Reduced-Reference Quality Assessment of Screen Content Images

Shiqi Wang, *Member, IEEE*, Ke Gu, Xinfeng Zhang, Weisi Lin, *Fellow, IEEE*, Siwei Ma, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— The screen content images (SCIs) quality influences the user experience and the interactive performance of remote computing systems. With numerous approaches proposed to evaluate the quality of natural images, much less work has been dedicated to reduced-reference image quality assessment (RR-IQA) of SCIs. Here, we propose an RR-IQA method from the perspective of SCI visual perception. In particular, the quality of the distorted SCI is evaluated by comparing a set of extracted statistical features that consider both primary visual information and unpredictable uncertainty. A unique property that differentiates the proposed method from previous RR-IQA methods for natural images is the consideration of behaviors when human subjects view the screen content, which motivates us to establish the perceptual model according to the distinct properties of SCIs. Validations based on the screen content IQA database show that the proposed algorithm provides accurate predictions across a wide range of SCI distortions with negligible transmission overhead.**

***Index Terms*— Image quality assessment (IQA), reduced reference (RR), screen content images (SCIs).**

I. Introduction

RECENTLY, the popularity of screen virtualization has been creating an ever-stronger demand for efficient screen content image (SCI) compression and quality assessment methods. In a variety of remote processing and virtual desktop application scenarios, the refreshed screen is rendered, compressed, and transmitted to the client side [1]–[3]. The screen virtualization can be achieved by user interaction with the local display interface, which is a mixture of natural image regions and textual content generated by computers. In these applications, objective image quality assessment (IQA) methods that are capable of accessing the perceived quality of SCIs are highly desired, as they can not only be used to monitor the screen quality of remote computing systems, but also provide a feasible way in devising and optimizing advanced image/video processing algorithms [4]–[6].

IQA models that are capable of automatically predicting the perceived quality of natural images have enjoyed popularity for decades. Most of the existing IQA methods require the full reference (FR) information, and popular FR methods include the structural similarity (SSIM) index [7], feature-similarity (FSIM) [8], gradient similarity (GSIM) [9], visual information fidelity (VIF) [10], and visual saliency induced index (VSI) [11]. However, SCIs exhibit quite different characteristics when compared with natural images. For instance, the computer generated SCIs are usually noise free and composed of thin edges with a limited number of colors. By contrast, natural images are composed of continuous-tone structures [12]. Another interesting observation is that the semantic information in SCI is mainly interpreted relying on the eye movements in the textual content. However, for natural images, the human visual system (HVS) is exquisitely adapted to extract the conceptual information from visual input with every new eye fixation [13]–[17]. These distinctive features of SCIs motivate us to carry out a further investigation on their perceptual characteristics, which are essential in developing trusted SCI IQA models.

In view of the importance of SCI quality assessment, in [18] a database of distorted SCIs with subjective quality rankings has been created, which includes seven common distortion types. It contains 980 distorted SCIs generated by corrupting 20 source SCIs in various scenarios. The results demonstrate that there is still a lack of accurate SCI quality assessment methods. This further inspires FR-IQA models that predict the SCI quality based on novel weighting strategies [18]–[20].

However, in typical remote computing systems, the reference image is not available at the client side. Thus, developing IQA algorithms that only require significantly less information of the original image is necessary and meaningful for applications in real scenarios. Traditionally, these algorithms can be categorized into reduced-reference (RR) and no-reference (NR) methods. Typically, NR algorithms are usually developed with the assumption of the distortion process [21]–[23]. Due to the absence of the information from the reference image, NR-IQA methods are usually less efficient in providing accurate predictions of the subjective quality. Fortunately, the RR-IQA achieves a good compromise between the FR and NR algorithms by comparing a few features [24]. Basically, the extracted features from the reference image are transmitted to the client side. In the literature, vari-

Manuscript received October 3, 2015; revised February 25, 2016 and March 19, 2016; accepted August 10, 2016. Date of publication August 25, 2016; date of current version January 5, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61322106 and Grant 61571017, in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800, and in part by the Singapore MoE Tier 1 Project under Grant M4011379 and Grant RG141/14. This paper was recommended by Associate Editor D. Marpe.

S. Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong. (e-mail: shiqwang@cityu.edu.hk).

K. Gu, X. Zhang, and W. Lin are with Nanyang Technological University, Singapore 639798 (e-mail: wangshiqi@ntu.edu.sg; guke@ntu.edu.sg; xfzhang@ntu.edu.sg; 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/TCSVT.2016.2602764

ous RR-IQA models have been proposed. In [25], the RR-IQA was developed with a wavelet domain natural image statistic model (WNISM). This idea was further extended to devise the divisive normalization domain RR-IQA [26] (DNT-RR) and SSIM-based RR-IQA [27] algorithms. In [28], the RR entropic differencing-based IQA method was presented. In [29], the VIF-based RR-IQA (VIF-RR) method was proposed, which employs the autoregression model to extract the features that summarize the perceptual information. By the analyses of the distributions of discrete cosine transform (DCT) coefficients, Ma *et al.* [30], [31] reorganized the DCT coefficients and predicted the perceptual quality relying on their city-block distance. In [32], a Fourier transform domain RR method was proposed based on the phase and magnitude information. In [33], the structural degradation model (SDM) was developed, where the image quality score was obtained based on the structural information divergence between the original and distorted images. However, most of these algorithms are still designed based on natural images, making them difficult to be straightforwardly applied in SCIs.

In order to develop an efficient RR-IQA method for SCIs, there are several challenges that need to be considered. First, the selected RR features should capture the perceptual properties of SCIs and be closely relevant with image distortions, such that the perceptual quality degradation can be clearly reflected by comparing these features directly. Second, the extracted features should consume as few bits as possible. Otherwise, they may impose a heavy burden to the SCI transmission. Third, the computational complexity of the RR feature calculation and comparison should be relatively low, which makes it practical for real-time screen quality monitoring.

Recent studies on human visual perception, including the free-energy principle [34] and the Bayesian brain theory [35], reveal that HVS actively infers the input scenes with the internal generative mechanism (IGM). In particular, the brain acts as an active inference system to understand the primary visual information [29], [36]. Here, the primary visual information can be regarded as the features that account for the high level tasks of HVS for scene understanding and recognition. However, due to the hypothesis that the IGM model could not be universal [34], there exists a gap between the real scene and the prediction model from the brain. Such a gap may cause "surprise" of the human subjects, and finally lead to the unpredictable uncertainty. In other words, the uncertainty indicates the information contained in the input image that cannot be explained by HVS. As such, the amount of uncertainty can be quantified by comparing the input with the orderly signals that are from the inference procedure of the brain.

Based on the IGM-based brain theory, it is reasonable to hypothesize that the image quality is closely relevant to the primary visual information and the amount of uncertainty [37]. Inspired by the SCI FR-IQA method proposed in [19], we develop a novel RR-IQA model by formulating the quality of SCI in terms of these two components. In particular, the perception of SCI is modeled in a unique way by assuming the HVS perception channel with both Gaussian blur and motion blur to extract the orderly signals of SCIs. As such, the amount of uncertainty can be quantified by the similarity between the input and orderly signals. In particular, lower similarity corresponds to larger amount of uncertainty and vice versa. Consequently, the SCI quality is evaluated by comparing the gradient-domain features that represent the primary visual information and the similarities that evaluate the unpredictable uncertainty. Experimental results demonstrate that the RR-IQA method is capable of delivering highly competitive prediction accuracy with relatively low overhead bits and computational cost.

II. Characteristics of SCIs

It is widely acknowledged that the main functionality of SCIs is to express the rich and meaningful information with the textual content. Therefore, before the introduction of the proposed algorithm, it is useful to discuss some interesting properties of SCIs, which are meaningful in the development of efficient IQA method.

To begin with, we study the characteristics of SCIs from the perspective of edge profile representation. To investigate the edge representation in textual content, we adopt the parametric model [38]–[40], which promotes us to adaptively decompose any edge profile and examine their properties in terms of contrast and structure, respectively. In particular, in the 1D domain, a step edge x_0 can be represented by a unit step function

$$u(x; b, l, x_0) = l \cdot U(x - x_0) + b \tag{1}$$

where $U(\cdot)$ is the step function, b is the edge basis, and l denotes the edge contrast. The actual edge composition in SCIs can be treated as a smooth transition of the unit edge, and this can be achieved by convolving the step edge $u(x; b, l, x_0)$ with Gaussian filter

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

Here, erf$(\cdot)$ is the error function and w represents the edge width. Such locally adaptive representation is able to explicitly decompose any new edge profile into three physically meaningful components. In particular, the parameter b specifies the base intensity of an edge. The parameter l represents the strength of the edge, and higher l value indicates a stronger edge. The edge structure is determined by w, and smaller w corresponds to sharper edge profile. The parameters are calculated by fitting (2) with local pixel values.

In Fig. 1, we demonstrate a typical SCI, together with its local edge contrast and structure maps. We can observe that the textual edges have higher contrast but thinner width compared with the natural content. It suggests us to study the quality of SCI by computing the gradient information, which simultaneously captures the edge contrast and structure around the edge regions. As such, the distortions that are sensitive to HVS in the textual content can be efficiently detected.

Subsequently, the distinct viewing behaviors of SCIs are briefly reviewed. Regarding the studies on the viewing behaviors of SCIs, Faraday's visual scanning model [17] for Web page images first explored how the visual information is

(a) (b) (c)

Fig. 1. Analyses of edge contrast and width map based on (2) [41]. (a) SCI. (b) Edge contrast map. (c) Edge width map.

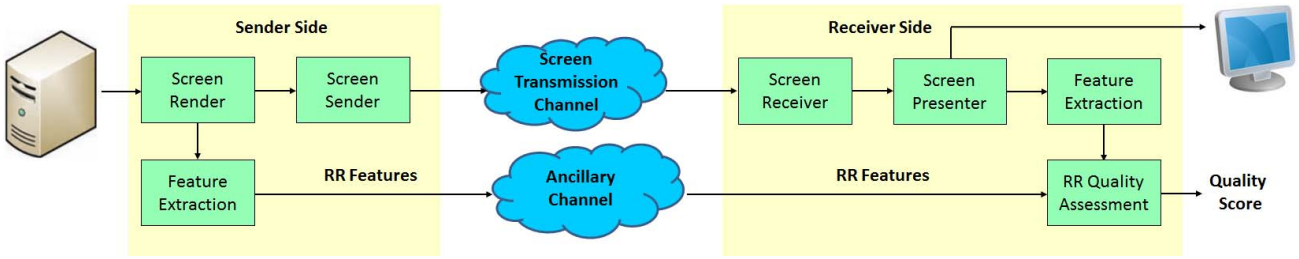


Fig. 2. General framework of an SCI RR-IQA system.

organized in a typical Web page. In particular, the viewing process is divided into two phases, including "searching" and "scanning." The "searching" phase takes place when the viewer attempts to identify a salient point in the image, and once the saliency point is detected, the "scanning" phase is subsequently applied to extract the information. Based on this strategy, Grier *et al.* [42] proposed a three-stage EHS (Expected Location, Heuristic Search, Systematic Search) theory to further explain the "search" procedure in typical Web pages. Recently, in [16], a Web page saliency database was created, and it is also observed that the textual regions in SCIs contain rich information and salient stimuli. These studies reveal that the SCIs have their own characteristics that make the viewing behaviors of them distinct from those of natural images.

III. Reduced-Reference SCI Quality Assessment

The idea of RR-IQA was first introduced in [43] as a pragmatic approach to monitor the real-time image/video quality over multimedia communication networks. We extend this philosophy to the interactive screen remoting system, as shown in Fig. 2. In particular, the server (sender side) and the client (receiver side) communicate with each other over a network through an interactive screen-remoting mechanism [1]. After receiving the client input, the servers render the new screen content and send the SCIs to the clients as a response. At the receiver side, the screen update model refreshes the display image with the received SCIs. Meanwhile, the features extracted from the sender side are transmitted and compared with that from the receiver side, such that the perceptual degradations in the interactive screen-remoting systems can be feasibly monitored. In particular, a feature extractor is applied to the captured SCI signal at the sender side. The extracted features are then transmitted to the receiver side through an error-free ancillary channel. Typically, the data rate in transmitting features as the side information is much lower than that of the SCI transmission channel. When the clients receive the distorted SCI via the error-prone channel, identical feature extraction process resembles that at the sender side. Finally, the divergence between the features that are extracted from the SCIs of the sender and receiver sides is employed to evaluate the image quality. As a result, the perceptual quality can be accurately predicted by the feature comparison. It is also worth mentioning that the RR-IQA methods for SCIs are meaningful as the generated screen content at the server side is usually noise-free. As such, the features extracted at the sender side can be regarded as the faithful source of the original SCI information. By contrast, the natural images are usually captured by physical sensors, which may inevitably introduce artifacts in the capturing process.

Given the distinct viewing behaviors of SCIs, it is worth mentioning that both statistics of natural images and properties of unnatural SCIs should be considered in the design of the proposed IQA model. The reason is that SCIs that serve as the input signals are eventually perceived by HVS. Moreover, it is widely believed that the natural environment is driving the function of HVS in the evolution process. Therefore, the natural scene statistics can be considered to account for the

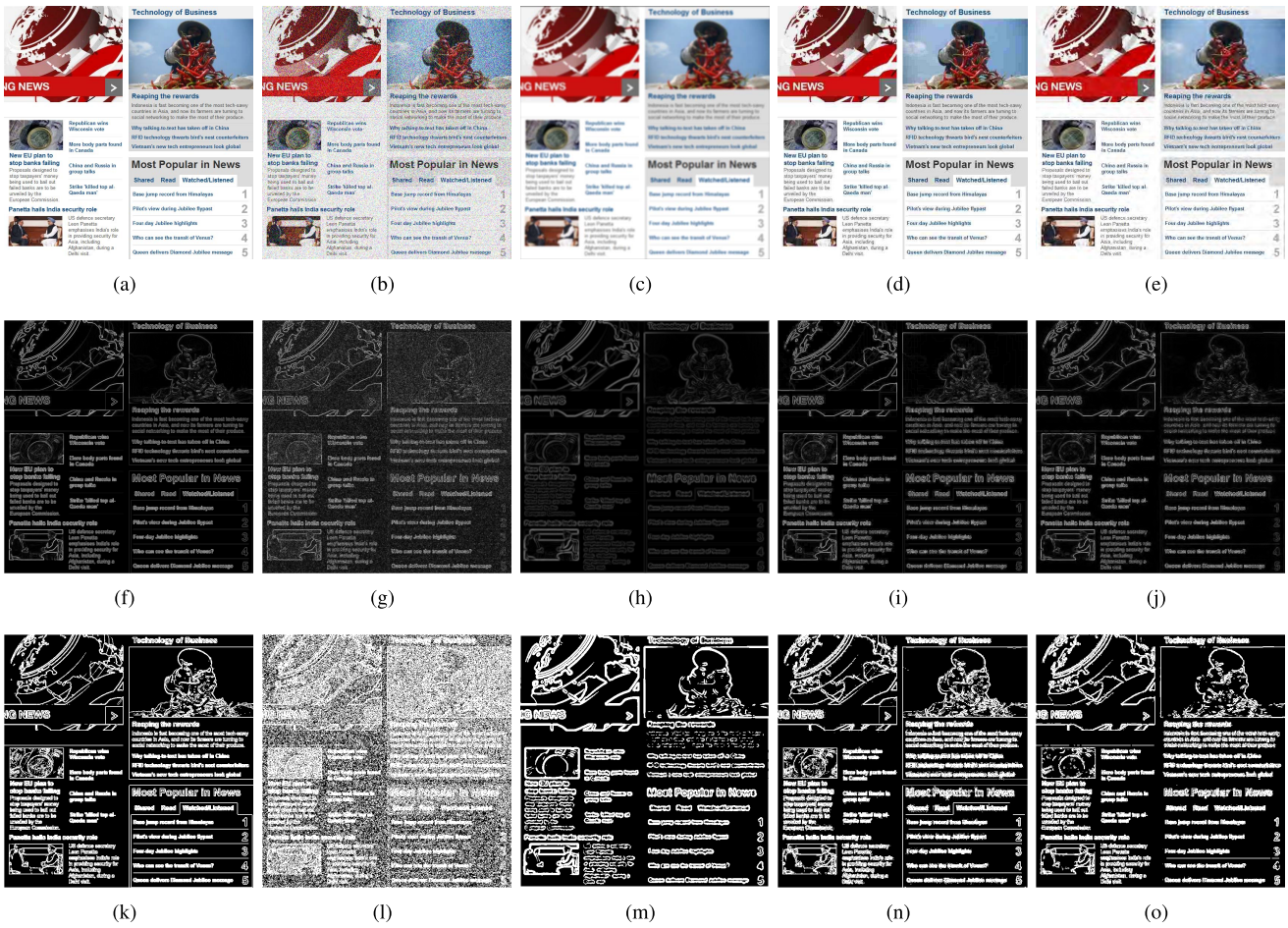


Fig. 3. Illustration of the GM and significance maps, where the top row shows the SCIs, the middle row shows the corresponding GM maps, and the bottom row shows the corresponding significance maps. (a), (f), and (k) Original. (b), (g), and (l) Gaussian noise. (c), (h), and (m) Gaussian blur. (d), (i), and (n) JPEG compression. (e), (j), and (o) JPEG2000 compression.

properties of HVS, and the properties of unnatural SCIs should be considered to model the input SCI signal.

The proposed RR-IQA approach is essentially based on the IGM, which assumes that perceptual quality relies on both the primary visual information and the uncertainty that cannot be explained by the HVS. Generally speaking, the primary visual information accounts for the high level vision tasks, such as image understanding and recognition [37]. Therefore, the distortions on the primary visual information may disturb the extraction of the information content, leading to the difficulty in image understanding. On the contrary, distortions on the uncertainty may cause uncomfortable viewing experience [29], [44]. The combination of primary visual information and the uncertainty leads to the final visual quality. Without loss of generality, in the following description, we detail the feature extraction process for the original image $\mathbf{X}$, and identical operations are also valid for the distorted image $\mathbf{Y}$.

Based upon our analyses, the textual content contains abundant high contrast edges, motivating us to study the quality of SCIs in gradient domain. The gradient domain representation was found to be an effective mechanism to account for the behaviors of HVS [8], [9], [19]. Moreover, many high level vision tasks, such as object recognition and visual cognition, have also benefited a lot from the gradient information [45], [46], which coincide with the main functionality of the primary visual information in IGM. Therefore, we employ the gradient magnitude (GM) to characterize the primary visual information

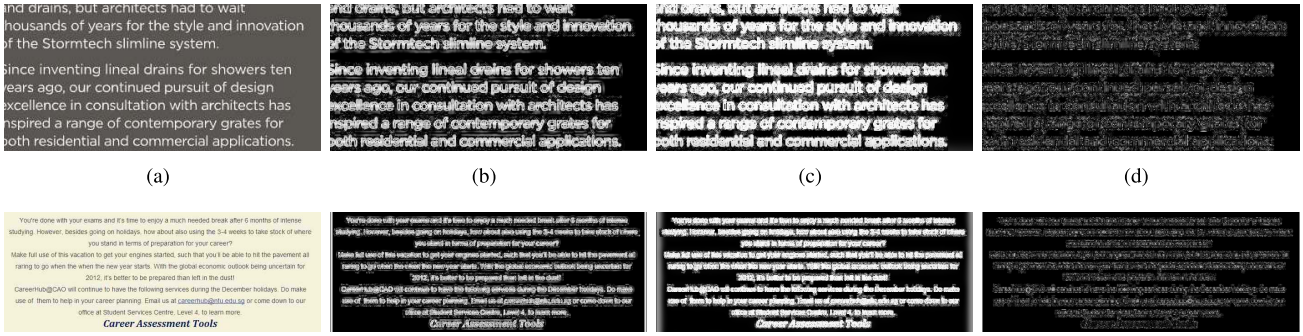
$$\mathcal{G}(\mathbf{X}) = \sqrt{\mathbf{g}_x^2(\mathbf{X}) + \mathbf{g}_y^2(\mathbf{X})} \tag{3}$$

where

$$\mathbf{g}_x(\mathbf{X}) = \mathbf{h}_x \otimes \mathbf{X} = \frac{1}{16}\begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix} \otimes \mathbf{X} \tag{4}$$

$$\mathbf{g}_y(\mathbf{X}) = \mathbf{h}_y \otimes \mathbf{X} = \frac{1}{16}\begin{bmatrix} +3 & +10 & -3 \\ 0 & 0 & 0 \\ +3 & -10 & -3 \end{bmatrix} \otimes \mathbf{X}. \tag{5}$$

Here, $\mathbf{h}_x$ and $\mathbf{h}_y$ denote the "Scharr" convolution masks that extract the gradient information from the image [47]. The GM maps for the original and distorted SCIs are shown in Fig. 3(f)–(j), which confirm that the GM information is capable of capturing the information loss caused by various types of distortions, such as Gaussian noise, Gaussian blur, JPEG, and JPEG2000 compression.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4. Illustration of S_G maps generated by comparing the input and the low-pass filtered SCIs with different parameters. (a) and (e) Distorted SCIs as the input. (b) and (f) $\sigma = 5.5$. (c) and (g) $\sigma = 11$. (d) and (h) $\sigma = 1.0$.

To incorporate the GM information in the RR-IQA algorithm, we further distinguish the significant and insignificant GMs by passing it through a nonlinear mapping. The design philosophy is that the significant GM corresponds to 1 and insignificant GM corresponds to 0. In this manner, the nonlinear mapping of the GM into the same dynamic range will largely facilitate the subsequent feature comparison process, such that the RR-IQA can be achieved in an efficient way by only comparing the significant and insignificant feature values. In general, the psychometric function with the sigmoid shape [48], [49] can be adopted to achieve such functionality. In this paper, we employ the Galton's ogive [50], which takes the form of a cumulative normal distribution function (CDF)

$$c(s) = \frac{1}{\sqrt{2\pi}\theta} \int_{-\infty}^{s} \exp\left[-\frac{(t-\tau)^2}{2\theta^2}\right] dt \tag{6}$$

where $c(s)$ denotes the detection probability density that distinguishes the significant and insignificant GMs. The input τ is the modulation threshold, s is the stimulus amplitude, and θ is the parameter that controls the slope of detection probability variation. In practical, θ is set as a constant value 0.05. Assuming the maximum gradient across the entire image to be

$$g_{\max} = \max\{\mathcal{G}(\mathbf{X})\} \tag{7}$$

and then τ is set as $\rho \cdot g_{\max}$ to adapt the characteristics of the input SCI. Here, ρ is chosen as 0.1 in practice.

Such a method has been widely adopted in the design of comparing two signals with different signal strengths [51]. After passing the GM through CDF, the significance map of the input SCI can be generated, which expresses useful information regarding the local primary information distribution across space. The significance map for the GM of the image $\mathbf{X}$ is denoted as $\mathcal{C}_G(\mathbf{X})$. In Fig. 3(k)–(o), the local significance maps are shown, and scrupulous observers may find that distortions can be detected when comparing the significance maps of the reference and distorted SCIs.

The uncertainty information is obtained by comparing the input with the orderly signals that are generated from the inference procedure of the brain. In the absence of any particular distortions, it is typically assumed that the input visual signal passes through the HVS channel before entering the brain. In the image perceiving process, the lens acts as a strong low-pass filter and high-frequency information may get lost [52]. Pamplona *et al.* [53] also demonstrated the strong low-pass filtering effect of the human eye using the power spectra. This inspires us to model the human visual perception process with low-pass filter. In particular, considering the characteristics of the HVS and SCIs, both Gaussian and motion low-pass filters are applied to quantify the uncertainty information [19]. The reasons of adopting the combination of such filters are manifold. First, the Gaussian filter is capable of achieving high contrast edge smoothing, and distortions around high contrast edges can be effectively reflected by comparing the input and Gaussian filtered SCIs [54]. Second, the motion blur is introduced to account for the viewing behavior of textual content, which relies on scanning in terms of eye movements to understand the information. Third, the combination of the uncertainty from Gaussian and motion blur channels characterizes how much uncertainty information will be generated during the phases of eye "fixation" and "saccade" [14], [15], [36].

Assuming the circular-symmetric Gaussian filter kernel to be

$$\mathbf{h}_g(i,j) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{i^2+j^2}{2\sigma^2}\right) \tag{8}$$

the Gaussian smoothed image is produced by convoluting it with the input image $\mathbf{X}$

$$\mathbf{X}_s = \mathbf{X} \otimes \mathbf{h}_g. \tag{9}$$

As such, the uncertainty information is computed by evaluating the similarity between $\mathbf{X}$ and $\mathbf{X}_s$. Here, we adopt the normalized version of gradient similarity [9], which holds the properties, such as symmetry, boundedness, and unique maximum

$$\mathcal{S}_G(\mathbf{X}) = f(\mathbf{X}, \mathbf{X}_s) = \frac{(\mathcal{G}(\mathbf{X}) - \mathcal{G}(\mathbf{X}_s))^2}{\mathcal{G}^2(\mathbf{X}) + \mathcal{G}^2(\mathbf{X}_s)}. \tag{10}$$

In general, S_G shall preserve the structure of the input SCI, such that distortions in the further comparison process can be well reflected. This requires the smoothing strength to be neither extremely strong (e.g., $\sigma > 10$) nor extremely weak (e.g., $\sigma < 1$). In Fig. 4, we demonstrate the distorted SCIs and their corresponding S_G maps using $\sigma = 1.0, 5.5, 11$. One can discern that the S_G map with $\sigma = 1.0$ cannot well preserve the original textual structure, and the S_G map with

Fig. 5. Illustration of the uncertainty maps from Gaussian blur and motion blur, where the top row shows the SCIs, the middle row shows the corresponding $\mathcal{S}_G$ maps, and the bottom row shows the corresponding $\mathcal{S}_M$ maps. (a), (d), and (g) Original. (b), (e), and (h) Gaussian noise. (c), (f), and (i) Gaussian blur.

$\sigma = 11$ may magnify the distortions. In this regard, here, the standard deviation σ is set to be 5.5 and the parameter sensitivities to the final quality prediction performance are evaluated in Section IV.

In analogy to the Gaussian blur, the motion blurred image $\mathbf{X}_m$ is generated by applying the motion blur convolution kernel $\mathbf{m}$ to the image $\mathbf{X}$. In particular, the motion blur convolution kernel is defined to be

$$\mathbf{m}(i,j) = \begin{cases} \dfrac{1}{t} & (i \cdot \sin\theta + j \cdot \cos\theta) = 0; \quad i^2 + j^2 \le \dfrac{t^2}{4} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where θ indicates the specific direction of motion and t denotes the amount of motion in pixels. As such, only the pixels along the motion direction are considered in the convolution process. The parameter values are empirically set as $t = 9$ and $\theta = 1$.

Again, the uncertainty information from motion blur channel is given by

$$\mathcal{S}_M(\mathbf{X}) = f(\mathbf{X}, \mathbf{X}_m) = \frac{(\mathcal{G}(\mathbf{X}) - \mathcal{G}(\mathbf{X}_m))^2}{\mathcal{G}^2(\mathbf{X}) + \mathcal{G}^2(\mathbf{X}_m)}. \tag{12}$$

In Fig. 5, we provide the SCIs and corresponding $\mathcal{S}_G$ and $\mathcal{S}_M$ maps. Two typical distorted SCIs are used for demonstration, which are Gaussian noised and Gaussian blurred versions of the original SCI. One can discern that the uncertainty maps $\mathcal{S}_G$ and $\mathcal{S}_M$ exhibit strong correlation with image distortion.

Consequently, the uncertainty information is computed by averaging $\mathcal{S}_G(\mathbf{X})$ and $\mathcal{S}_M(\mathbf{X})$ [19]

$$\mathcal{S}(\mathbf{X}) = \frac{\mathcal{S}_G(\mathbf{X}) + \mathcal{S}_M(\mathbf{X})}{2}. \tag{13}$$

Again, the uncertainty information is passed into the CDF to generate the significant and insignificant uncertainty, resulting the corresponding significance map $\mathcal{C}_S(\mathbf{X})$. Finally, the visual primary information and uncertainty are combined together for quality evaluation

$$\mathcal{Q}(\mathbf{X}) = \mathcal{C}_G(\mathbf{X}) \circ \mathcal{C}_S(\mathbf{X}) \tag{14}$$

where "$\circ$" denotes the Schur product [55] in terms of the entrywise operation. It ensures a local combination to reflect the quality over space.

The uncertainty information caused by the encountered surprise when perceiving the real scene is highly relevant to the "saliency" and "information content." In particular, for textual content, the high contrast edges that convey meaningful information and meanwhile produce high perceptual contrast are also obvious in the uncertainty map. Although the captured uncertainty may not be as accurate as specifically designed saliency prediction algorithms in predicting the saliency points, the concept of uncertainty lays a perceptually meaningful groundwork for saliency modeling [56]. In this manner, combining the maps of primary visual information and uncertainty can also be interpreted as using the "saliency" or "information content" to weight the primary visual information.

In analogy to the feature extraction for the original image $\mathbf{X}$, identical operation can be applied to the distorted image $\mathbf{Y}$ to obtain $\mathcal{Q}(\mathbf{Y})$. However, it is difficult to transmit the $\mathcal{Q}(\mathbf{X})$ directly to the receiver side for comparison, which may impose a high transmission burden in terms of the RR data rate. Therefore, to achieve a good compromise between prediction accuracy and the RR data rate, we establish the histogram that represents the distribution of $\mathcal{Q}$. In particular, the range of $\mathcal{Q}$ ($[r_{\min}, r_{\max}]$) is divided into n equal sized intervals, and the histogram bin corresponding to each interval is determined by the number of elements in set χ_i

$$h_i = |\chi_i|, \chi_i = \{x | \mathcal{Q}(x) \in I_i\} \tag{15}$$

where

$$I_i = \left[r_{\min} + \frac{(i-1) \cdot (r_{\max} - r_{\min})}{n}, r_{\min} + \frac{i \cdot (r_{\max} - r_{\min})}{n} \right). \tag{16}$$

As such, the established histogram for image $\mathbf{X}$ is determined by

$$H_{\mathbf{X}}(i) = \frac{h_i}{\sum_{j=1}^{n} h_j}. \tag{17}$$

In words, each bin value belonging to the histogram represents the corresponding probability of the interval. The established histogram for image $\mathbf{Y}$ [$H_{\mathbf{Y}}(i)$] is calculated in the same

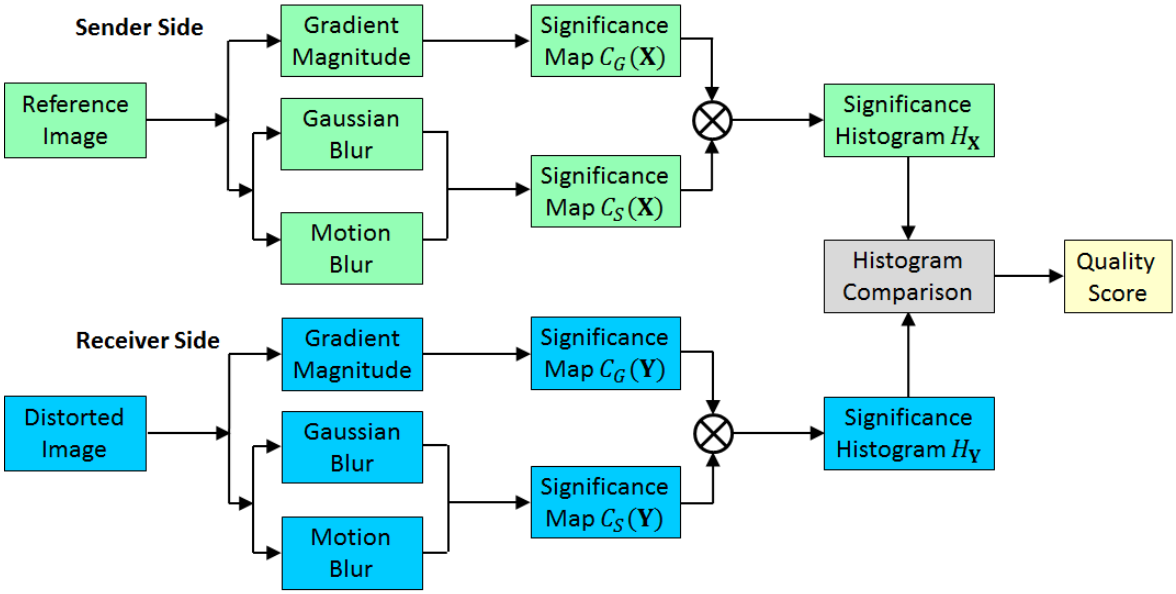


Fig. 6. Diagram of the proposed SCI RR-IQA model.

manner, and the final score of the proposed RR-IQA algorithm is computed by comparing the two histograms as

$$D(\mathbf{X},\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{|H_\mathbf{X}(i) - H_\mathbf{Y}(i)|}{H_\mathbf{X}(i) + H_\mathbf{Y}(i) + \varepsilon}\right) \tag{18}$$

where ε is introduced to avoid the instability when $H_\mathbf{X}(i) + H_\mathbf{Y}(i)$ is close to zero.

The diagram of the proposed method is shown in Fig. 6. Identical feature extractions of the reference and distorted SCIs are performed at the sender and the receiver sides, respectively. To extract the features, first, the GM map is created. Subsequently, the similarities between the input SCI and the Gaussian/motion blurred versions are computed to generate the uncertainty information. The significance maps of the GM and uncertainty information are combined together, and the histogram of the combined significance map is established to facilitate the feature transmission. At the receiver side, the two histograms are further compared to obtain the final quality score.

IV. Validations

A. Performance Evaluation on SIQAD

To validate the performance of the proposed algorithm, the screen IQA database (SIQAD) [18] that is collected from SCIs, including Web pages, slides, and Portable Document Format (PDFs), is employed. In particular, SIQAD includes 980 distorted images based on 20 references in total. Each reference image is distorted with seven distortion types at seven distortion levels. The distortion types include Gaussian noise, Gaussian blur, motion blur, contrast change, JPEG compression, JPEG2000 compression, and layer segmentation-based coding.

We compare the proposed method with both FR and RR IQA algorithms. The popular FR algorithms include SSIM [7], peak signal-to-noise ratio (PSNR), FSIM [8], VSI [11], GSIM [9], VIF [10], and visual signal-to-noise ratio (VSNR) [57]. The specifically designed screen content image perceptual quality assessment (SPQA) [18] is included in the comparison as well. It is also worth noting that the two versions of SSIM implementations are compared, which are denoted to be SSIM_1 [58] and SSIM_2 [59]. The difference between them lies in whether to employ the appropriate scale to preprocess the reference and distorted SCIs. Practical RR-IQA methods, such as the WNISM [25], DNT-RR [26], VIF-RR [29], Fourier transform based (FTB) [32], and SDM [33], are compared as well. As suggested in [32], the fifth score of FTB is employed ($Q^{(4)}_{\text{Phase}}$), which approximately requires 1/4096 of the reference image information. In the literature, these algorithms can efficiently achieve high accuracy prediction of natural image quality. Five evaluation metrics to assess the performance of IQA measures are reported, including Spearman rank correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC), mean absolute error (MAE), root mean-squared error (RMSE), and Kendall's rank correlation coefficient (KRCC). A better objective IQA measure should have higher PLCC, SRCC, and KRCC, while lower MAE and rms values.

A nonlinear mapping between objective and subjective scores is performed to compute PLCC. In particular, assuming the objective score to be r, the logistic regression function is applied to obtain the mapped score, which is given by [60] as

$$q(r) = \beta_1\left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(r-\beta_3)}}\right) + \beta_4 r + \beta_5 \tag{19}$$

where β_1–β_5 are the model parameters obtained numerically using a nonlinear regression process. Assuming that the total number of images is N, the mapped score for the ith image is q_i, and the corresponding subjective score is o_i, the PLCC is then computed to access the accuracy of the prediction

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q})\cdot(o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 \cdot \sum_i (o_i - \bar{o})^2}} \tag{20}$$

TABLE I
PERFORMANCE COMPARISONS WITH RR AND FR ALGORITHMS

SIQAD	PSNR	$SSIM_1$	$SSIM_2$	VIF	SPQA	FSIM	VSI	GSIM	VSNR	DNT-RR	VIF-RR	WNISM	FTB	SDM	**Proposed**
SRCC	0.5608	0.5836	0.7566	0.8069	0.8416	0.5819	0.5381	0.5483	0.5703	0.5054	0.6082	0.5188	0.4575	0.6020	**0.7655**
PLCC	0.5869	0.5912	0.7561	0.8206	0.8584	0.5902	0.5568	0.5686	0.5966	0.5291	0.5758	0.5857	0.4691	0.6034	**0.8014**
KRCC	0.4226	0.4235	0.5583	0.6082	0.6591	0.4250	0.3874	0.4054	0.4381	0.3615	0.4431	0.3540	0.3268	0.4322	**0.5756**
RMSE	11.590	11.545	9.3676	8.1795	7.3421	11.555	11.890	11.775	11.487	12.147	11.703	11.602	12.641	11.414	**8.5620**
MAE	9.0393	9.0934	7.3133	6.5261	5.7213	9.0116	9.2875	9.1663	8.8284	9.7913	9.5197	9.4566	10.132	9.0139	**6.8021**

TABLE II
STATISTICAL SIGNIFICANCE EVALUATION BASED ON PLCC

IQA Model	PSNR	$SSIM_1$	$SSIM_2$	VIF	SPQA	FSIM	VSI	GSIM	VSNR	DNT-RR	VIF-RR	WNISM	FTB	SDM
$\alpha = 0.05$	1	1	1	-	0	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	1	1	-	-	0	1	1	1	1	1	1	1	1	1

where $\bar{q}$ and $\bar{o}$ represent the mean score of q and o over the test set.

After converting the objective scores, MAE and RMSE are calculated to measure the prediction accuracy, which are given by

$$\text{MAE} = \frac{1}{N}\sum |q_i - o_i|$$
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum (q_i - o_i)^2}. \tag{21}$$

SRCC is employed to evaluate the prediction monotonicity. Assuming that the difference between the ith image's ranks in subjective and objective evaluations is υ_i, it is defined as

$$\text{SRCC} = 1 - \frac{6\sum_{i=1}^{N}\upsilon_i^2}{N(N^2-1)}. \tag{22}$$

Moreover, the KRCC is given by

$$\text{KRCC} = \frac{2(N_c - N_d)}{N(N-1)} \tag{23}$$

where N_c and N_d denote the number of concordant and discordant pairs in the database, respectively.

Here, all the 980 distorted images are included in the evaluation. The test results are given in Table I. To reduce the transmission overhead, the number of bins n is set to be 5, indicating that only four feature values need to be transmitted, as the sum of the probability in all of the bins equals to unity. In particular, we quantize each feature into 12 b, and in total, the additional overhead for each SCI is only 48 b. As such, the number of the overhead bits is among the lowest in the RR-IQA methods. To the best of our knowledge, only the VIF-RR method requires less transmission bits than our approach, which costs 30 b to transmit the features. From Table I, it can be seen that the proposed scheme clearly outperforms the RR-IQA algorithms with the state-of-the-art performance. Since FR and RR-IQA algorithms are applied in different scenarios and RR algorithms require much less information than the FR methods, it is usually unfair to directly compare the RR method with the FR method. However, the FR-IQA algorithms can still supply us useful references on the current status of the SCI IQA research. It is also interesting to find that the proposed method outperforms most of the FR-IQA algorithms, and is only inferior to VIF and SPQA. The performance improvements originate from the design philosophy, in which the characteristics of SCIs are considered.

B. Statistical Significance Analyses

Following the standardized procedures in [61], we carry out the statistical significant analyses to obtain the significance of the difference in terms of the PLCC and RMSE based on the statistical hypothesis testing, respectively. The motivation of conducting these evaluations is to know whether the confidence in the estimation of the proposed algorithm's performance allows us to draw the statistically sound conclusion of superiority or inferiority compared with the state-of-the-art methods.

The significance of the difference between the correlation coefficients is obtained by performing Fisher's z-transformation to convert the PLCC into the normally distributed variable [62]. The hypothesis testing is performed by hypothesizing that there is no significant difference between the proposed and one of the compared methods. Two significance levels are applied, and the corresponding α values are set to be 0.05 and 0.01, respectively. The results are demonstrated in Table II, where a symbol "−" indicates that two IQA methods are statistically indistinguishable with each other, "1" denotes that the proposed IQA method is statistically better than the corresponding one in the column, and "0" denotes that the IQA method of the column is better than the proposed one. It can be observed that the proposed method is statistically superior to all RR-IQA algorithms, inferior to SPQA, and indistinguishable with the VIF method when $\alpha = 0.05$. Moreover, when the α value is lowered down to 0.01 to decrease the probability of type I error, the proposed method is also statistically indistinguishable with $SSIM_2$.

Subsequently, following the method described in [61], we perform the statistical significance analyses on the difference between RMSE. In particular, the F-distribution is employed to compare whether the two methods are statistically significantly different. Again, we use the same notations as in PLCC significance test, and the results are provided in Table III. It can be observed that the proposed model is statistically superior to the RR IQA algorithms. Compared with the FR-IQA algorithms, for both $\alpha = 0.01$ and $\alpha = 0.05$, the proposed method is statistically superior to most of the FR-IQA methods, inferior to SPQA, and indistinguishable with VIF.

TABLE III

STATISTICAL SIGNIFICANCE EVALUATION BASED ON RMSE

IQA Model	PSNR	$SSIM_1$	$SSIM_2$	VIF	SPQA	FSIM	VSI	GSIM	VSNR	DNT-RR	VIF-RR	WNISM	FTB	SDM
$\alpha = 0.05$	1	1	1	-	0	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	1	1	1	-	0	1	1	1	1	1	1	1	1	1

TABLE IV

DISTORTION TYPE BREAKDOWN FOR PLCC AND SRCC COMPARISONS

	Gaussian Noise		Gaussian Blur		Motion Blur		Contrast Change		JPEG		JPEG2000		Layer Coding	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
DNT-RR	0.8189	0.8211	**0.8946**	**0.8875**	**0.7928**	**0.7903**	**0.7846**	**0.6719**	0.4697	0.4328	0.6060	0.5849	0.5585	0.5555
VIF-RR	0.8657	0.8479	**0.8830**	**0.8715**	0.7350	0.7214	0.7570	0.6493	0.6912	0.6803	**0.7647**	**0.7588**	**0.7321**	**0.7347**
WNISM	0.8570	0.8442	0.8524	0.8370	0.6618	0.6606	0.7402	0.6142	0.2627	0.1742	0.3543	0.2810	0.2551	0.1880
FTB	0.7185	0.7165	0.7358	0.7400	0.5984	0.5866	0.5207	0.1112	0.5696	0.5474	0.5122	0.5155	0.5498	0.5213
SDM	**0.8694**	**0.8635**	0.7836	0.8199	0.5434	0.5307	**0.7831**	**0.6617**	**0.7203**	**0.7331**	0.6635	0.6292	0.7092	**0.7496**
Proposed	**0.8798**	**0.8664**	0.8810	0.8715	**0.8465**	**0.8434**	0.6812	0.5291	**0.7638**	**0.7605**	**0.6807**	**0.6617**	**0.7110**	0.7116

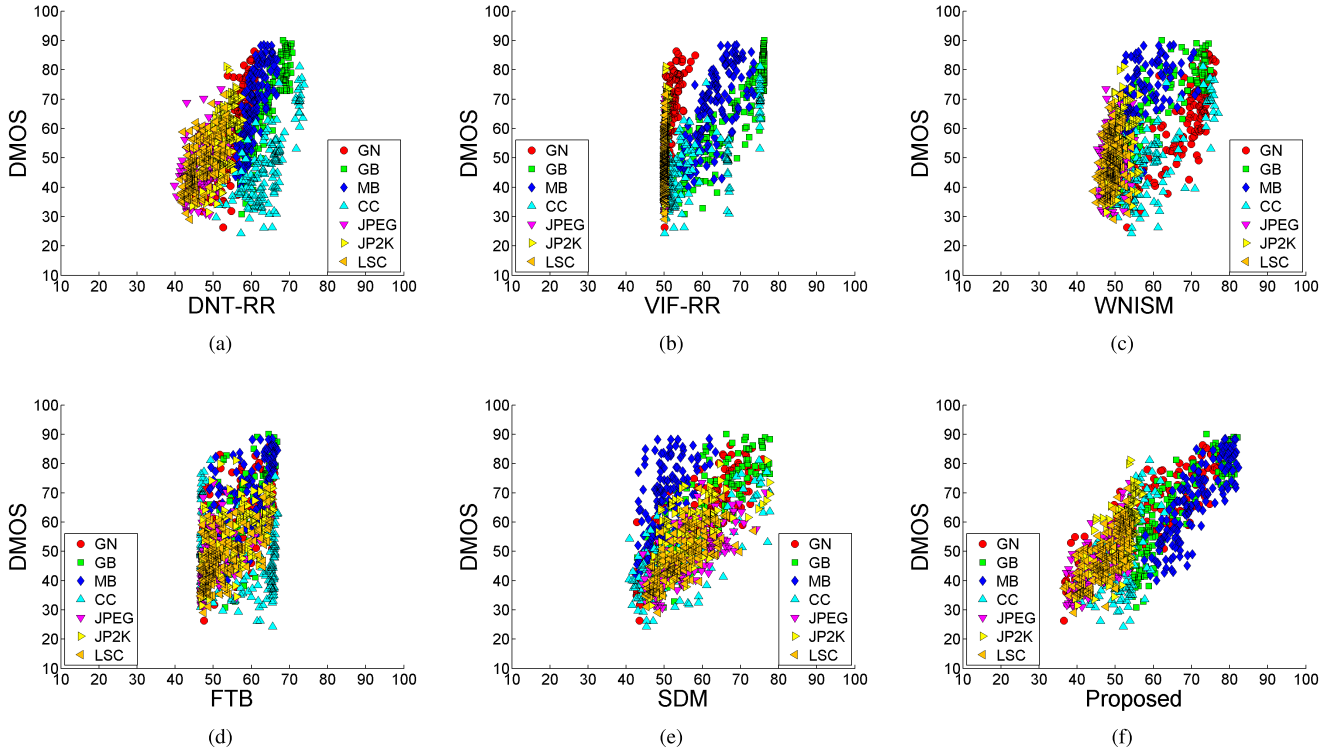


Fig. 7. Scatter plots of the difference mean opinion score versus objective RR-IQA scores (after nonlinear regression) for all SCIs. (a) DNT-RR, (b) VIF-RR, (c) WNISM, (d) FTB, (e) SDM, (f) Proposed.

C. Performance Comparison on Individual Distortion Types

In this section, the breakdown prediction performance is examined for individual distortion types. The performance is provided in Table IV. It can be observed that in most of the cases, the proposed method is among the best from the perspectives of prediction accuracy and monotonicity. The scatter plots between human ratings and the objective scores after nonlinear regression are demonstrated in Fig. 7. Different colors are used for different distortion types. One can discern that the proposed method has stronger ability in cross-distortion[1] quality prediction, which further verifies the robustness and efficiency of the algorithm. Moreover, the scatter plots for the compressed SCIs (JPEG, JPEG2000, and layer segmentation-based coding) are shown in Fig. 8, which suggest that the proposed algorithm is capable of delivering trusted quality prediction scores on the evaluation of compression artifacts.

[1]"Cross-distortion quality prediction" indicates that the IQA measure is able to achieve good prediction performance when different distortion types are involved. It requires the IQA model to be general and flexible to handle broader types of distortions.

D. Performance Comparison on Individual Content Types

In SIQAD, the 20 source SCIs can be further divided into three types from the perspective of application scenario: Web page, slides, and PDF files (digital magazines). The SCIs belonging to each content type are shown in Fig. 9. Moreover, the IQA performance is further examined by comparing the proposed method with the state-of-the-art RR-IQA algorithms. The results are illustrated in Table V, and we can see that

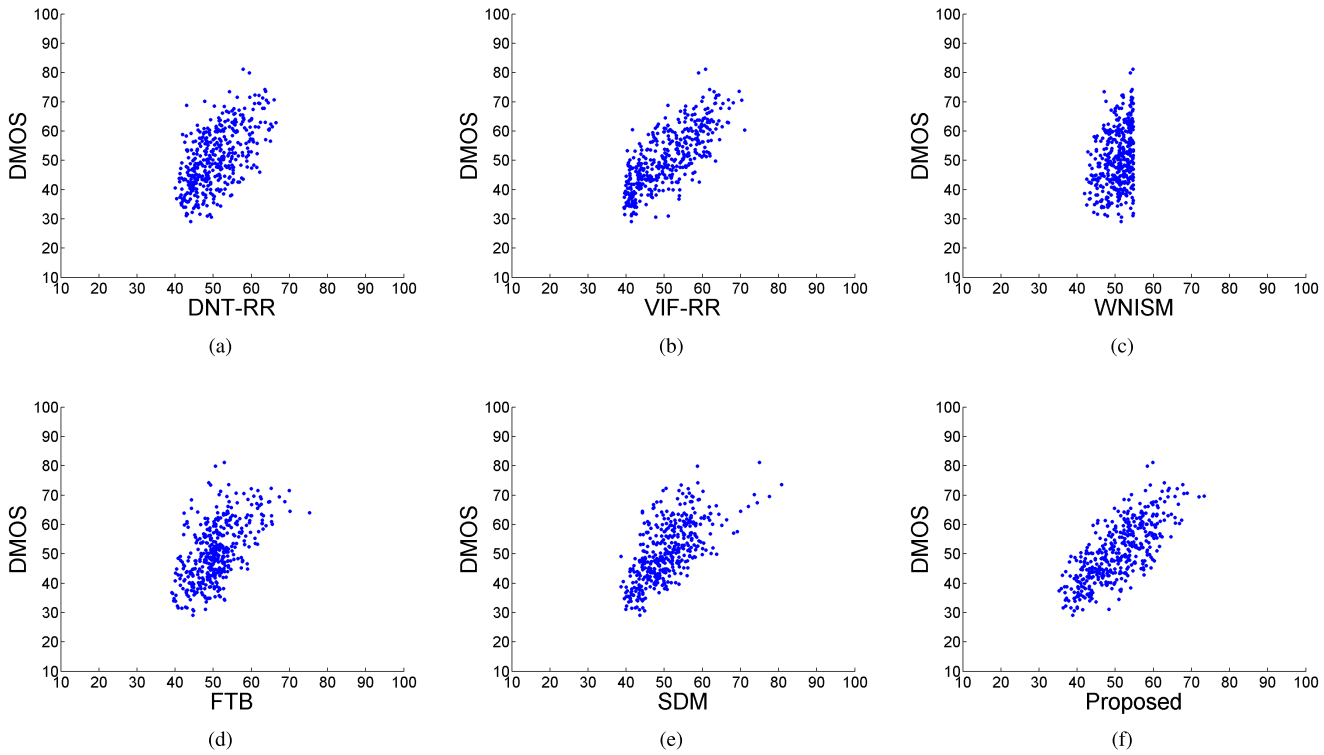


Fig. 8. Scatter plots of the DMOS versus objective RR-IQA scores (after nonlinear regression) for compressed SCIs (including JPEG, JPEG2000, and layer segmentation-based coding). (a) DNT-RR, (b) VIF-RR, (c) WNISM, (d) FTB, (e) SDM, (f) Proposed.

Fig. 9. SCIs belonging to different content types. Red rectangular: Web page. Green rectangular: PDF. Blue rectangular: slides.

TABLE V
PERFORMANCE COMPARISONS WITH RR-IQA ALGORITHMS ON INDIVIDUAL CONTENT TYPES

SCI		DNT-RR	VIF-RR	WNISM	FTB	SDM	Proposed
Webpage	SRCC	0.5353	0.6579	0.5556	0.4592	0.5301	**0.8411**
	PLCC	0.5606	0.6186	0.6008	0.4647	0.5528	**0.8619**
	KRCC	0.3888	0.4892	0.3846	0.3257	0.3838	**0.6521**
	RMSE	11.6564	11.0605	11.2526	12.4642	11.7301	**7.1387**
	MAE	9.4685	9.1471	9.0649	10.0879	9.2630	**5.7670**
Slides	SRCC	0.4703	0.5993	0.5772	0.4883	0.7042	**0.7427**
	PLCC	0.5234	0.6488	0.6862	0.5396	0.7112	**0.7997**
	KRCC	0.3365	0.4437	0.4045	0.3721	0.5082	**0.5659**
	RMSE	12.2494	10.9388	10.4570	12.1029	10.1056	**8.6315**
	MAE	9.9742	8.4214	8.6821	9.6359	7.8479	**7.0576**
PDF	SRCC	0.4645	0.5578	0.4950	0.4459	0.6291	**0.7064**
	PLCC	0.4892	0.5673	0.5884	0.4695	0.6249	**0.7514**
	KRCC	0.3289	0.3999	0.3343	0.3255	0.4499	**0.5181**
	RMSE	12.3638	11.6737	11.4618	12.5161	11.0666	**9.3532**
	MAE	9.9008	9.3657	9.4657	9.8397	8.8083	**7.3986**

the proposed algorithm can achieve the best quality prediction performance for all of the three content types.

E. Parameter Sensitivity Evaluation

Generally speaking, a trusted quality measure should be able to tolerate small parameter value changes. Therefore, we conduct an experiment to investigate the impact of the parameters used in Gaussian and motion blur on the quality prediction performance. First, the standard deviation in the Gaussian smooth kernel σ is examined by varying it from 3.5 to 7.5, and the results are tabulated in Table VI. We can observe that the performance of our IQA measure is barely affected.

TABLE VI

PARAMETER SENSITIVITY TESTING WITH THE VARIATION OF σ

σ	3.5	4.5	**5.5**	6.5	7.5
SRCC	0.7617	0.7649	**0.7655**	0.7653	0.7651
PLCC	0.7972	0.8008	**0.8014**	0.8013	0.8014
KRCC	0.5705	0.5746	**0.5756**	0.5750	0.5749
RMSE	8.6421	8.5729	**8.5620**	8.5635	8.5621
MAE	6.9130	6.8299	**6.8021**	6.8235	6.8004

TABLE VII

PARAMETER SENSITIVITY TESTING WITH THE VARIATION OF t

t	7	8	**9**	10	11
SRCC	0.7531	0.7567	**0.7655**	0.7658	0.7625
PLCC	0.7874	0.7943	**0.8014**	0.8041	0.8015
KRCC	0.5612	0.5664	**0.5756**	0.5770	0.5721
RMSE	8.8246	8.6958	**8.5620**	8.5089	8.5600
MAE	7.0773	6.9795	**6.8021**	6.7546	6.8236

TABLE VIII

PARAMETER SENSITIVITY TESTING WITH THE VARIATION OF θ

θ	-3	-1	**1**	3	5
SRCC	0.7666	0.7629	**0.7655**	0.7665	0.7667
PLCC	0.8023	0.7993	**0.8014**	0.8028	0.8025
KRCC	0.5759	0.5718	**0.5756**	0.5765	0.5766
RMSE	8.5452	8.6026	**8.5620**	8.5350	8.5408
MAE	6.8349	6.8813	**6.8021**	6.8295	6.8198

Subsequently, the parameters used in the motion blur are examined by changing t from 7 to 11 and θ from -3 to 5. From Tables VII and VIII, it can be seen that the proposed method is able to achieve considerably stable performance. These results further provide useful evidence that the proposed method is robust and tolerant to the varying parameter values used in Gaussian and motion blur.

F. Performance Evaluation on Transmission Loss

Furthermore, we evaluate the performance of the proposed method using the distorted SCIs degraded by the transmission loss. In particular, two data sets are built together with the subjective testing results. Distortion types in the two data sets are JPEG and H.264/Advanced Video Coding (AVC) transmission errors. The 20 reference SCIs in SIQAD are employed. To create the distorted versions of these SCIs, each SCI is first compressed by the two codecs, JPEG and H.264/AVC, and then the coding blocks are randomly discarded. The mean pixel values from surrounding blocks are used to infer the discarded block. To be consistent with the compression process, the discarded block sizes are 8×8 and 16×16 for JPEG and H.264/AVC, respectively. Four distortion levels depending on the probability of the transmission loss are included, from 10% to 40% with an interval of 10%. In total, 80 distorted SCIs are generated from the 20 reference SCIs for each data set. Twenty human subjects were invited to participate. In particular, they were asked to view the distorted SCIs with a viewing distance around 2–2.5 screen heights, and ten-category discrete scale was employed to record the subjective opinions. This process is in consistent with that in developing the SIQAD. After collecting the raw scores, the average values are calculated to generate the final MOS for each distorted SCI.

TABLE IX

PERFORMANCE COMPARISONS WITH STATE-OF-THE-ART RR-IQA ALGORITHMS FOR JPEG TRANSMISSION LOSS

	RRDNT	VIR-RR	WNISM	FTB	SDM	Proposed
SRCC	0.8550	0.9016	0.6693	0.6343	0.8496	0.9171
PLCC	0.8687	0.9328	0.6737	0.6489	0.8503	0.9343
KRCC	0.6561	0.7069	0.4978	0.4533	0.6529	0.7565
RMSE	0.8153	0.5931	1.2166	1.2526	0.8664	0.5870
MAE	0.6580	0.4824	0.9413	1.0442	0.6922	0.4300

TABLE X

PERFORMANCE COMPARISONS WITH STATE-OF-THE-ART RR-IQA ALGORITHMS FOR H.264/AVC TRANSMISSION LOSS

	RRDNT	VIR-RR	WNISM	FTB	SDM	Proposed
SRCC	0.8670	0.9114	0.4640	0.5756	0.8802	0.9488
PLCC	0.8934	0.9545	0.4913	0.6355	0.8541	0.9608
KRCC	0.6535	0.7077	0.3083	0.4089	0.7000	0.8070
RMSE	0.6765	0.4493	1.3118	1.1628	0.7832	0.4176
MAE	0.5188	0.3583	1.0954	0.9126	0.6532	0.3135

The experimental results compared with the state-of-the-art algorithms are tabulated in Tables IX and X, from which we can observe that the proposed algorithm achieves superior performance in terms of both prediction accuracy and monotonicity. These results further demonstrate strong quality prediction capability of the proposed method for SCI distortions.

G. Complexity Comparison

Table XI tabulates the execution time of different IQA methods. In particular, these IQA methods are run on the SIQAD. The testing environments are Intel I7-4790 CPU@3.60 GHz, 8-GB random access memory, and MATLAB R2014 platform. The average running time is recorded. For the RR-IQA methods, the computation of feature extraction at both the sender and receiver sides, and the feature comparison operations are included. One can discern that that the computational complexity of the proposed model is among the lowest in the compared IQA methods. Moreover, compared with the other RR-IQA algorithms, our method significantly saves the computational time, which enables its applications in real scenarios.

H. Comparisons of the RR Data Rate to the SCI Data Rate

We have conducted an experiment to compare the data rate of transmitting the features to that of transmitting the SCI. As explained in Section IV-A, the RR data rate (R_{RR}) used to transmitted the features is 48 b per picture. Assuming that the coding bits of each SCI are R_{SCI}, then the percentage of the data rate used in transmitting the feature information is computed by

$$P_{\mathrm{RR}} = \frac{R_{\mathrm{RR}}}{R_{\mathrm{SCI}} + R_{\mathrm{RR}}}. \tag{24}$$

In practice, we use the High Efficiency Video Coding screen content coding extension codec (HM $-$ 15.0 $+$ REXt $-$ 8.0 $+$ SCM $-$ 2.0rc1) to compress each reference SCI in SIQAD

TABLE XI
RUNNING TIME OF DIFFERENT IQA METHODS

IQA Model	PSNR	SSIM	FSIM	VSI	GSIM	VSNR	DNT-RR	VIF-RR	WNISM	FTB	SDM	Proposed
Time (s)	0.002	0.016	0.279	0.156	0.020	0.240	4.314	12.731	1.867	0.531	0.391	0.197

TABLE XII
PERCENTAGE OF THE BIT RATE P_{RR} USED IN TRANSMITTING THE FEATURE INFORMATION FOR EACH SCI IN SIQAD

CIM1			CIM2			CIM3			CIM4		
QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40
0.008%	0.016%	0.036%	0.008%	0.013%	0.026%	0.008%	0.013%	0.031%	0.008%	0.014%	0.029%
CIM5			CIM6			CIM7			CIM8		
QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40
0.009%	0.015%	0.029%	0.006%	0.011%	0.027%	0.012%	0.021%	0.037%	0.010%	0.021%	0.050%
CIM9			CIM10			CIM11			CIM12		
QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40
0.006%	0.012%	0.032%	0.008%	0.014%	0.031%	0.006%	0.013%	0.031%	0.005%	0.009%	0.021%
CIM13			CIM14			CIM15			CIM16		
QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40
0.009%	0.017%	0.037%	0.007%	0.013%	0.035%	0.007%	0.012%	0.024%	0.006%	0.012%	0.030%
CIM17			CIM18			CIM19			CIM20		
QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40	QP=20	QP=30	QP=40
0.011%	0.022%	0.050%	0.006%	0.012%	0.025%	0.006%	0.012%	0.029%	0.008%	0.016%	0.036%

(CIM1∼CIM20) at three QP points (QP = 20, 30, 40), ranging from high bit rate to low bit rate coding. Advanced coding tools that have been specifically developed for SCI compression are enabled, such as intra-block copy and palette mode [63]. The results are shown in Table XII, from which we can observe that P_{RR} is 0.02% on average. As such, the RR data rate can be regarded as negligible in the transmission of SCI stream.

I. Discussion

Practically, deploying the SCI RR quality assessment method in the interactive screen-remoting system requires the algorithm to be both effective and efficient, especially in the trend of increasing proliferation of high-volume screen visualization data for the purpose of real-time quality monitoring and high-fidelity display maintaining. Though it is difficult to simultaneously achieve both of them, we can observe that our model is able to achieve a good compromise from the validations. In particular, the proposed method outperforms the state-of-the-art methods for the overall database and most of the individual distortion types, while the transmission overhead is only 48 b/pic and the computational complexity is among the lowest.

As one of the first attempts on this topic, the proposed method also has several limitations that should be improved in the future. First, the current method is applicable to the SCI. In practice, how to extend it to screen content video should be further investigated. In particular, the high correlation among screen video frames may further reduce the transmission and computational overhead of the extracted features by exploiting the inter-prediction like techniques for feature prediction. Second, it is worth mentioning that the design of RR method does not make any assumption on the image distortion types, making it have the potential to be used for general-purpose applications. However, currently, the testing is based on eight distortion types. In the future, more distortion types will be involved for verification as well. Moreover, statistical features that exhibit more robust cross distortion type prediction will be studied in the future. Third, the methodology of selecting the model parameters from the functionalities of HVS is worth further exploring. Though the investigation of viewing behaviors for the particular screen content is still at the starting stage, the fundamentally interesting differences between SCIs and natural images from the perspective of psychological studies may bring more inspirations to the model parameter selection in the future. Finally, how to design screen content enhancement algorithm using the statistical features from the proposed RR-IQA algorithm is a topic worth further investigating. This poses new challenges to IQA and restoration research for SCIs and opens up new space for future exploration.

V. CONCLUSION

We have specifically developed an RR-IQA model that automatically predicts the quality of SCIs. Statistical features obtained from the primary visual information and the amount of uncertainty are combined in an efficient way to reflect the perceived quality. The RR data rate (48 b/pic) is negligible compared with the compressed SCI bitstream, and the computational complexity is among the lowest in the state-of-the-art IQA algorithms. Experimental results show that the proposed method is well correlated with subjective evaluations of SCI quality, suggesting that it is promising at handling computer generated unnatural images.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and anonymous reviewers for their valuable comments, which significantly helped them to improve the quality of this paper.

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] Z. Wang, K. Zeng, A. Rehman, H. Yeganeh, and S. Wang, "Objective video presentation QoE predictor for smart adaptive video streaming," *Proc. SPIE*, vol. 9599, p. 95990Y, Sep. 2015.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[8] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[9] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[10] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[11] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[12] 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.

[13] E. C. Potter, B. Wyble, C. E. Hagmann, and E. S. McCourt, "Detecting meaning in RSVP at 13 ms per picture," *Attention, Perception, Psychophys.*, vol. 76, no. 2, pp. 270–279, Feb. 2014.

[14] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends Cognit. Sci.*, vol. 7, no. 11, pp. 498–504, 2003.

[15] A. Thiele, P. Henning, M. Kubischik, and K.-P. Hoffmann, "Neural mechanisms of saccadic suppression," *Science*, vol. 295, no. 5564, pp. 2460–2462, 2002.

[16] C. Shen and Q. Zhao, "Webpage saliency," in *Proc. ECCV*, 2014, pp. 33–46.

[17] P. Faraday, "Visually critiquing Web pages," in *Multimedia*. Springer, 2000, pp. 155–166.

[18] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.

[19] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[20] 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.*, [Online]. Available: http://ieeexplore.ieee.org/document/7478441/

[21] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.

[22] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, "No-reference perceptual image quality metric using gradient profiles for JPEG2000," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 502–516, Aug. 2010.

[23] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[24] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[25] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.

[26] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[27] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

[28] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[29] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[30] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 10, pp. 1441–1456, Oct. 2012.

[31] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[32] M. Narwaria, W. Lin, I. V. McLoughlin, S. Emmanuel, and L. T. Chia, "Fourier transform-based scalable image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364–3377, Aug. 2012.

[33] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 1095–1098.

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

[35] D. C. Knill and A. Pouget, "The Bayesian brain: The role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, no. 12, pp. 712–719, 2004.

[36] 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.

[37] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

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

[39] 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.

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

[41] 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, Aug. 2016.

[42] R. Grier, P. Kortum, and J. Miller, "How users view Web pages: An exploration of cognitive and perceptual mechanisms," in *Human Computer Interaction Research in Web Design and Evaluation*. Hershey, PA, USA: Information Science, 2007, pp. 22–41.

[43] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "Objective video quality assessment system based on human perception," *Proc. SPIE*, vol. 1913, pp. 15–26, Sep. 1993.

[44] J. Wu, F. Qi, and G. Shi, "Self-similarity based structural regularity for just noticeable difference estimation," *J. Vis. Commun. Image Represent.*, vol. 23, no. 6, pp. 845–852, 2012.

[45] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 886–893.

[46] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[47] B. Jähne, H. Haussecker, and P. Geissler, Eds., *Handbook of Computer Vision and Applications*, vol. 2. New York, NY, USA: Academic, 1999.

[48] J. P. Guilford, *Psychometric Methods* (McGraw-Hill Series in Psychology), 2nd ed. New York, NY, USA: McGraw-Hill, Dec. 1954.

[49] Y. Le Grand, *Light, Colour and Vision*. London, U.K.: Chapman & Hall, 1957.

[50] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, vol. 72. Bellingham, WA, USA: SPIE, 1999.

[51] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[52] R. G. Boothe, *Perception of the Visual Environment*. New York, NY, USA: Springer, 2001.

[53] D. Pamplona, J. Triesch, and C. A. Rothkopf, "Power spectra of the natural input to the visual system," *Vis. Res.*, vol. 83, pp. 66–75, May 2013.

[54] G. Zhai, A. Kaup, J. Wang, and X. Yang, "Retina model inspired image quality assessment," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Nov. 2013, pp. 1–6.

[55] C. Davis, "The norm of the Schur product operation," *Numer. Math.*, vol. 4, no. 1, pp. 343–344, 1962.
[56] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Visual saliency detection with free energy theory," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1552–1555, Oct. 2015.
[57] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[58] Z. Wang. *SSIM Index Implementation Version 1*, accessed Jun. 2016. [Online]. Available: https://ece.uwaterloo.ca/~z70wang/research/ssim/ssim.m
[59] Z. Wang. *SSIM Index Implementation Version 2*, accessed Jun. 2016. [Online]. Available: https://ece.uwaterloo.ca/~z70wang/research/ssim/ssim_index.m
[60] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[61] *Series P: Terminals and Subjective and Objective Assessment Methods*, document ITU-T P.1401, ITU-T, 2012.
[62] M. R. Spiegel, *Schaum's Outline of Theory and Problems of Statistics*. New York, NY, USA: McGraw-Hill, 1961.
[63] 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.

Shiqi Wang (M'15) received the B.S. degree in computer science from Harbin Institute of Technology, Harbin, China, in 2008 and the Ph.D. degree in computer application technology from Peking University, Beijing, China, in 2014.

He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. His research interests include visual quality assessment and compression.

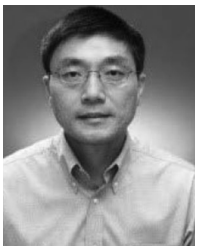
Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively.

He is with Nanyang Technological University, Singapore, as a Research Fellow. His research interests include quality assessment, contrast enhancement, and visual saliency detection.

Dr. Gu is a Special Session Organizer for VCIP 2016.

Xinfeng Zhang received the B.S. degree in computer science from Hebei University of Technology, Tianjin, China, in 2007 and the Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014.

He is a Research Fellow with Nanyang Technological University, Singapore. His research interests include image and video processing, and image and video compression.



Weisi Lin (SM'98–F'16) received the Ph.D. degree from Kings College, London University, London, U.K., in 1993.

He was the Laboratory Head of Visual Processing, Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He has authored over 300 scholarly publications and received over 4 million dollars in research grant funding. He holds seven patents. 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.

Dr. Lin has served as an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE SIGNAL PROCESSING LETTERS, and *Journal of Visual Communication and Image Representation*. He is also on six IEEE Technical Committees and Technical Program Committees of a number of international conferences.

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-doctorate position with University of Southern California, Los Angeles, CA, USA. He then 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, Bunkyō, Japan, in 1991.

He was a Professor of Computer Science with Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of Computer Science with Peking University, Beijing. 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.

Dr. Gao served or serves on the Editorial Board for several journals, such as IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, *Journal of Image Communications* (EURASIP), and *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as IEEE ICME and ACM Multimedia, and also was on the advisory and technical committees of numerous professional organizations.