.*, vol. 70, no. 3, pp. 177–200, Nov. 1998.
[36] M. Uzair and R. D. Dony, "Estimating just-noticeable distortion for images/videos in pixel domain," *IET Image Process.*, vol. 11, no. 8, pp. 559–567, Aug. 2017.
[37] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-11, ITU, Geneva, Switzerland, 2002.

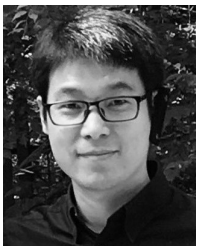
ZHIPENG ZENG received the B.S. degree in communication engineering and the M.S. degree in information and communication engineering from the School of Information Science and Engineering, Huaqiao University, Xiamen, China, in 2016 and 2019, respectively. He is currently a Visiting Student with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. His current research interests include image processing and video coding.

HUANQIANG ZENG (S'10–M'13–SM'18) received the B.S. and M.S. degrees from Huaqiao University, Xiamen, China, and the Ph.D. degree from Nanyang Technological University, Singapore, all in electrical engineering. He was a Postdoctoral Fellow with The Chinese University of Hong Kong, Hong Kong. He is currently a Full Professor with the School of Information Science and Engineering, Huaqiao University, Xiamen, China. He has published more than 90 articles in well-known journals and conferences, including IEEE TIP, TCSVT, and TIT. His current research interests include image processing, video coding, machine learning, and computer vision.

Dr. Zeng has also been actively serving as the Technical Program Committee Member of multiple flagship international conferences. He received three best poster/paper awards (in International Forum of Digital TV and Multimedia Communication, in 2018, and in Chinese Conference on Signal Processing 2017/2019). He has also been actively serving as the General Co-Chair of the IEEE International Symposium on Intelligent Signal Processing and Communication Systems 2017 (ISPACS2017), the Technical Program Co-Chair of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2017 (APSIPA-ASC2017), the Area Chair of the IEEE International Conference on Visual Communications and Image Processing (VCIP2015), and the Reviewer of numerous international journals and conferences. He has been actively serving as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE ACCESS, and *IET Electronics Letters* and a Guest Editor of the *Journal of Visual Communication and Image Representation*, *Multimedia Tools and Applications*, and the *Journal of Ambient Intelligence and Humanized Computing*.

JING CHEN (M'17) received the B.S. and M.S. degrees in computer science from Huaqiao University, Xiamen, China, and the Ph.D. degree in control engineering from Xiamen University, Xiamen. She was a Visiting Scholar with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, from 2011 to 2012. She is currently an Associate Professor with the School of Information Science and Engineering, Huaqiao University. Her current research interests include image processing and video coding. She is a Traffic Video Technical Committee Member of the China Society of Image and Graphics. She received the Best Paper Award from the Chinese Conference on Signal Processing, in 2017. She was the Financial Co-Chair and the Session Chair of the IEEE International Symposium on Intelligent Signal Processing and Communication Systems, in 2017. She has been serving as a Reviewer of the *Journal of Visual Communication and Image Representation* and *Multimedia Tools and Applications*.

JIANQING ZHU received the B.S. degree in communication engineering and the M.S. degree in communication and information system from the School of Information Science and Engineering, Huaqiao University, Xiamen, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2015. He is currently an Associate Professor with the College of Engineering, Huaqiao University, Quanzhou, China. His current research interests include computer vision and pattern recognition, with a focus on image and video analysis, particularly, person re-identification, object detection, and video surveillance. He was a recipient of the Best Biometrics Student Paper award from the International Conference on Biometrics, in 2015.



YUN ZHANG (M'12–SM'16) received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2014, he was a Postdoctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong. From 2010 to 2017, he was an Assistant Professor and an Associate Professor with the Shenzhen Institutes of Advanced Technology (SIAT), CAS, Shenzhen, China, where he is currently a Professor. His current research interests include video compression, 3D video processing, and visual perception.

KAI-KUANG MA (S'80–M'84–SM'95–F'13) received the B.E. degree in electronic engineering from Chung Yuan Christian University, Chungli, Taiwan, the M.S. degree in electrical engineering from Duke University, Durham, NC, USA, and the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC, USA.

From 1992 to 1995, he was a Member of the Technical Staff with the Institute of Microelectronics (IME), Singapore, involved in digital video coding and the MPEG standards. From 1984 to 1992, he was with IBM Corporation at Kingston, NY, USA, and Research Triangle Park, Durham, NC, USA, involved in various DSP and VLSI advanced product development. He is currently a Full Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He has published extensively and holds one USA patent on fast motion estimation algorithm. His current research interests include digital image/video processing and computer vision, including digital image/video coding and standards, image/video segmentation, denoising and enhancement, and interpolation and super-resolution. His current research interests on computer vision include image matching and registration, scene analysis and recognition, and human-computer interaction.

Dr. Ma has been serving as an Editorial Board Member of several leading international journals in his research area, such as a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2016 to 2019, an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, since 2015, the IEEE SIGNAL PROCESSING LETTERS, from 2014 to 2016, and the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2007 to 2010, an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, from 1997 to 2012, the IEEE TRANSACTIONS ON MULTIMEDIA, from 2002 to 2009, the *International Journal of Image and Graphics*, from 2003 to 2015, and the *Journal of Visual Communication and Image Representation*, from 2005 to 2015. He is an elected member of three IEEE Technical Committees, including the Image and Multidimensional Signal Processing (IMDSP) Committee, the Multimedia Communications Committee, and the Digital Signal Processing Committee. He has been serving as a Technical Program Committee member, a reviewer, and a Session Chair of multiple IEEE international conferences. He is a member of Sigma Xi and Eta Kappa Nu. He served as a Singapore MPEG Chairman and the Head of Delegation, from 1997 to 2001. On the MPEG contributions, two fast motion estimation algorithms (Diamond Search and MVFAST) produced from his research group have been adopted by the MPEG-4 standard, as the reference core technology for fast motion estimation. He was the General Chair of organizing a series of international standard meetings (MPEG and JPEG) and JPEG2000 and MPEG-7 workshops held in Singapore, in 2001. He was the Chairman of the IEEE Signal Processing Singapore Chapter, from 2000 to 2002. He is a General Co-Chair of ISPACS2017, ASIPA2017, ACCV2016 Workshop, VCIP-2013; a Technical Program Co-Chair of ICIP-2004, ISPACS-2007, IIH-MSP-2009, and PSIVT-2010; and an Area Chair of ACCV-2009 and ACCV-2010. He was elected as a Distinguished Lecturer of the IEEE Circuits and Systems Society, from 2008 to 2009.