Just Noticeable Difference Estimation for Screen Content Images

Shiqi Wang, *Member, IEEE*, Lin Ma, *Member, IEEE*, Yuming Fang, Weisi Lin, *Fellow, IEEE*, Siwei Ma, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— We propose a novel just noticeable difference (JND) model for a screen content image (SCI). The distinct properties of the SCI result in different behaviors of the human visual system when viewing the textual content, which motivate us to employ a local parametric edge model with an adaptive representation of the edge profile in JND modeling. In particular, we decompose each edge profile into its luminance, contrast, and structure, and then evaluate the visibility threshold in different ways. The edge luminance adaptation, contrast masking, and structural distortion sensitivity are studied in subjective experiments, and the final JND model is established based on the edge profile reconstruction with tolerable variations. Extensive experiments are conducted to verify the proposed JND model, which confirm that it is accurate in predicting the JND profile, and outperforms the state-of-the-art schemes in terms of the distortion masking ability. Furthermore, we explore the applicability of the proposed JND model in the scenario of perceptually lossless SCI compression, and experimental results show that the proposed scheme can outperform the conventional JND guided compression schemes by providing better visual quality at the same coding bits.**

***Index Terms*— Just noticeable difference, screen content image, parametric edge modeling.**

I. Introduction

RECENTLY, we have witnessed the rapid development services of cloud computing, which enables the thin-clients such as laptop, tablet, PDA, smart phone, etc to provide an even further computing experiences by remotely accessing the computationally intensive and graphically rich resources from cloud. To achieve this, the remote screen is usually compressed and transmitted to the local terminals to facilitate the remote computing process. Typical applications include remote computing platforms [1], [2], mobile browsers [3] and remote desktop sharing systems [4]. In these scenarios, the time variant interface can be regarded as a kind of screen content image (SCI), which is a combination of pictorial regions and computer generated textual content [5]. Therefore, the quality of SCI can largely influence the user experience of the remote systems.

The SCIs exhibit different statistical properties compared with natural images. Typically, the SCIs are rendered with sharp edges and thin lines [6], while the natural images are usually featured by continuous-tone content with smooth edges and thick lines. Moreover, the capturing of natural images will introduce noise due to the limitations of the image sensors, while the SCIs are usually noise free as they are purely generated by computer. In view of these distinct features of SCIs, it is meaningful to further investigate the perceptual characteristics, which are critical in guiding and optimizing the screen content compression and processing systems.

The near-threshold properties of the human visual system (HVS) indicate that the HVS cannot sense any pixel variations. The just noticeable difference (JND) characterizes the minimum visibility threshold below which the pixel level variations cannot be perceived by the HVS. Appropriate JND models can be efficiently applied on improving the performance of image/video compression and processing systems [7]–[12]. Generally, JND models can be estimated in pixel domain or transform domain [13]. Pixel domain JND models are usually formulated with the consideration of luminance adaptation (refers to the masking effect of the HVS toward background luminance) and contrast masking (refers to the visibility reduction of one visual signal at the presence of another). In [8] and [10], the overlapping effect of luminance adaptation and spatial contrast masking was investigated to derive the JND model. In [14], the edge masking and texture masking were further distinguished due to the entropy masking property. In [15], the disorderly concealment effect was introduced to model JND. Transform domain JND models take advantage of human vision sensitivities of different frequency components, and the contrast sensitivity function (CSF) is frequently applied in identifying the base thresholds. In [16], the discrete cosine transform (DCT) based DCTune JND model was proposed with the consideration of

Manuscript received June 29, 2015; revised December 26, 2015 and May 6, 2016; accepted May 9, 2016. Date of publication May 26, 2016; date of current version June 23, 2016. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800 and in part by the National Natural Science Foundation of China under Grant 61322106 and Grant 61571017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter Tay. *(Corresponding author: Lin Ma.)*

S. Wang is with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore 639798 (e-mail: wangshiqi@ntu.edu.sg).

L. Ma is with Huawei Noah's Ark Lab, Hong Kong (e-mail: forest.linma@gmail.com).

Y. Fang is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: fa0001ng@e.ntu.edu.sg).

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

S. Ma and W. Gao are with the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing 100871, China (e-mail: swma@pku.edu.cn; wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2573597

the masking effect. Höntsch and Karam [17] further modified the DCTune model by replacing a single pixel with a foveal region. Recently, a spatial-temporal JND model for grey scale image/video in DCT domain was introduced in [18], which incorporated the luminance adaptation, contrast masking and Gamma correction to estimate JND. In [19], adaptive size transform based JND model was proposed, which extends the 8×8 DCT to 16×16 DCT by including both the spatial and temporal HVS properties.

To adapt the applicability of JND in various scenarios, the conventional JND models were extended to just noticeable color difference (JNCD) [20] and foveated JND (FJND) [12] to exploit more masking effects. Moreover, the concept of just noticeable blur (JNB) was introduced in [21]. It indicated that HVS is able to mask blurriness around an edge up to a certain threshold. In [22], JNB was further extended to deduce a no reference blur metric with cumulative probability.

To our best knowledge, none of the existing JND models are developed for SCIs. The distinct properties of SCIs such as thin edges, limited color representation indicate that specifically designed JND model for SCIs is highly desirable. Moreover, in [23] it is observed that the extent of the visual field used to extract information in SCIs is much smaller than that of natural images. In view of these distinct features of SCIs, in [24] and [25] a database with subjective quality ranking of distorted SCIs was created. It demonstrates that the state-of-the-art perceptual models are still quite limited in predicting the visual quality of the SCIs, and inspires a series of specifically developed full-reference SCI quality assessment measures recently [26], [27].

In this work, we develop a novel pixel domain JND model for SCIs. A parametric edge model [28] is employed, in which we represent any edge profile in a unique and adaptive way by three conceptually independent components: edge luminance, edge contrast and edge structure. Such edge representation strategy is adapted to the input signal. Subsequently, the visibility thresholds for these components are studied in different ways. In particular, the sensitivity of luminance is derived based on luminance adaptation, the contrast masking effect is accounted for by the divisive normalization philosophy, and the tolerable structural distortion is approximated based on the allowable edge width variation. The final JND model of textual content is generated by employing the combination of these masking effects to reconstruct the edge profile. Extensive experimental results exhibit that the proposed model outperforms the state-of-the-art schemes in terms of the distortion masking ability.

Furthermore, inspired by the recent developments of perceptual image and video coding [13], [29]–[31], we incorporate the proposed JND model in perceptually lossless SCI compression. In particular, this is achieved in the framework of high efficiency video coding (HEVC) screen content coding extension [32], [33], and the subjective results provide the meaningful evidence that the proposed JND model is efficient in reducing the coding bits while maintaining the just noticeable difference level of visual quality.

The remainder of this paper is organized as follows. In Section II, we analyze the characteristics of SCIs. Section III details the parametric edge modeling and the visibility thresholds derived in terms of luminance, contrast and structure. In Section IV, we propose the JND model for SCI. Section V shows the experimental results by comparing the proposed JND model with conventional methods. In Section VI, we further explore its application on SCI compression. Finally, the paper is concluded in Section VII.

II. Analyses on Characteristics of SCIs

Rather than providing a naturalness looking, the functionality of computer generated SCIs is to convey semantic information with high contrast edges. In view of this, we provide comprehensive analyses on the unique characteristics of SCIs, including the frequency energy falloff statistics, sharpness of edges and free-energy principle based image uncertainty.

A. Frequency Energy Falloff Statistics

The natural scene statistics (NSS) is based on the hypothesis that the HVS is highly adapted to the statistics of the natural visual environment, such that the departure from such statistics characterizes image unnaturalness. One typical phenomenon in NSS is that the amplitude spectrum of natural images falls with the spatial frequency approximately proportional to $1/f^p$ law [34], where f is the spatial frequency and p is an image dependent constant. As SCIs do not have the property of "naturalness", we examine this property by decomposing the images with Fourier transform and computing the frequency energy. For typical natural image, pure textual image and compound image composed of both textual and natural content, the frequency energy falloff statistics are demonstrated in Fig. 1. We can observe that the falloff statistics for natural image approximately accord with the $1/f^p$ relationship due to the straight line in log-log scale. However, for SCIs with pure textual or compound content there are peaks at high frequency.

B. Sharpness of Edges

In contrast with natural images, SCIs are featured with thin edges. To illustrate this property, the JNB based blur metric [21] which takes the advantage of edge width is computed for both SCIs and natural images. Specifically, the blur metric is defined as follows,

$$S = \frac{L}{D_s}, \tag{1}$$

where L denotes the total number of processed blocks in the image and D_s denotes the sharpness measure. For the same number of processed blocks, a smaller D_s value corresponds to a sharper image. In particular, it is represented by,

$$D_s = \left(\sum_{R_b} \left| D_{R_b} \right|^{\beta} \right)^{\frac{1}{\beta}}, \tag{2}$$

where R_b is the edge block and D_{R_b} is computed by the width of edge e_i (denoted as $\Omega(e_i)$) and the just noticeable

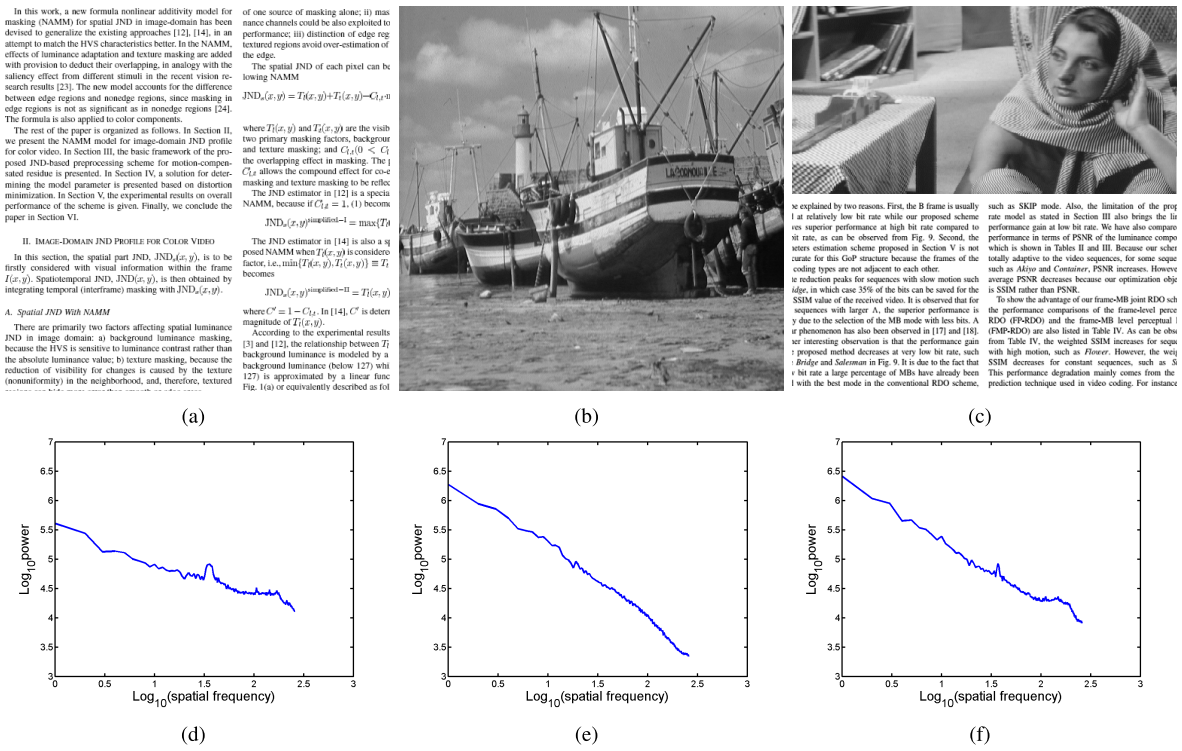


Fig. 1. Frequency energy falloffs of the pure textual image, natural image and compound image that is composed of both textual and natural content. (a) Pure textual image; (b) Natural image; (c) Compound image; (d) Frequency energy falloff of pure textual image; (e) Frequency energy falloff of the natural image; (f) Frequency energy falloff of the compound image.

width $\Omega_{JNB}(e_i)$ in the current block,

$$D_{R_b} = \left(\sum_{e_i \in R_b} \left| \frac{\Omega(e_i)}{\Omega_{JNB}(e_i)} \right|^{\beta} \right)^{\frac{1}{\beta}}. \quad (3)$$

The edge width is measured by the number of pixels with increasing grayscale values in one direction while decreasing grayscale values in the other direction. The Ω_{JNB} is derived from subjective tests and is measured to be 5 for low contrast and 3 for high contrast edges, respectively. The β value is set to be 3.6 in this test.

The design philosophy of the sharpness metric is employing the width of the edge to represent the probability of blur detection. In [21], the blur detection probability is modeled in the form of the exponential function [35],

$$P(e_i) = 1 - \exp\left(- \left| \frac{\Omega(e_i)}{\Omega_{JNB}(e_i)} \right|^{\beta} \right). \quad (4)$$

This implies that the probability of the detected blur is reflected by the width of the edge to the just noticeable edge width Ω_{JNB}. Specifically, when $\Omega(e_i) = \Omega_{JNB}(e_i)$, the probability of blur detection equals to 63%. In this manner, a lower $\Omega(e_i)$ indicates a larger value of S, corresponding to a sharper input image.

The blur metric scores for SCIs and natural images are shown in Table I. The SCIs are obtained from the database [25]. It is observed that the SCIs exhibit sharper appearance because of larger values of S, which verifies that the SCIs usually contain more thinner edges compared with natural images.

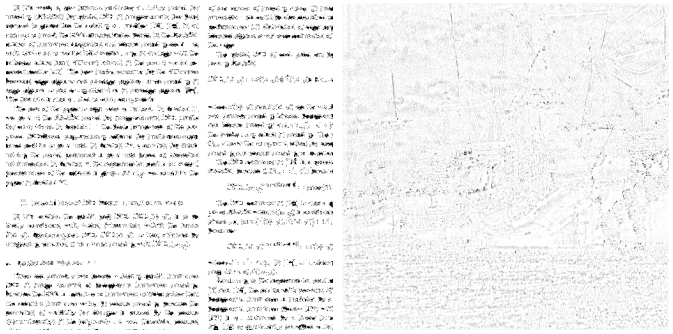
TABLE I
SHARPNESS EVALUATIONS OF NATURAL IMAGES AND SCIS

Natural Images	S	SCIs	S
airplane	4.06	cim1	7.97
baboon	4.85	cim2	7.97
barbara	3.45	cim3	7.98
boat	4.92	cim4	8.96

C. Investigation of "Surprise" Based on Free-Energy Principle

The basic premise of the free-energy principle based brain theory [36] is that the cognitive process is manipulated by an internal generative model (IGM) that can actively infer the orderly information of input visual signals and avoid the disorderly content. The free-energy characterizes the upper bound of the "surprise" for the image data, or equivalently, measures the discrepancy between the image data and the best explanation by the brain generative model [37]. In this subsection, we examine the free-energy of SCI to further investigate how human brain perceives the screen visuals.

To compute the free-energy, statistical modeling of the free-energy should be established. Here we apply the linear



(a) (b)

Fig. 2. Visualization of the residual map between the original and the AR predicted images (residuals are enlarged for better visualization, light/dark regions represent low/high residual values, respectively). (a) Residual map of SCI in Fig. 1 (a). (b) Residual map of the natural image in Fig. 1 (b).

autoregressive (AR) model as follows,

$$x'_n = \chi^k(x_n)a + e_n, \tag{5}$$

where x'_n is the predicted pixel value of x_n and $\chi^k(x_n)$ is the vector of pixels consisting of k neighborhoods. Parameter e_n is the zero mean Gaussian noise. The parameters of the AR model are estimated by [15],

$$a_i = \frac{H_m(x_n; x_i)}{\sum_j H_m(x_n; x_j)}, \tag{6}$$

where $H_m(x_n; x_j)$ denotes the mutual information between x_n and x_j.

In [37], it is shown that the free-energy can be characterized by the the total description length of the image by adding the entropy of the prediction residuals with the model cost. As the model cost can be regarded as a constant over the entire image, here only the local prediction residuals are evaluated. The residual maps that represent the local free-energy of SCI and natural image in Fig. 1 (a)&(b) are shown in Fig. 2, which illustrate that the free-energy exhibits much stronger in textual content. This further demonstrates that the textual content is more informative and contains more "surprise".

According to the free-energy principle based brain theory, to maintain a low free energy, the system can automatically change the way that the environment is sampled to minimize it. As the acquired information or "surprise" is directly determined by how large the visual field is, and to lower the uncertainties in the representation of the natural scene, the HVS therefore tries to adaptively select smaller visual field when perceiving the textual content [23]. Moreover, the above analyses indicate that the SCIs contain less homogeneous content and the JND thresholds for neighbouring pixels may vary significantly. These distinct properties imply that the traditional patch based JND models that predict the masking amplitudes relying on the background luminance/activities may not be appropriate for SCIs. Instead, we model the textual content in SCIs with the local edge profile. This gives rise to the motivation of the proposed JND estimation method that adopts the parametric model to depict the edge in terms of three conceptually independent components: luminance, contrast and structure.

III. Visibility Threshold in Parametric Edge Modeling

It is generally believed that the HVS understands an image mainly based on its low-level features, such as edges and zero-crossings [38], [39]. In particular, the edges in an SCI carry important information for human perception. To explore the edge construction, here we adopt a parametric model [28], [40], [41] to depict edges in SCI. In this section, we will detail this model and derive the visibility thresholds of each modeling portions.

A. Parametric Edge Modeling

We adopt the one-dimensional notation to describe the edge model. In general, a step edge x_0 can be characterized by a unit step function with the edge basis as

$$u(x; b, c, x_0) = c \cdot U(x - x_0) + b, \tag{7}$$

where $U(\cdot)$ denotes the unit step function, b denotes the edge basis and c represents the edge contrast. However, such infinitely sharp edges or step edges do not exist in practical images. Even for the textual blocks in a typical SCI, the number of major colors is usually more than three [42]. This implies that there are usually smooth transitions in the edge construction. Therefore, following [28] and [41], a typical edge is treated as the distorted version of the isolated ideal step edge with a point spread function,

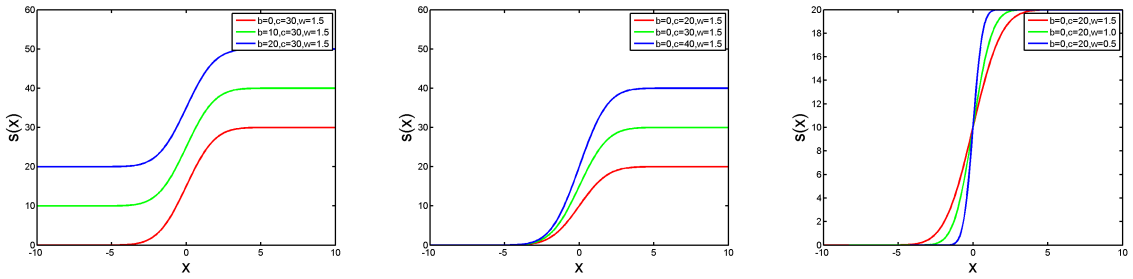
$$s(x; b, c, w, x_0) = u(x; b, c, x_0) * psf(x; w). \tag{8}$$

The Gaussian function which is usually used as an approximation to the point spread function of an optical system is employed to model $psf(x; w)$. Therefore, the real edge in SCI can be regarded as a smooth version of the unit step edge, which can be characterized by convolving the step edge $u(x; b, c, x_0)$ with Gaussian filter $g(x; w) = (1/\sqrt{(2\pi w^2)}) \exp((-x^2/2w^2))$,

$$s(x; b, c, w, x_0) = b + \frac{c}{2}\left(1 + \mathrm{erf}\left(\frac{x - x_0}{w\sqrt{2}}\right)\right). \tag{9}$$

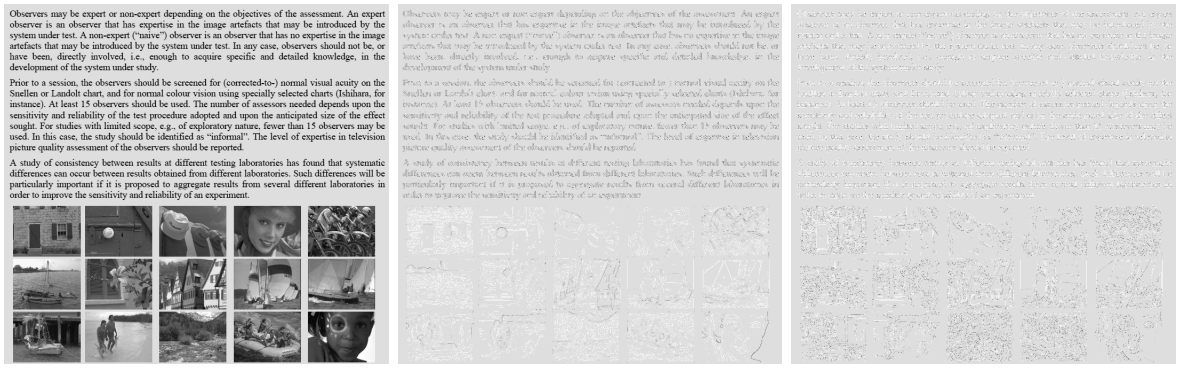
Here erf$(\cdot)$ denotes the error function and w is the standard deviation of the Gaussian function $g(x; w)$ that controls the edge structure. As the edge structure is actually determined by the edge width, the parameter w can also be regarded as the parameter that reflects the edge width. Such a representation is adaptive, which allows us to decompose any edge profile into luminance, contrast and structure controlled by parameters b, c and w, respectively. As shown in Fig. 3 (a), the parameter b determines the base intensity of an edge. In Fig. 3 (b), one can discern that the parameter c reflects the strength of the edge, such that higher c corresponds to a stronger edge. Fig. 3 (c) exhibits that the edge structure is controlled by w, which corresponds to a specific shape, and the edge profile will become sharper as w becomes smaller.

These parameters are derived by fitting the function (9) with the local pixel information. To this end, edge detection should be firstly performed. The edge detection shares the similar procedures as Canny edge detection [43]. In particular, it consists of Gaussian smoothing and differentiation, which



(a) (b) (c)

Fig. 3. Illustration of one-dimensional edge model with different settings of b, c and w.



(a) (b) (c)

Fig. 4. Illustration of edge contrast and width map (unified background color for better visualization). (a) SCI. (b) Edge contrast map. (c) Edge width map.

filters the signal by convolving the $s(x; b, c, w, x_0)$ with the derivative of the Gaussian filter, leading to

$$d(x; c, w, \sigma_d) = \frac{c}{\sqrt{2\pi\,(w^2 + d^2)}} \exp\left(\frac{-(x - x_0)^2}{2(w^2 + \sigma_d^2)}\right). \quad (10)$$

After filtering the edge with the derivative of Gaussian, it can be detected by finding the local maxima in the filtered output. Subsequently, these parameters are estimated by sampling the response at three locations $x = (0, a, -a)$. Let $d_1 = d(0; c, w, \sigma_d)$, $d_2 = d(a; c, w, \sigma_d)$ and $d_3 = d(-a; c, w, \sigma_d)$, we have

$$\begin{aligned} w &= \sqrt{a^2/ln(l_1) - \sigma_d^2} \\ c &= d_1 \cdot \sqrt{2\pi a^2/ln(l_1)} \cdot l_2^{\frac{1}{4a}} \\ b &= s(x_0) - c/2, \end{aligned} \quad (11)$$

where $l_1 = d_1^2/d_2 d_3$ and $l_2 = d_2/d_3$. Generally, small sampling distance a may not well reflect the whole edge structure, and large sampling distance may exceed the range of the edge profile. Following the work in [41], here we set $a = 1$.

In Fig. 4, we demonstrate the edge contrast and width map of a typical SCI, and one can discern that the edges in textual content have higher contrast but thinner width than those in pictorial content. This is in line with our observation in Section II. The parametric model further supports the reconstruction of edge luminance, contrast and structure by modifying b, c and w. This enables us to analyze and derive the visibility thresholds from these three aspects, as demonstrated in Fig. 5. As such, the JND of the pixels along the edge profile can be derived based on the edge profile reconstruction with the tolerable luminance, contrast and structure variations.

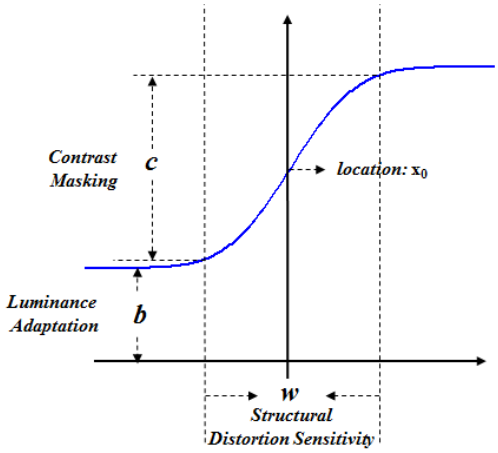


Fig. 5. Luminance adaptation, contrast masking and structural distortion sensitivity in an edge profile.

B. Luminance Adaptation

In the literature, conventional luminance adaptation models usually rely on the background luminance from an $N \times N$ patch to derive the visibility sensitivity. In [7] and [44], it is observed that the approximated slope of the line that relates

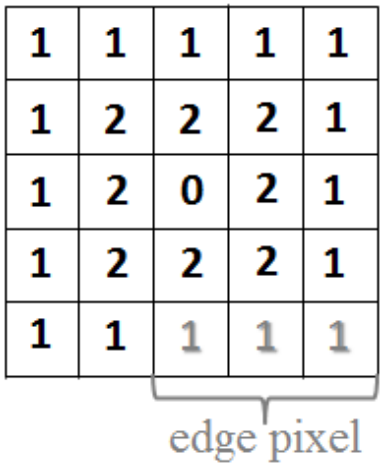


Fig. 6. The weighted low-pass filter for calculating average background luminance (the edge pixels are excluded).

the visibility threshold tends to increase slightly as the background luminance increases, which is qualitatively consistent with Weber's law. The reason for adopting such methods to compute the background luminance is that the natural images are usually composed of regions with homogeneous texture. However, this property does not always hold for the textual content of SCIs, which motivates us to further investigate its distinct luminance adaptation effect.

Compared with natural images, the pictorial regions in SCIs may usually contain more large uniformly flat areas, while the non-pictorial regions are composed of strong and thin edges. Therefore, it is desirable to separate edge pixels from the pixels in pictorial regions in determining the background luminance. With edge detection in (10), pixels in each SCI can be divided into two sets: edge and pictorial. This allows us to measure the luminance masking in different ways.

Following the one dimensional notation, we assume the position of a pixel is p. For pictorial pixel, the background pixel $\overline{I(p)}$ is computed within a 5×5 window [10], as illustrated in Fig. 6. It is worth noting that the edge pixels are excluded in the calculation of background luminance, as they usually do not belong to the homogeneous regions. In contrast, when p belongs to the edge pixel set, the mean luminance along the edge profile is $\overline{I(p)} = b + c/2$, as illustrated in Fig. 5.

The luminance masking effect in [44] is formulated as follows,

$$T_l(p) = \begin{cases} \alpha_1 \cdot \left(1 - \sqrt{\frac{\overline{I(p)}}{127}}\right) + \beta & \text{if } \overline{I(p)} \leq 127 \\ \alpha_2 \cdot \left(\overline{I(p)} - 127\right) + \beta & \text{otherwise,} \end{cases} \tag{12}$$

where the parameters are set as α_1=17, $\alpha_2 = 3/128$ and $\beta = 3$.

To further investigate the luminance adaptation effect in SCI, a subjective test was conducted to study the relationship between $\overline{I(p)}$ and the visibility threshold that is controlled by three parameters α_1, α_2 and β. Specifically, 20 subjects (12 males, 8 females, aging from 18 to 28) participated in this study. As illustrated in Fig. 7, 10 sample SCIs from database [25] are used in the testing. The screen resolution

Fig. 7. Demonstration of the SCIs used in the subjective tests to obtain the optimal parameters.

is 1920×1080 and the viewing distance is about two times of the screen height. Each image was altered by ten sets of parameters as follows,

$$\begin{aligned} &\psi_1 = \{8, 1, 1/128\}, \quad \psi_2 = \{10, 1, 1/128\} \\ &\psi_3 = \{13, 2, 2/128\}, \quad \psi_4 = \{15, 2, 2/128\} \\ &\psi_5 = \{17, 2, 2/128\}, \quad \psi_6 = \{19, 3, 3/128\} \\ &\psi_7 = \{21, 3, 3/128\}, \quad \psi_8 = \{23, 3, 3/128\} \\ &\psi_9 = \{25, 4, 4/128\}, \quad \psi_{10} = \{27, 4, 4/128\}, \end{aligned} \tag{13}$$

where each set corresponds to the parameter vector $\psi = \{\alpha_1, \beta, \alpha_2\}$. Given the parameter set ψ and the original image I, the luminance masking threshold can be obtained via (12), and finally the pixel p in the altered image from luminance adaptation is generated by,

$$\bar{I}_l(p) = I(p) + r \cdot T_l(p), \tag{14}$$

where r is randomly set as $+1$ or -1.

The subjective test is based on a two-alternative-forced-choice (2AFC) method. This method is widely used in psychophysical studies [45], where in each trial, a subject is shown two images of the same scene (i.e., the original image and its noise-contaminated version) and is asked (forced) to choose the one he/she thinks to have better quality. At each level, the percentage by which the subjects are in favor of original image is recorded, and the parameters corresponding to a probability of 75% in favor of the original SCIs is adopted in the final JND modeling [46]. Finally, as observed in Fig. 8, the parameter set ψ_5 is finally chosen. It is interesting to find that the curve generated by the derived parameters indicates a lower tolerance than the results derived in [7]. This can be explained by the reason that most areas of the SCIs are occupied by sharp, thin edges and noise-free pictorial regions, both of which have lower tolerance on the luminance change.

C. Edge Contrast Masking

It is generally believed that HVS perceives contrast rather than absolute level of light. This trend is more obvious for SCIs as the relative changes in luminance convey important information to HVS. In this work, we aim to derive the visibility threshold on the noticeable edge contrast variation by incorporating the divisive normalization framework, which has shown to be a useful model that accounts for the masking effect in the HVS [47], [48]. In divisive normalization, it is assumed that the contrast change is dependent on the difference of the normalized contrast but not adaptive to the

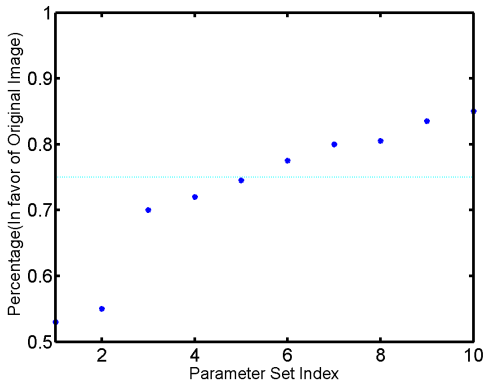


Fig. 8. The relationship between the parameter set index in (13) and the percentage of the preference on original SCIs.

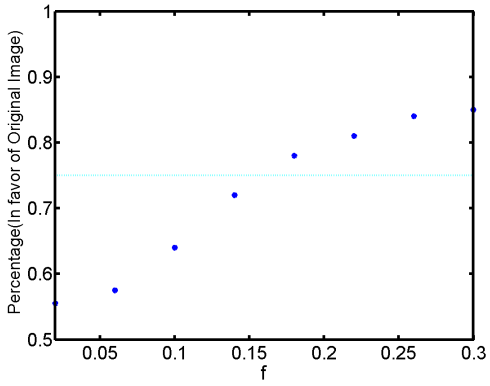


Fig. 9. The relationship between f and the percentage of the preference on original SCIs.

absolute values themselves. Assume $c(p)$ and $c'(p)$ are the original and changed contrast of the edge profile that pixel p belongs to, this process can be represented as,

$$f = \frac{|c(p) - c'(p)|}{c(p) + c'(p)}, \tag{15}$$

where the parameter f indicates the perceived contrast change. This implies that visibility threshold of contrast variation can be gauged by the determination of the threshold on f. To explore this, a similar subjective test as in Section III-B was conducted, where the distortion type is replaced with contrast changed. The contrast changed SCIs are generated by reconstructing the edge profile using the alternated contrast with (9). The testing conditions are the same as those in Section III-B and a set of f values between [0.02 0.3] with an interval 0.04 are examined. The plot between f and the percentage in favor of the original image is demonstrated in Fig. 9, indicating that the visibility threshold $f_{th} = 0.14$ is an appropriate value.

Given f_{th}, the tolerable changed contrast is computed as,

$$\begin{aligned} T_{c+}(p) &= \frac{1 + f_{th}}{1 - f_{th}} \cdot c(p) \\ T_{c-}(p) &= \frac{1 - f_{th}}{1 + f_{th}} \cdot c(p), \end{aligned} \tag{16}$$

where $T_{c+}(p)$ and $T_{c-}(p)$ indicate the thresholds corresponding to the increased and decreased contrast.

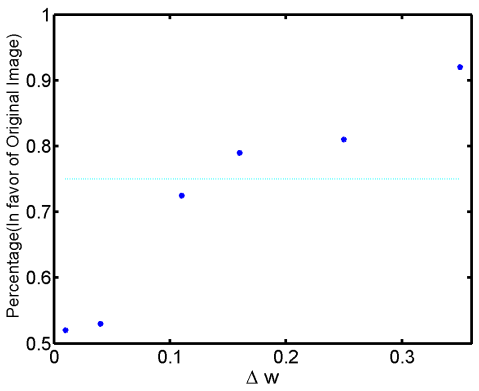


Fig. 10. The relationship between Δw and the percentage of the preference on original SCIs.

Moreover, (16) implies that people are more tolerating to the contrast increase because of the following relationship

$$T_{c+}(p) - c(p) > c(p) - T_{c-}(p). \tag{17}$$

This is in accordance with a general rule in image quality assessment (IQA) and psychology studies that humans are more likely to remember unpleasant experiences (e.g. contrast decrease) compared to pleasant moment (e.g. contrast increase) [49], [50].

D. Edge Structural Distortion Sensitivity

As demonstrated in Fig. 3 (c), the edge width is a deterministic factor that reflects the structure of the edge profile and furthermore the sharpness/blurriness of the image [40]. In (4), the probability of detecting the blur distortion is determined by the edge width and the threshold up to which HVS can mask the blurriness. However, setting a constant threshold on the width of the edge may not be appropriate for SCI, as two high contrast edge profiles within the same SCI may have apparently different widths. Instead, here we investigate the visibility threshold on the relative change of w,

$$\Delta w = w_t - w, \tag{18}$$

where w_t is calculated via parameter estimation based on the distorted SCI. To generate these distorted versions, Gaussian smoothing is performed by 2D convolution on the original SCIs with different kernels. As such, the widths of edges are increased, and the average Δw across each image is computed. A subjective experiment is subsequently conducted to study the visibility threshold on Δw, where the testing conditions follow the descriptions in Section III-B as well. Specifically, each SCI is alternated by six levels of distortions with the convolution kernels as follows,

$$\sigma = [0.3, 0.5, 0.65, 0.8, 1, 1.2], \tag{19}$$

where σ is the standard deviation in the Gaussian function. The plot between Δw and the percentage in favor of the original image is demonstrated in Fig. 10. One can discern that when Δw is set around 0.1 (corresponding to $\sigma = 0.65$), the probability of 75% in favor of the original SCIs can be achieved.

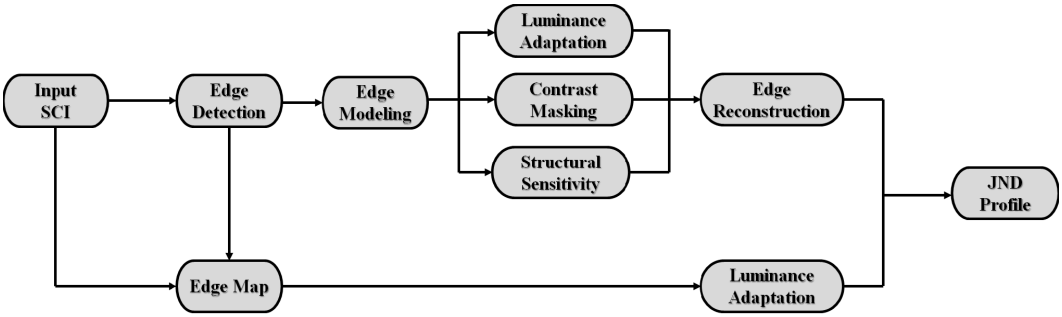


Fig. 11. The proposed JND estimation process.

IV. Proposed JND Model for SCI

In this section, we discuss how to incorporate the three masking effects that co-exist in SCIs into a unified JND model. We firstly study the JND for edge profile. Following the description of Section III-A that adopts the one-dimensional notation, the visibility threshold on luminance adaptation for pixel p can be calculated as,

$$T_{el}(p) = s(p; T_l(p) + b, c, w, x_0) - s(p; b, c, w, x_0). \quad (20)$$

In other words, the luminance adaptation effect of the pixels on the edge profile is determined by the $T_l(p)$.

In a similar fashion, contrast masking effect is accounted for by the reconstruction of edge profiles with the visibility threshold on contrast,

$$\begin{aligned} T_{ec}(p) = \min\{&|s(p; b, T_{c+}, w, x_0) - s(p; b, c, w, x_0)|, \\ &|s(p; b, T_{c-}, w, x_0) - s(p; b, c, w, x_0)|\}, \end{aligned} \quad (21)$$

where the min operation ensures that the lower value is chosen as the just noticeable threshold.

Given $T_{el}(p)$ and $T_{ec}(p)$, the nonlinear additivity model for masking (NAMM) serves as an effective approach to account for the overlapping effect by adding two masking effects [10], [13], [14],

$$T_{ns}(p) = T_{el}(p) + T_{ec}(p) - C_{ns} \cdot \min\{T_{el}(p), T_{ec}(p)\}, \quad (22)$$

where C_{ns} $(0 < C_{ns} < 1)$ is a constant that deduces the overlapping between $T_{el}(p)$ and $T_{ec}(p)$.

Regarding the visibility threshold of the structural distortion, it is computed as

$$T_s(p) = |s(p; b, c, w + \Delta w, x_0) - s(p; b, c, w, x_0)|. \quad (23)$$

In this manner, pixels in different positions from the same edge profile correspond to different non-structural and structural visibility thresholds. The combination of structural and non-structural masking effects leads to the overall edge profile JND model,

$$T_e(p) = T_s(p) + T_{ns}(p) - C \cdot \min\{T_s(p), T_{ns}(p)\}. \quad (24)$$

Again, the parameter C $(0 < C < 1)$ is a constant that deduces the overlapping between $T_s(p)$ and $T_{ns}(p)$. In this work, we set both C and C_{ns} to be 0.2. It is also worth mentioning that the edge profile JND model accounts for the strong and thin edges in textual regions, as well as the edges in natural images. This originates from the design philosophy of the visibility threshold derivation, where both strong and weak, thick and thin edges are considered.

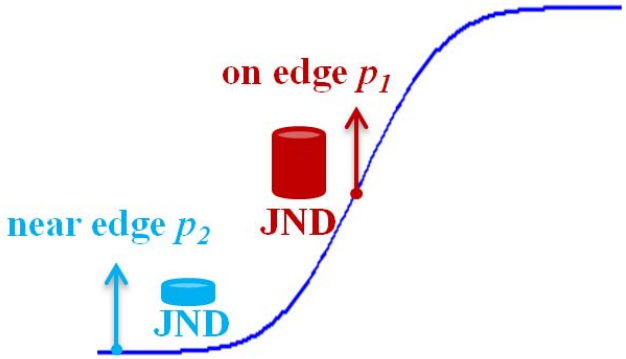


Fig. 12. Illustration of on edge and near edge pixels.

The JND derivation process of SCI is summarized in Fig. 11. In this framework, the SCI content is firstly distinguished based on edge detection in (10) to obtain the edge pixel set S_E, which is composed of pixels locating along the edge profile. Subsequently, the parameters of each edge profile are derived, and the tolerable thresholds of luminance adaptation, contrast masking and structural sensitivity are obtained. Finally, the edge and pictorial JND profiles are established and merged together,

$$T(p) = \begin{cases} T_e(p) & p \in S_E \\ T_l(p) & \text{otherwise.} \end{cases} \quad (25)$$

A distinct property of the proposed JND model is that it estimates the JND value for textual content at a finer scale by distinguishing the pixels on and near the sharp transitions in an edge profile. As illustrated in Fig. 12, the on edge pixel p_1 and near edge pixel p_2 share similar background as they are close to each other, resulting in similar JND values when applying the traditional JND models, which estimate the JND profile relying on the background luminance and activities. However, in the proposed model they are distinguished by the detected edge, implying that $p_1 \in S_E$ while $p_2 \notin S_E$. Therefore, different models are applied to derive the JND, such that $T(p_1) = T_e(p_1)$ and $T(p_2) = T_l(p_2)$. The JND for pixels on edge profile is derived from the nonlinear additivity model by incorporating the masking effects of luminance, contrast and structure, which possibly leads to $T(p_1) > T(p_2)$. This is conceptually consistent with the characteristics of HVS, as when much noise is injected into the edge boundary, the

Fig. 13. Demonstration of SCIs used in the validation. (The first 10 SCIs are from database [25] and the last 5 SCIs are from HEVC range extension test sequences.)

TABLE II

JND ENERGY COMPARISONS OF DIFFERENT MODELS

Image Num	Yang	Zhang	Liu	Wu	Proposed
1	33.417	9.425	57.052	111.031	65.274
2	20.730	10.869	53.387	104.190	57.671
3	47.522	10.495	78.857	122.625	92.575
4	19.173	10.934	45.357	142.733	61.904
5	33.040	11.716	63.143	145.686	74.146
6	24.700	11.454	77.270	286.680	90.632
7	48.232	9.535	69.897	146.915	81.837
8	42.640	10.256	63.059	381.513	68.348
9	37.958	9.375	67.321	185.263	91.568
10	35.363	11.908	83.878	245.212	97.727
11	89.165	13.182	102.063	171.899	97.359
12	65.978	13.259	81.417	112.085	83.734
13	82.964	16.032	103.745	194.717	99.412
14	27.805	8.456	36.328	55.330	34.042
15	19.647	12.773	74.775	270.849	83.148
Average	**41.889**	**11.311**	**70.503**	**178.449**	**78.625**

spurious signals near sharp transitions may easily get noticed. Therefore, the near-threshold tolerable distortion for p_2 should be smaller than p_1, benefiting from contrast masking effect by the strong edges.

V. EXPERIMENTAL RESULTS

In this section, extensive experiments are carried out to evaluate the performance of the proposed JND model. As shown in Fig. 13, fifteen SCIs from both the screen content quality assessment database [25] and HEVC range extension test sequences are used for testing. It is noted that to achieve cross-validation, the SCIs in this experiments are different from the SCIs used in the subjective testing of Section III. Application scenarios of these test SCIs include web browsing, cloud CAD and word editing, etc. The JND noise is injected into SCIs and the performance is compared with four existing JND models, including Yang *et al.*'s [10], Liu *et al.*'s [14], Zhang *et al.*'s [51] and Wu *et al.*'s methods [15].

A. Comparisons on Distortion Masking Ability

In the first experiment, the noise is injected into the SCI to evaluate the HVS error tolerance ability. The distorted image is generated by adding the JND profile to the original image in pixel domain,

$$\bar{I}(p) = I(p) + r \cdot T(p), \tag{26}$$

where r is randomly set as $+1$ or -1.

We evaluate the performance of the proposed JND model from two perspectives. Firstly, the error tolerance ability is evaluated in terms of the energy of the JND signal by averaging the $T(p)^2$ over the whole SCI. In other words, the mean square error (MSE) between $\bar{I}(p)$ and $I(p)$ is used to measure the tolerated error. For the proposed and state-of-the-art methods, the JND energy comparisons are demonstrated in Table II. It is observed that except Wu *et al.*'s method, the proposed scheme yields higher JND energy, implying stronger ability in error tolerance. Secondly, a subjective study is further conducted to assess the quality of the noise-contaminated SCIs, in which 20 subjects (13 males, 7 females) were invited. It is noted that these 20 subjects are different from those subjects invited in Section III. The testing conditions are identical as described in Section III. The guidelines in the subjective experiments are shown in Table III, and the subjects were asked to offer their opinions on the subjective quality of the images following the rating criterions. Specifically, two images were randomly juxtaposed on the same screen. The subjective scores are finally converted with the unified order that the right is the original image as the reference and the left is the JND-injected version. Specific instructions and training sessions were given before the test. The average subjective values are calculated to demonstrate the image visual quality, which is illustrated in Table IV. It is observed that the proposed method has

TABLE III

RATING CRITERIONS FOR SUBJECTIVE EVALUATION

Description	Score
The right one is much worse than the left one	-3
The right one is worse than the left one	-2
The right one is slightly worse than the left one	-1
The right one has the same quality as the left one	0
The right one is slightly better than the left one	1
The right one is better than the left one	2
The right one is much better than the left one	3

TABLE IV

SUBJECTIVE EVALUATION SCORES (ORIGINAL VS. JND NOISE CONTAMINATED SCIS)

Image Num	Yang	Zhang	Liu	Wu	Proposed
1	0.15	0.15	0.65	1.45	0.35
2	0.35	0.05	0.45	1.65	0.20
3	0.20	0.15	0.75	1.10	0.55
4	0.75	0.15	0.45	2.60	0.25
5	0.65	0.40	0.50	1.05	0.50
6	0.40	0.30	0.70	2.00	0.40
7	0.80	0.35	0.55	1.35	0.50
8	0.25	0.15	0.35	1.80	0.25
9	0.35	0.45	0.50	2.05	0.45
10	0.45	0.25	0.95	1.05	0.30
11	0.30	0.10	0.40	1.70	0.25
12	0.50	0.20	0.50	1.85	0.50
13	0.20	0.15	0.55	2.50	0.05
14	0.15	0.05	0.60	1.60	0.40
15	0.25	0.20	0.70	2.65	0.25
Average	**0.38**	**0.21**	**0.57**	**1.76**	**0.35**

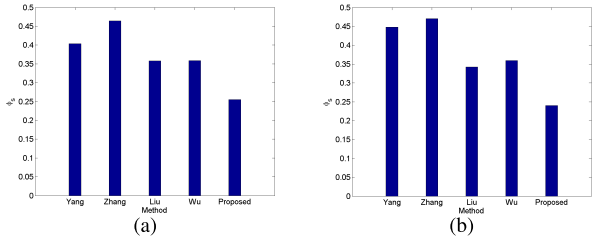


Fig. 14. The relative strength of JND for non-edge pixels using different JND models. (a) The first SCI in Fig. 13; (b) The second SCI in Fig. 13.

lower mean value than Yang *et al.*'s and Liu *et al.*'s methods, demonstrating that it has better distortion masking ability. Though Zhang *et al.*'s method achieves the best visual quality, the average JND energy of Zhang *et al.*'s method is much lower than the proposed method. Moreover, the subjective quality of Wu *et al.*'s model is lowest, as the AR model cannot achieve accurate prediction on textual content, making it difficult to estimate the visibility thresholds. These results demonstrate the superior performance of the proposed model in terms of distortion masking ability.

Furthermore, we examine the JND thresholds derived for edge pixel set S_E and non-edge pixel set S_{NE}. In particular, given the JND map, the relative strength of JND for the non-edge pixels is defined as

$$\phi_s = \frac{\sum_{p\in S_{NE}} T(p)/|S_{NE}|}{\sum_{p\in S_E} T(p)/|S_E| + \sum_{p\in S_{NE}} T(p)/|S_{NE}|}, \tag{27}$$

where $|\cdot|$ indicates the number of elements in a set. Therefore, ϕ_s evaluates the percentage of noise injected to the non-edge regions, and higher ϕ_s corresponds to higher noise for non-edge pixels given the same level of JND thresholds. In Fig. 14, the ϕ_s values for the first and second SCIs in Fig. 13 are presented, and the results show that the proposed scheme has lower ϕ_s compared to the state-of-the-art methods. This is also conceptually consistent with the discussions in Section IV that the JND thresholds for edge pixels are higher than the surrounding non-edge pixels for better distortion masking.

B. Comparisons on Distortion Shaping Accuracy

In the second experiment, we aim to examine whether the proposed JND model is better at guiding the shaping of the noise distribution. Specifically, identical amounts of JND energy are maintained for different models by injecting the noise with JND profile regulation,

$$\tilde{I}(p) = I(p) + \beta \cdot r \cdot T(p), \tag{28}$$

where $\tilde{I}(p)$ denotes the noise-contaminated pixel with β ensuring that each distorted SCI shares the same JND energy.

A subjective test was further conducted to compare the quality of SCIs contaminated by the proposed and conventional JND models. Again, the scores are converted with the unified order that the right is the SCI produced by the proposed method and the left is the SCI generated with one conventional JND model. The comparison results of the subjective viewing test are shown in Table V. One can discern that the proposed model outperforms the other four models in most of the cases. The results demonstrate that the proposed model is effective in guiding the shaping of noise, such that more noise can be injected into the insensitive pixels, leading to better visual quality. It is also observed that occasionally for some comparisons the proposed scheme has lower quality. One reason may be attributed to the proposed distortion shaping strategy that may not always follow the saliency of the human viewers.

TABLE V
SUBJECTIVE EVALUATION SCORES (THE PROPOSED MODEL VS. FOUR CONVENTIONAL JND MODELS)

Image Num \ Proposed vs.	Yang	Zhang	Liu	Wu
1	0.95	1.05	0.40	0.45
2	0.45	0.70	0.45	-0.20
3	-0.15	0.65	0.25	0.25
4	0.95	0.70	0.70	1.20
5	0.10	0.85	-0.35	-0.10
6	0.20	1.20	0.00	1.55
7	0.45	0.55	-0.55	-0.20
8	0.60	0.30	0.80	0.30
9	0.95	0.15	0.50	1.30
10	0.70	0.90	0.75	0.75
11	0.35	-0.10	0.20	1.90
12	0.00	0.05	0.15	1.65
13	-0.30	0.30	0.55	2.10
14	0.75	0.85	1.10	1.15
15	1.20	1.20	0.65	0.90
Average	**0.48**	**0.62**	**0.37**	**0.87**

Moreover, the noise injected SCIs that are generated by different JND models with identical JND energy are demonstrated in Figs. 15&16. Scrupulous observes may find that the proposed method is better at preserving the edge structure and shaping the noise, such that the distortions get less noticeable compared with the conventional approaches. These results further confirm that the proposed JND model is more appropriate for SCI content.

VI. JND BASED PERCEPTUALLY LOSSLESS SCI COMPRESSION

In this section, we incorporate the proposed JND profile into the screen content coding (SCC) extensions of HEVC [32], targeting at perceptually lossless SCI compression. In the SCC extension of HEVC, many advanced coding tools have been proposed to improve the coding performance [33]. For example, inspired by the fact that SCIs contain limited number of colors rather than the continuous color tone in natural content, the base color and index map representation for SCIs has been intensively studied in the literature [52]–[55]. Moreover, to account for the repeated patterns occurred in the SCI, the intra motion compensation approach was developed for screen content coding [56]. To further reduce the redundancy among the color components, the adaptive color-space transform was incorporated into the coding loop in [57].

Perceptually lossless refers that no discernable visual difference compared with the original SCIs is incurred during the compression process [13]. Specifically, with the proposed JND modeling, perceptually lossless implies that,

$$D = \sum D(p) = 0, \tag{29}$$

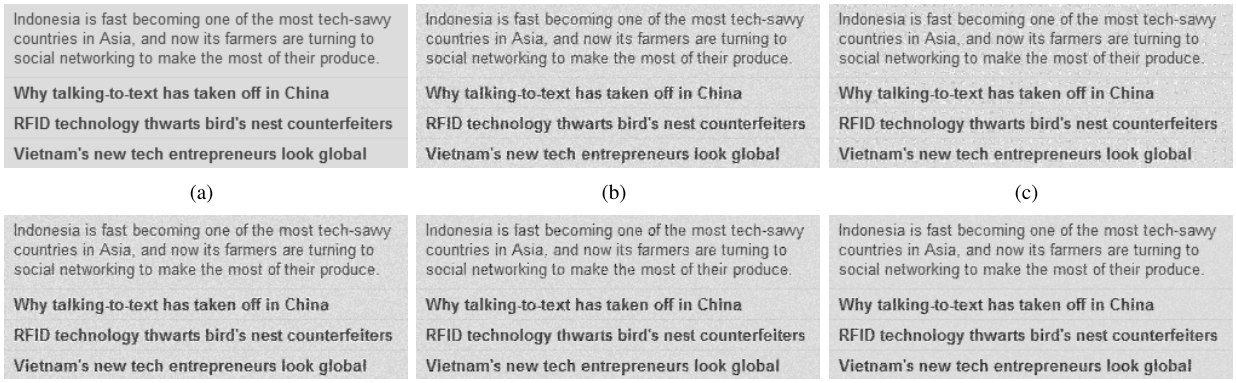


Fig. 15. Visual quality comparison of the distorted SCIs generated by different JND models for webpage textual content (cropped for better visualization). (a) Original. (b) Yang *et al.*'s method. (c) Zhang *et al.*'s method. (d) Liu *et al.*'s method. (e) Wu *et al.*'s method. (f) Proposed method.

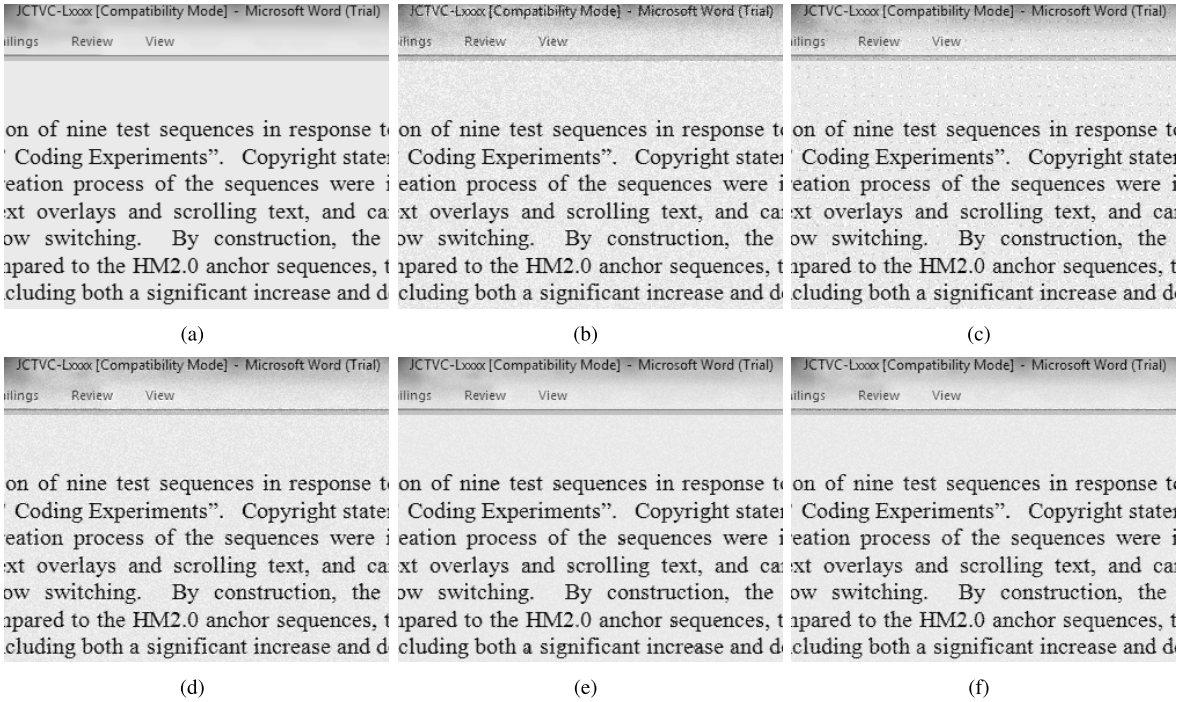


Fig. 16. Visual quality comparison of the distorted SCIs generated by different JND models for document textual content (cropped for better visualization). (a) Original. (b) Yang *et al.*'s method. (c) Zhang *et al.*'s method. (d) Liu *et al.*'s method. (e) Wu *et al.*'s method. (f) Proposed method.

where

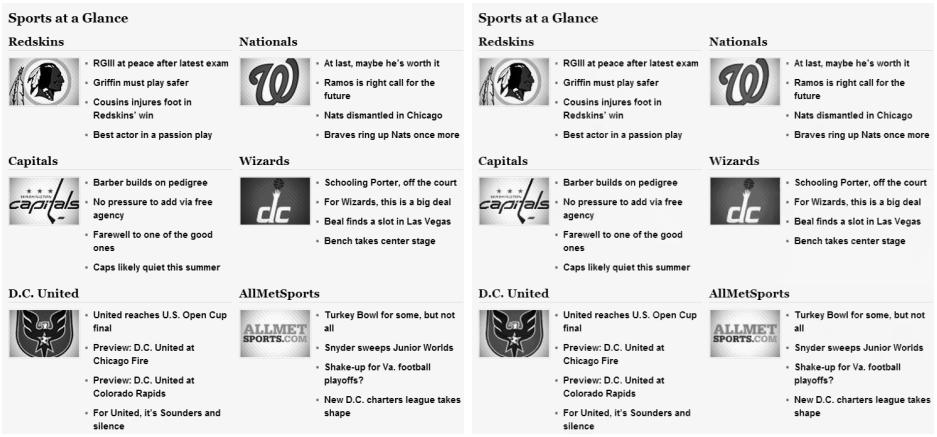
$$D(p) = \begin{cases} 0 & |I_o(p) - I_d(p)| \le T(p) \\ (I_o(p) - I_d(p))^2 & \text{otherwise.} \end{cases} \tag{30}$$

Here $I_o(p)$ and $I_d(p)$ indicate the original and compressed pixel values for p. As such, all the irreversible information loss lies within the range of JND threshold.

To achieve lossless SCI compression, the lossy modules such as transform and quantization are bypassed in HEVC SCC extension. Following this framework, the prediction residual is firstly shrunk by the JND profile and then lossless compressed with the employed codec. Assume the prediction residual for pixel p is $R(p)$, it is compressed in the following way,

$$\hat{R}(p) = \begin{cases} 0 & |R(p)| \le T(p) \\ R(p) - T(p) & R(p) > T(p) \\ R(p) + T(p) & \text{otherwise.} \end{cases} \tag{31}$$

In the implementation, we incorporate the perceptually lossless coding scheme into the newly developed HEVC extension codec HM-16.4+SCM-4.0 [58]. The test SCIs are identical with those employed in Section V. The performance in terms of the coding bits for each pixel is demonstrated in Table VI. Specifically, "Anchor" indicates the lossless compression of SCIs and the proposed scheme is the perceptually lossless coding with the proposed JND model. It is observed that on average 24% rate reduction is achieved. The bit savings originate from the shrinkage of the residual energy.



(a) (b)

Fig. 17. An example of perceptually lossless coding for SCI. (a) Original SCI. (b) The perceptually lossless coded SCI.

TABLE VI
CODING BITS COMPARISON

Image Num	Anchor (bpp)	Proposed (bpp)	Saving
1	1.46	1.17	20.26%
2	0.81	0.75	8.25%
3	1.87	1.16	38.22%
4	1.15	0.85	25.52%
5	2.30	1.37	40.60%
6	2.56	1.82	29.02%
7	1.77	1.28	27.78%
8	1.23	0.90	27.08%
9	1.68	1.20	28.56%
10	2.24	1.49	33.60%
11	0.19	0.18	6.61%
12	0.30	0.21	30.25%
13	0.14	0.11	17.57%
14	0.58	0.50	14.38%
15	0.47	0.42	9.75%
Average	**1.25**	**0.89**	**23.83%**

TABLE VII
SUBJECTIVE EVALUATION OF PERCEPTUALLY LOSSLESS COMPRESSION (THE PROPOSED MODEL VS. FOUR CONVENTIONAL MODELS)

Score \ Proposed vs.	Yang	Zhang	Liu	Wu
Average	0.24	0.31	0.22	0.37

Fig. 17 shows the original and perceptually lossless coded SCI with the proposed method. Note that for this SCI the bit saving is around 25%. However, since our proposed scheme is based on JND profile optimization for perceptually lossless coding, the perceptual distortions of both SCIs are equivalent to zero. It can be observed that though the perceptually lossless compression has undergone a irreversible information loss process, little discernable artifacts can be observed.

To further validate our scheme, we carry out a subjective quality evaluation test to compare the perceptually lossless compressed SCIs by incorporating the proposed and conventional JND models. Specifically, we also follow the subjective testing guidelines in Section V-B, where two SCIs generated by the proposed and one conventional JND method are compared. Again, the JND profiles generated by the conventional methods are regulated by enforcing the scaling parameter β as in (28), ensuring that exactly the same bit rate savings as reported in Table VI are achieved. As such, better reconstructed SCI quality corresponds higher compression performance. The subjective testing results obtained by averaging the 15 SCIs are reported in Table VII, which demonstrate that the SCIs generated by the proposed JND model have better quality. These results clearly demonstrate that the proposed method can efficiently achieve the perceptually lossless coding with better perceptual lossless coding performance.

VII. CONCLUSION

We have proposed a JND model that is specifically designed for screen content images. The novelty of the model lies in computing the JND at a finer scale by introducing a parametric edge model, which provides a feasible way to estimate the visibility thresholds of three conceptually independent components including luminance, contrast and structure. We demonstrate the effectiveness of the JND model and compare it with conventional schemes by subjective testing. The JND model is further incorporated into the SCI compression, which provides evidence that the proposed model can significantly improve the coding efficiency.

The SCI JND model can play a variety of roles in the transport of screen visuals. First, it can be used to optimize the screen compression algorithms in the interactive screen-remoting system. In this work, we incorporate the JND model into the prediction coding process to shrink the residual energy. Additionally, it can also influence various modules in the compression pipeline, such as data discarding (via JND based quantization), codec optimization (via JND inspired rate-distortion optimization) and postprocessing (via JND guided loop filter). Second, it can further help the design of meaningful quality metrics to dynamically monitor and adjust the SCI quality. Third, it can be applied in the screen streaming system to optimize the transmission strategies and further improve the visual quality-of-experience.

REFERENCES

[1] H. Shen, Y. Lu, F. Wu, and S. Li, "A high-performanance remote computing platform," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2009, pp. 1–6.

[2] R. A. Baratto, L. N. Kim, and J. Nieh, "Thinc: A virtual display architecture for thin-client computing," *ACM SIGOPS Oper. Syst. Rev.*, vol. 39, no. 5, pp. 277–290, 2005.

[3] H. Shen, Z. Pan, H. Sun, Y. Lu, and S. Li, "A proxy-based mobile Web browser," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 763–766.

[4] *Virtual Network Computing (VNC)*, accessed on May. 2016. [Online]. Available: http://www.realvnc.com

[5] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud-mobile convergence," *IEEE Multimedia Mag.*, vol. 18, no. 2, pp. 4–11, Feb. 2011.

[6] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 173–185, Jan. 2013.

[7] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 900–909, Jul. 2005.

[8] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742–752, Jun. 2005.

[9] M. Bouchakour, G. Jeannic, and F. Autrusseau, "JND mask adaptation for wavelet domain watermarking," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun./Apr. 2008, pp. 201–204.

[10] 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.

[11] T.-H. Huang, C.-K. Liang, S.-L. Yeh, and H. H. Chen, "JND-based enhancedment of perceptibility for dim images," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 1752–1755.

[12] 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.

[13] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025–2043, Sep. 2013.

[14] 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.

[15] 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.

[16] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Sid Int. Symp. Dig. Tech. Papers*, vol. 24, p. 946, Dec. 1993.

[17] I. Höntsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 213–222, Mar. 2002.

[18] 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.

[19] L. Ma, K. N. Ngan, F. Zhang, and S. Li, "Adaptive block-size transform based just-noticeable difference model for images/videos," *Signal Process., Image Commun.*, vol. 26, no. 3, pp. 162–174, Mar. 2011.

[20] C.-H. Chou and K.-C. Liu, "A perceptually tuned watermarking scheme for color images," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2966–2982, Nov. 2010.

[21] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[22] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.

[23] M. Castelhano and K. Rayner, "Eye movements during reading, visual search, and scene perception: An overview," in *Proc. Cognit. Cult. Influences Eye Movements*, 2008, pp. 175–195.

[24] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," in *Proc. IEEE 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 257–262.

[25] H. Yang, Y. Fang, Y. Yuan, and W. Lin, "Subjective quality evaluation of compressed digital compound images," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 105–114, Jan. 2014.

[26] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Comput. Graph. Appl.*, to be published.

[27] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[28] P. J. L. van Beek, "Edge-based image representation and coding," Ph.D. dissertation, Dept. Elect. Eng., Inf. Theory Group, Delft Univ. Technol., The Netherlands, 1995.

[29] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[30] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, Apr. 2013.

[31] X. Liu, D. Zhai, J. Zhou, X. Zhang, D. Zhao, and W. Gao, "Compressive sampling-based image coding for resource-deficient visual communication," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2844–2855, Jun. 2016.

[32] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.

[33] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 50–62, Jan. 2016.

[34] D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vis. Res.*, vol. 37, no. 23, pp. 3367–3383, Dec. 1997.

[35] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.*, vol. 21, no. 3, pp. 409–418, Mar. 1981.

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

[37] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[38] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London. B, Biol. Sci.*, vol. 207, pp. 187–217, Feb. 1980.

[39] M. C. Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proc. Roy. Soc. London Ser. B, Biol. Sci.*, vol. 235, no. 1280, pp. 221–245, 1988.

[40] J. Guan, W. Zhang, J. Gu, and H. Ren, "No-reference blur assessment based on edge modeling," *J. Vis. Commun. Image Represent.*, vol. 29, pp. 1–7, May 2015.

[41] W. Zhang and W.-K. Cham, "Single-image refocusing and defocusing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 873–882, Feb. 2012.

[42] Z. Pan, H. Shen, Y. Lu, S. Li, and N. Yu, "A low-complexity screen compression scheme for interactive screen sharing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 949–960, Jun. 2013.

[43] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[44] 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.

[45] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 616–625, Oct. 2012.

[46] *Measuring Image Quality: Sarnoff's JNDmetrix Technology, Sarnoff JNDmetrix Technology Overview*, Sarnoff Corp., Princeton, NJ, USA, Jul. 2002.

[47] J. M. Foley, "Human luminance pattern-vision mechanisms: Masking experiments require a new model," *J. Opt. Soc. Amer. A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[48] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.

[49] Y. Kawayoke and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 385–388.

[50] K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Process.*, vol. 70, no. 3, pp. 279–294, 1998.

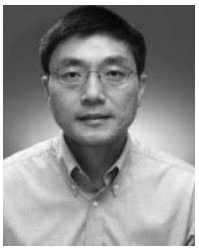
[51] X. H. Zhang, W. S. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Process.*, vol. 85, no. 4, pp. 795–808, 2005.

[52] C. Lan, G. Shi, and F. Wu, "Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 946–957, Apr. 2010.
[53] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Screen content coding based on HEVC framework," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1316–1326, Aug. 2013.
[54] S. Wang, J. Fu, Y. Lu, S. Li, and W. Gao, "Content-aware layered compound video compression," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 145–148.
[55] Z. Ma, W. Wang, M. Xu, and H. Yu, "Advanced screen content coding using color table and index map," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4399–4412, Oct. 2014.
[56] D.-K. Kwon and M. Budagavi, "RCE3: Results of test 3.3 on intra motion compensation," in *Proc. 14th Meeting*, vol. 25. Vienna, Austria, 2013, p. 2013.
[57] L. Zhang, J. Chen, J. Sole, M. Karczewicz, X. Xiu, and J.-Z. Xu, "Adaptive color-space transform for HEVC screen content coding," in *Proc. IEEE Data Compress. Conf.*, Apr. 2015, pp. 233–242.
[58] JCTVC. *HM16.4+SCM-4.0rc1*, accessed on May, 2016. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.4+SCM-4.0rc1

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from Peking University, in 2014. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. He is currently a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

Lin Ma (M'13) received the B.E. and M.E. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, and the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong (CUHK), in 2013. He was a Research Intern with Microsoft Research Asia from 2007 to 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from 2008 to 2009. He was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, from 2011 to 2011. He is currently a Researcher with the Huawei Noah's Ark Laboratory, Hong Kong. His research interests lie in the areas of deep learning and multimodal learning, specifically for image and language, image/video processing, and quality assessment. He received the best paper award in the Pacific-Rim Conference on Multimedia 2008. He also received the Microsoft Research Asia Fellowship in 2011. He was a finalist to the HKIS Young Scientist Award in engineering science in 2012.

Yuming Fang received the B.E. degree from Sichuan University, and the M.S. degree from the Beijing University of Technology, China, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. He was a (Visiting) Researcher with National Tsinghua University, Taiwan, the IRCCyN Laboratory, École Polytechnique de l'Université de Nantes, Nantes, France, and the University of Waterloo, Waterloo, Canada. He is currently an Associate Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, computer vision, and 3D image/video processing. He was a Special Session Organizer or Session Chair in many international conferences, such as QoMEX 2014, VCIP 2015, and ICME 2016.



Weisi Lin (SM'98–F'16) received the Ph.D. degree from Kings College, London University, London, U.K., in 1993. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, and served as a Laboratory Head of Visual Processing with the Institute for Infocomm Research. He authored over 300 scholarly publications, held seven patents, and received over U.S. $4 million in research grant funding. He has maintained active long-term working relationship with a number of companies. His research interests include image processing, video compression, perceptual visual and audio modeling, computer vision, and multimedia communication. He served as an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*. He is also on six IEEE technical committees and the technical program committees of a number of international conferences. He was the Lead Guest Editor for a special issue on perceptual signal processing of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING in 2012. He is a Chartered Engineer in the U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He co-chaired the IEEE MMTC special interest group on the quality of experience. He was an Elected Distinguished Lecturer of APSIPA in 2012/2013.

Siwei Ma (S'03–M'12) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. From 2005 to 2007, he held a post-doctoral position with the University of Southern California, Los Angeles. Then, he joined the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, where he is currently a Professor. He has authored over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Japan, in 1991. He is a Professor of Computer Science with Peking University, China. Before joining Peking University, he was a Professor of Computer Science with the Harbin Institute of Technology from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences. He has authored extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. He served or serves on the Editorial Board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.