# No-Reference Quality Assessment of Screen Content Pictures

Ke Gu, Jun Zhou, *Member, IEEE*, Jun-Fei Qiao, *Member, IEEE*, Guangtao Zhai, *Member, IEEE*,
Weisi Lin, *Fellow, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

*Abstract*—**Recent years have witnessed a growing number of image and video centric applications on mobile, vehicular, and cloud platforms, involving a wide variety of digital screen content images. Unlike natural scene images captured with modern high fidelity cameras, screen content images are typically composed of fewer colors, simpler shapes, and a larger frequency of thin lines. In this paper, we develop a novel blind/no-reference (NR) model for accessing the perceptual quality of screen content pictures with big data learning. The new model extracts four types of features descriptive of the picture complexity, of screen content statistics, of global brightness quality, and of the sharpness of details. Comparative experiments verify the efficacy of the new model as compared with existing relevant blind picture quality assessment algorithms applied on screen content image databases. A regression module is trained on a considerable number of training samples labeled with objective visual quality predictions delivered by a high-performance full-reference method designed for screen content image quality assessment (IQA). This results in an opinion-unaware NR blind screen content IQA algorithm. Our proposed model delivers computational efficiency and promising performance. The source code of the new model will be available at: https://sites.google.com/site/guke198701/publications.**

*Index Terms*—**Screen content image, image quality assessment (IQA), no-reference (NR), opinion-unaware (OU), scene statistics model, hybrid filter, image complexity description, big data.**

## I. INTRODUCTION

S CREEN content pictures have become quite common over the last several years. Numerous consumer applications, such as online gaming, mobile web browsing, vehicle navigation, remote control, cloud computing and more, involve computer-generated screen content images. Figures 1(a)-(b)

K. Gu and J.-F. Qiao are with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn; junfeiq@bjut.edu.cn).

J. Zhou and G. Zhai are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn; zhoujun@sjtu.edu.cn).

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

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
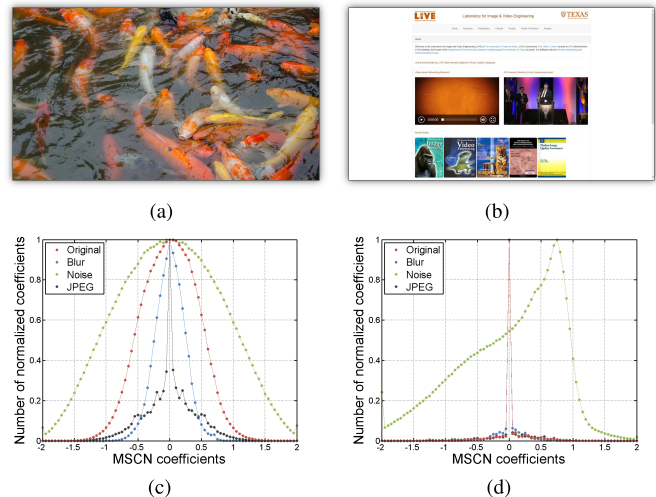
Fig. 1. Small Comparison of naturalistic and screen content images: (a)-(b) original pristine natural and screen content images; (c)-(d) histograms of MSCN coefficients of pristine and distorted versions of (a)-(b) corrupted by Gaussian blur, additive noise and JPEG compression.

depict two typical images, one of a natural scene and the other of a screen content scene, captured using a digital camera and a screenshot tool, respectively. There are important differences between camera-captured images of natural scenes and computer-generated screen content images. Natural scene images have rich and complex distributions of luminance and color that are governed by statistical laws, while screen content images generally contain fewer and simpler luminance and color variation and structures.

The study of screen content image quality is a new and interesting topic. In [1], Gu *et al.* conducted a performance comparison of mainstream Full-Reference (FR) Image Quality Assessment (IQA) methods on screen content image databases, including NQM [2], SSIM [3], MS-SSIM [4], VSNR [5], FSIM [6], GSI [7], GMSD [8], LTG [9], and VSI [10]. *Full reference* refers to the situation where a reference image is available when predicting quality. Their results implied that existing FR metrics, despite attaining superior performance when evaluating the quality of natural scene images, fail on the screen content IQA problem.

Similar problems are encountered using blind/*No-Reference* (NR) IQA models. Motivated by well-known *Natural Scene Statistics* (NSS) models [11], a variety of blind picture quality models [12], including BLIINDS-II [13], BRISQUE [14], C-DIIVINE [15], LPSI [16], NIQE [17] and IL-NIQE [18], have been developed. *No reference* refers to the situation where no information contained in any reference

image is used to infer quality. Unfortunately, all were found to work ineffectively on the screen content IQA problem [19]. Two main reasons for this are that NSS models fail when 1) pristine natural scene images are contaminated; 2) images are of computer graphic or document contents, not resulting from a natural source. To offer a more straightforward illustration, we applied the decorrelating method of [14] on the two images in Figs. 1(a)-(b) by computing their *Mean Subtracted Contrast Normalized* (MSCN) coefficients and plotting their histograms in Figs. 1(c)-(d). Clearly, the MSCN coefficients of the pristine natural scene image nicely follow the NSS model; that is, the histogram of MSCN coefficients exhibits a Gaussian-like appearance [11]. By contrast, the undistorted screen content image yields a quite different Laplacian-like MSCN distribution.

We plotted the empirical probability density functions (histograms) of distorted versions of the natural scene and screen content images in Figs. 1(c)-(d). Three distortions: Gaussian blur, additive noise and JPEG compression, were applied to the original natural scene and screen content images. From Fig. 1(c), it may be observed that as revealed in [11], each type of distortion changes the distribution of the MSCN coefficients in a particular way; for example, blur distortion narrows the histogram towards a Laplacian-like distribution, whereas additive white noise widens the histogram. However, as shown in Fig. 1(d), distortions such as blur and blockiness may not affect the statistical distribution as compared with that of the undistorted screen content image. This dichotomy is increased when the original image is less naturalistic, viz., contains more artificial content, such as text, and less photographic content. Further, the distribution is transduced into an odd shape when additive noise is injected. Thus NSS models appear to be inadequate for the design of blind IQA models of screen content images, which suggests that enriching and complementing NSS features with other descriptors of less naturalistic content may be beneficial.

To address this important yet challenging problem, we propose a novel IQA framework which consists of four elements. The first element is a description of the image complexity, which is affected by artifacts. For instance, image complexity increases when high-frequency noise is injected, whereas it declines as low-frequency blur is introduced. The second element models the normalized bandpass statistics of screen content images, using them to measure the statistical departure of corrupted images from a pristine state. This element is included to mainly account for distortions of those portions of screen content that are naturalistic photographs. The third and fourth elements measure global brightness and surface quality, and picture detail, respectively. A total of 15 features are extracted from each input image signal. To convert features into quality scores, we deploy a large set of training data as samples to learn a regression module. The training samples are composed of about 100,000 screen content images captured on public webpages and labeled using a high-performance screen content FR-IQA model that we describe later. By using a fixed regression module, our approach belongs to the class of *Opinion-Unaware* (OU) NR-IQA models that require no training on human scored images.

The layout of this paper is as follows. Related work is considered in Section II. Section III presents a detailed description of the quality-aware features used and the regression module. Experiments conducted on two screen content image databases [20], [21] are provided to validate the designed features and the proposed OU-NR-IQA model against state-of-the-art quality models, as described in Section IV. We draw some concluding remarks in Section V.

## II. RELATED WORKS

Research on screen content images is relatively new, especially with regards to quality assessment. We first review representative work in the area of screen content IQA.

### A. Image Databases

A basic tool we will use here is the *Screen Image Quality Assessment Database* (SIQAD) [20]. This database, which is the first of its kind, is made up of 20 pristine screen content images and 980 corresponding images distorted by seven categories of distortions: *Gaussian Blur* (GB), *Contrast Change* (CC), *Gaussian Noise* (GN), *Motion Blur* (MB), *JPEG2000 Compression* (J2C), *JPEG Compression* (JC), and *Layer segmentation-backed Coding* (LC) [22], each at seven levels of distortions.

We also use the more specific *Quality Assessment of Compressed Screen content images* (QACS) [21], which captures the effects of compression on the quality of screen content images. The QACS database contains 492 compressed images generated by corrupting 24 undistorted screen content images using two advanced coding methods: *High-Efficiency Video Coding* (HEVC) [23] and the new *Screen Content Compression* (SCC) algorithm, which claims to improve on HEVC when applied to screen content [24].

### B. Quality Measures

Creating picture quality databases of even moderate size, such as SIQAD and QACS, consumes great expense of time and labor. Subjective evaluation is unrealistic in real-time application scenarios. Thus, significant effort has been applied to the development of objective IQA models which are capable of quickly and accurately predicting image quality.

Currently, only a few screen content IQA models have been developed, including four FR models and one NR model. The pioneering first model is the FR *Screen content Perceptual Quality Assessment* (SPQA) model [20], which finds perceptual differences of pictorial and textual areas between distorted and undistorted images. Another FR method uses adaptive window sizes within the classical SSIM framework [25]. A small kernel is used for textual regions while a large kernel is used for pictorial areas. The *Structure-Induced Quality Metric* (SIQM) is an FR model that measures structural degradation predicted by SSIM [26]. Further along this line, the *Saliency-guided Quality Measure of Screen content* (SQMS) model incorporates gradient magnitude structural information and a model of visual saliency [1]. SQMS currently delivers the best correlation performance against human judgments among state-of-the-art FR-IQA models that predict
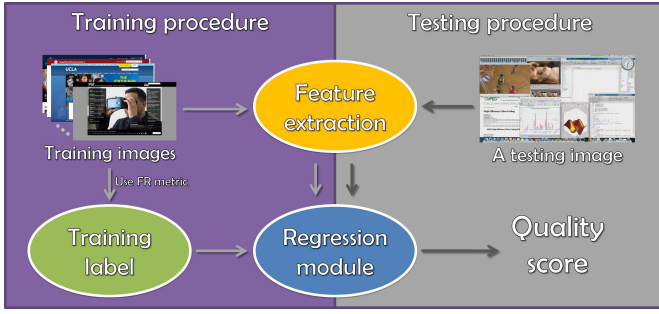
Fig. 2. A general framework for creating blind IQA models without training on human opinion scores.
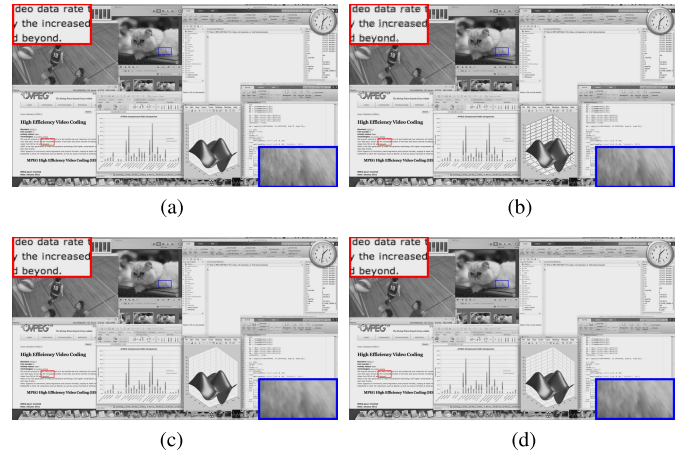


Fig. 3. Comparison of different filters: (a) a lossless screen content image; (b)-(d) processed images created using AR model, BL filter and hybrid filter, respectively.

the quality of screen content images. We will use this model later as a proxy for human predictions to label distorted screen content images.

A no-reference model which is also free of training on human scores was dubbed *Blind Quality Measure for Screen content images* (BQMS) [19]. In this method, 13 features are extracted under a statistical model of screen content pictures, built using 1,000 high-quality "webpage" and "screen snap" images collected from the "Google Images" website. A fixed regression module was learned on 100,000 distorted images assessed/labeled by FR SIQM scores, eliminating the need for subjective tests to create the model. Experimental results validated the competitive performance of the BQMS model against recently proposed FR and NR algorithms.

## III. METHODOLOGY

Figure 2 depicts a general framework for the design of opinion-unaware NR-IQA models via big data learning. This framework could be used to transform any blind IQA model into one that does not require human ratings, such as recent blind IQA models designed to handle multiple distortions [27], [28], infrared images [29], authentic distortions [30], contrast distortions [31], tone-mapped images [32], [33], dehazed images [34], etc. By contrast with NIQE [17] and IL-NIQE [18] which gauge the distance between a query image and a corpus of uncorrupted natural images to infer visual quality, this general framework can be used to derive both general-purpose and distortion-specific IQA models. Our proposed blind quality model is based on this general framework.

### A. Feature Selection

*1) Image Complexity Description:* Image complexity is an important factor to be considered when devising screen content IQA models, since it relates to the effects of gaze direction and spatial masking. Autoregressive (AR) models have been successfully used in the past to estimate image complexity [35], [36], where they have been found to be highly sensitive to distortions and hence, effective for supporting image quality prediction [55], [56]. We measure image complexity by computing an error map between an image and its predicted output generated by an AR model of the input image **s** in a local manner

$$y_q = \mathcal{Q}^n(x_q) \, \mathbf{a} + \tilde{t}_q \qquad (1)$$

where $q$ is the index of a query pixel; $y_q$ is the value of a pixel at location $x_q$; $\mathcal{Q}^n(y_q)$ is composed of the $n$ neighboring pixels of $x_q$; $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$ is a vector of AR parameters; and $\tilde{t}_q$ is the residual error. Then the predicted image is

$$\hat{y}_q = \mathcal{Q}^n(x_q) \, \hat{\mathbf{a}} \qquad (2)$$

where $\hat{\mathbf{a}}$ is determined based on the method in [35].

We present a visual example of a screen content image and its associated AR predicted output in Figs. 3(a)-(b). As can be seen, the AR model performs quite well on textured regions (highlighted by a blue rectangle) [37], but less well near image edges owing to introduced ringing artifacts (highlighted by a red rectangle). An alternative approach would be to deploy the bilateral (BL) filter, which has edge-preserving power and is computationally simple, to modify the AR model towards protecting edges and inhibiting ringing artifacts [38]. The BL filter is defined by

$$y_q = \mathcal{Q}^n(x_q) \, \mathbf{b} + \hat{t}_q \qquad (3)$$

where $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$ are a set of coefficients produced by BL filtering; $\hat{t}_q$ is the error; and $\mathbf{b}$ is the BL filter response. The parameters used in the BL filter follow the assignment in [38], to produce the result shown in Fig. 3(c). The BL filter delivers sharper results near luminance edges than does the AR-based predictor, but it fails to maintain texture details. To obtain the best properties of both models, we devised a hybrid filter that systematically combines the AR and BL filters:

$$\hat{y}_q = \frac{\mathcal{Q}^n(x_q) \, \hat{\mathbf{a}} + \kappa \, \mathcal{Q}^n(x_q) \, \mathbf{b}}{1 + \kappa} \qquad (4)$$

where $\kappa$ adjusts the relative strength of the responses of the AR and BL filters. We fixed this value at $\kappa = 9$, since its associated hybrid filter can yield the output image which exhibits a good tradeoff between the AR and BL predictors, as shown in Fig. 3(d). More analysis about how $\kappa$ was determined will be provided in Section IV. While a simple linear weighting function with fixed weights is used, an adaptive weighting scheme may work better and will be studied in future work.

Next we compute the residual error map $\Delta y_q = y_q - \hat{y}_q$, where large absolute values correspond to pixels that are not accurately predicted, as in highly complex textured regions, while small absolute values correspond to less complex or smooth regions. The image complexity feature is then defined to be the entropy of the residual error map $E_r$:

$$E_r = - \int_i p_i \log p_i \, di \qquad (5)$$

where $p_i$ is the probability density of the $i$-th grayscale in the error map $\Delta y_q$.

Early psychophysical masking experiments [39] and neuropsychological recordings [40] indicated that mechanisms selective to narrow ranges of spatial frequencies and orientations are functionally intrinsic in the human visual system (HVS). These observations have evolved into multiscale cortical models that pervade modern perceptual modeling and visual processing algorithms. Therefore, we also measure the image complexity at a decreased resolution (by subsampling with a stride of 16 pixels along each cardinal direction after applying a $16 \times 16$ square moving low-pass filter. We denote the reduced resolution complexity as $E_d$. More scales were not taken into account since they were found to have little additional impact on performance. Thus, the overall image complexity description is the pair $F_c = \{E_r, E_d\}$.

*2) Screen Content Statistics:* We make measurements of the degradation of image structure in the following way. Given an image $\mathbf{s}$, we denote $\mu_{\mathbf{s}}$, $\sigma_{\mathbf{s}}$ and $\tilde{\sigma}_{\mathbf{s}}$ as local mean and variance maps:

$$\mu_{\mathbf{s}} = \sum_{r=1}^{R} w_r s_r \qquad (6)$$

$$\sigma_{\mathbf{s}} = \left[ \sum_{r=1}^{R} w_r (s_r - \mu_s)^2 \right]^{\frac{1}{2}} \qquad (7)$$

$$\tilde{\sigma}_{\mathbf{s}} = \left[ \sum_{r=1}^{R} (s_r - \mu_s)^2 \right]^{\frac{1}{2}} \qquad (8)$$

where $\mathbf{w} = \{w_r | r = 1, 2, \ldots, R\}$ is a normalized Gaussian window. The structural degradation is then measured by

$$S_\mu(\mathbf{s}) = \frac{1}{D} \sum \left( \frac{\sigma_{(\mu_{\mathbf{s}}, \mathbf{s})} + \delta}{\sigma_{(\mu_{\mathbf{s}})} \sigma_{\mathbf{s}} + \delta} \right) \qquad (9)$$

$$S_\sigma(\mathbf{s}) = \frac{1}{D} \sum \left( \frac{\sigma_{(\sigma_{\mathbf{s}}, \tilde{\sigma}_{\mathbf{s}})} + \delta}{\sigma_{(\sigma_{\mathbf{s}})} \sigma_{(\tilde{\sigma}_{\mathbf{s}})} + \delta} \right) \qquad (10)$$

where $D$ is the number of pixels in $\mathbf{s}$; $\delta$ is an additional fixed positive stabilizing constant; and $\sigma_{(\alpha, \beta)}$ is the local empirical covariance map between $\alpha$ and $\beta$:

$$\sigma_{(\alpha, \beta)} = \sum_{r=1}^{R} w_r (\alpha_r - \mu_\alpha)(\beta_r - \mu_\beta). \qquad (11)$$

Our approach to modeling quality perception is patch based [3]. We deploy two normalized Gaussian window functions to capture microstructure and macrostructure, respectively. This is motivated by the observation that screen content pictures usually include both pictorial and textual parts simultaneously. As in [3] we apply a Gaussian window function of
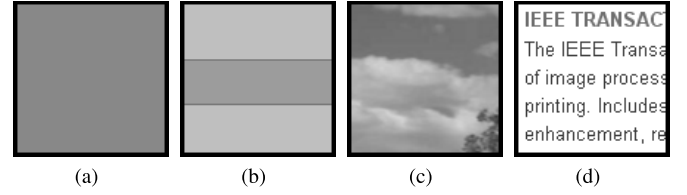


Fig. 4. Comparison of structural degradation information and image complexity measure on four different kinds of image patches: (a) Smooth patch with $E_r = 0.3916$ and $S_{(\mu,3,i)} = 0.9975$; (b) Edge patch with $E_r = 0.8021$ and $S_{(\mu,3,i)} = 0.9868$; (c) Textural patch with $E_r = 2.6455$ and $S_{(\mu,3,i)} = 0.9808$; (d) Textual patch with $E_r = 3.0838$ and $S_{(\mu,3,i)} = 0.9077$.

size $11 \times 11$ and standard deviation 1.5 to capture the structure of pictorial parts.

In order to also capture detailed structures in the textual parts of the images, which often contain fine lines, we also measure (6)-(11) using a smaller Gaussian function of size $3 \times 3$ and unit standard deviation [25]. Hence, we compute (6)-(11) using two windows. Furthermore, we make a detailed analysis of compressed image blocks. When an image is corrupted by block-based (JPEG) compression, using $8 \times 8$ codeblocks, the $6 \times 6$ interior of a coded block is often smoothed by the zeroing of high-frequency block DCT coefficients, whereas block artifacts are commonly introduced along the block edges. So we process the interiors and edges of blocks differently, when extracting structural degradation information. Other distortions, such as noise and blur, affect the block interiors and edges almost uniformly [41]. This analysis yields eight structural degradation features, denoted $S_{(a,b,c)}$, where $a = \{\mu, \sigma\}$ indicates information type, $b = \{3, 11\}$ indicates kernel size, and $c = \{i, e\}$ indicates block interiors and edges, respectively.

Using the concept of image structural similarity in [3], we use Eq. (9), we to measure the variations between structures in the image $\mathbf{s}$ and an associated blurred version $\mu_{\mathbf{s}}$ of it. Similarly, in Eq. (10), we first remove the mean from $\mathbf{s}$ and $\mu_{\mathbf{s}}$ to generate $\sigma_{\mathbf{s}}$ and $\tilde{\sigma}_{\mathbf{s}}$, then compute the structural differences between them. We suppose that image complexity should have a negative correlation with structural degradation information defined in Eqs. (9)-(10). That is to say, a high-complexity image generally has low structural degradation, and vice versa. As shown in Fig. 4, four representative image patches that belong to different types are selected for comparison. The values of their associated image complexity $E_r$ and one of structural degradation information $S_{(\mu,3,i)}$ are presented in the figure. In the example, as the image complexity rises, the structural degradation information reduces.

To further demonstrate our supposition, a total of 800 screen content pictures were gathered with screenshot tools to examine the correlation between structural degradation features and image complexity features. These images are composed of homepages of well-known journals, conferences and workshops, a matlab interface, international and domestic college and web portals, webstore platforms, google maps, webpage gaming and more. No overlap exists between the captured screen content images and the 44 source images in the SIQAD and QACS databases used for testing. Sixteen representative screen content images are shown in Fig. 5. Eight structural
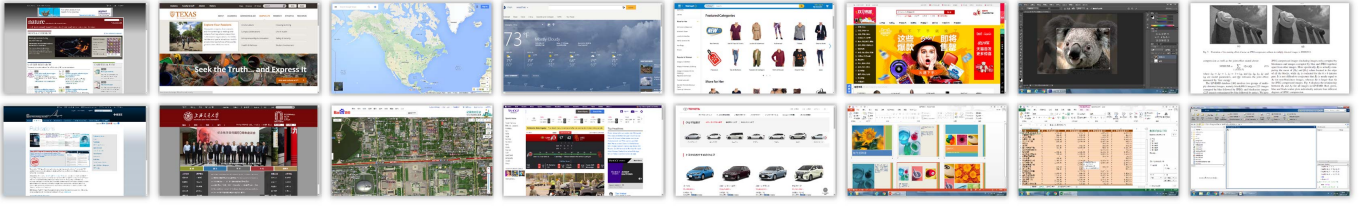
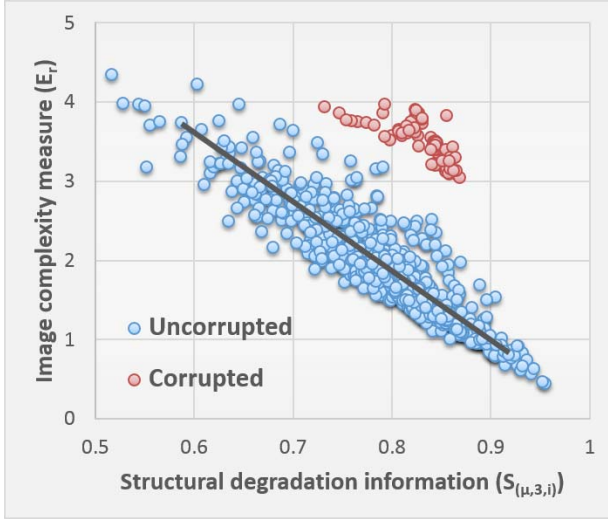Fig. 5.    Sixteen representative images of screen content scene collected using screenshot tools.



Fig. 6.    Representative scatter plot of image complexity feature $E_r$ versus structural degradation information $S_{(\mu,3,i)}$ on uncorrupted (blue points) and corrupted (red points) screen content images.

degradation features $S_{(a,b,c)}(\mathbf{s}_0)$, where $\mathbf{s}_0$ indicates an uncorrupted screen content image, and the image complexity features $E_r(\mathbf{s}_0)$ were compared using the captured screen content images. One exemplified scatter plot is shown in Fig. 6. Blue points are associated to uncorrupted screen content images. As may be seen, there is an evident near-linear relationship on uncorrupted images between the image complexity feature $E_r$ and the structural degradation $S_{(\mu,3,i)}$. This motivates exploring the possibility of predicting visual distortions by measuring the departure of a corrupted screen content image from this linear relationship observed on good quality screen content images. Therefore we attempt to fit the linear regression model:

$$E_r(\mathbf{s}_0) = \begin{bmatrix} A_{(a,b,c)} \\ B_{(a,b,c)} \end{bmatrix}^T \begin{bmatrix} S_{(a,b,c)}(\mathbf{s}_0) \\ 1 \end{bmatrix} \qquad (12)$$

where $[A_{(a,b,c)}, B_{(a,b,c)}]$ indicates one of 8 parameter pairs corresponding to $(a, b, c)$. We use the least square method to estimate these parameters.

Structural degradation features capture variations in image structure, whereas image complexity measurements are responsive to image details. Thus, structural degradation features and image complexity features exhibit differing sensitivities to the levels and types of distortion. Generally, we find that the approximate linear relationship between uncorrupted screen content picture features will be disturbed when distortions are introduced, as shown by the red points in Fig. 6.

Based on this notion, define $T_{(a,b,c)}(\mathbf{s}) = E_r(\mathbf{s}) - (A_{(a,b,c)} \cdot S_{(a,b,c)}(\mathbf{s}) + B_{(a,b,c)})$. The values of $T_{(a,b,c)}(\mathbf{s})$ computed on high quality images should approach zero, while on corrupted images $T_{(a,b,c)}(\mathbf{s})$ will depart from zero with increasing distance when the distortion grows. We then define features predictive of screen content distortions to be $T_{(a,b,c)}$, where $a = \{\mu, \sigma\}$, $b = \{3, 11\}$, and $c = \{i, e\}$.

*3) Global Measurement of Brightness and Surface Quality:* The above-described features are effective for gauging many visual degradations, but are not able to capture undesirable brightness shifts or contrast alterations. Of these, contrast alteration is more difficult to address as it also affects the image complexity: an enhanced contrast may increase the image complexity and vice versa. Thus, we seek features that are insensitive to noise, blur and other artifacts, but are sensitive to contrast adjustment. To this end, we deploy the sample mean of the image $\mathbf{s}$, denoted as $O_1$:

$$O_1 = E(\mathbf{s}) = \frac{1}{D} \sum_{d=1}^{D} s_d. \qquad (13)$$

This feature captures brightness shifts resulting from errors of improper post-processing technologies. We also measure the sample skewness of the image $\mathbf{s}$:

$$O_3 = \frac{E[(\mathbf{s} - O_1)^3]}{\sqrt{E^3[(\mathbf{s} - O_1)^2]}}. \qquad (14)$$

As shown in [42], this feature has a positive correlation with image contrast. For illustration consider the example in Fig. 7. The processed screen content image with greater skew appears glossier and darker than its corresponding original version. In [42], a heuristic model was presented that relates the perception of surface quality to skewness. They suggested a neural mechanism supportive of the model: an accelerating nonlinearity responsive to on- and off-center visual neurons could be used to calculate skewness and thus predict the perceived image surface quality. To summarize, we measure features related to global brightness and surface quality and denote them $F_{bs} = \{O_1, O_3\}$.

*4) Detail Assessment of Sharpness and Corners:* The last thirty years have witnessed on explosive growth of picture compression technologies. Compression generally introduces complex interplays of multiple distortions. We use two classes of features designed to sense two major types of compression distortion: local sharpness loss and blocking. The first factor senses loss of sharpness [43], [44], [45]. Similar to [44], we measure the log-energy of wavelet subbands (9/7 Danbechies DWT filters) of an image at three

(a) Original screen image      (b) Histogram of image (a)      (c) Enhanced screen image      (d) Histogram of image (c)
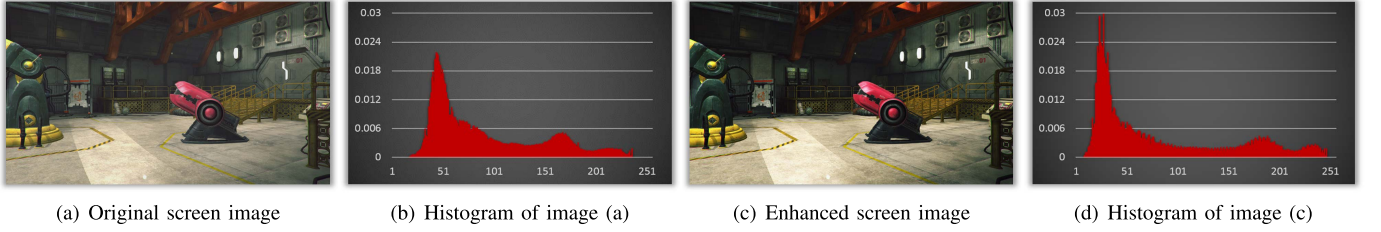
Fig. 7. Illustration of skewness: (a)-(b) original screen content image with skewness 0.8662 and its histogram; (c)-(d) enhanced screen content image with skewness 0.9084 and its histogram. It is apparent that (d) has a longer tail than (b).
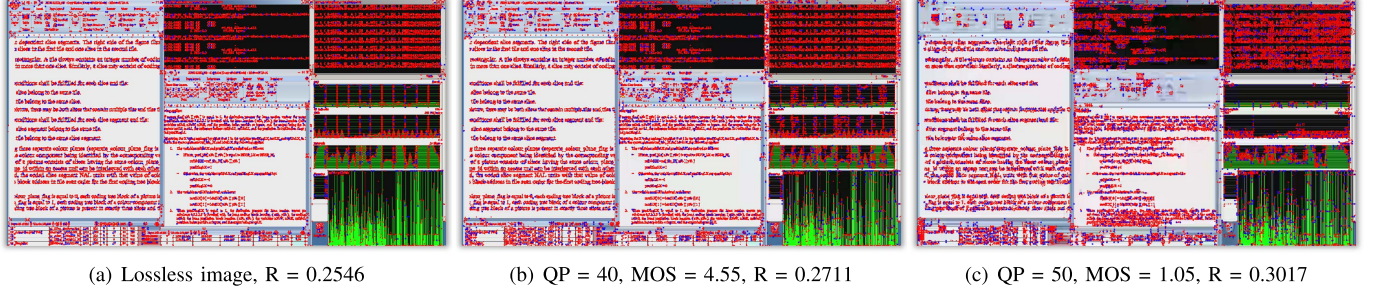


(a) Lossless image, R = 0.2546      (b) QP = 40, MOS = 4.55, R = 0.2711      (c) QP = 50, MOS = 1.05, R = 0.3017

Fig. 8. Comparison of images with different compression levels. QP: the quality parameter of compression. MOS: mean opinion score. R will be defined in (19). (a) Lossless image; (b)-(c) Compressed image with QP = 40 and QP = 50.

scales, $\{LL_3, LH_n, HL_n, HH_n\}$, where $n = 1, 2, 3$. At each decomposition level, the log-energy is calculated

$$L_{m,n} = \log_{10}\left[1 + \frac{1}{M_n}\sum_h m_n^2(h)\right] \quad (15)$$

where $h$ is the pixel index; $m$ indexes $LH$, $HL$, and $HH$; and $M_n$ indicates the number of wavelet coefficients at the $n$-th level. The log-energy at each decomposition level is measured as

$$L_n = \frac{L_{LH,n} + L_{HL,n} + \gamma L_{HH,n}}{2 + \gamma} \quad (16)$$

where we fixed $\gamma = 8$ to impose a larger impact on the $HH$ subbands and more discussions about how to determine this parameter will be given in Section IV. Only the 2nd and 3rd levels are used to capture sharpness-related information. We have found that using all three levels does not yield any gain in performance. The second type of compression feature measures blockiness via a corner detection technique. It was shown in [46] that compression blockiness is closely correlated with corners. Fig. 8 exemplifies how corners can arise and vary with compression on a screen content image. Certainly, this type of computer-created content contains many sharp edges and regular patterns, hence genuine corners may often arise in screen content images. However, pseudo corners arise due to blockiness from compression. We take the strategy that while genuine corners may be found anywhere, detected pseudo corners only occur at block boundaries. Thus define the image matrix $\mathbf{S} = (s_{ij})_{\tau \times \upsilon}$, where $\tau$ and $\upsilon$ indicate image height and width respectively. Corners are first detected using the Shi-Tomasi detector [47]. Denote the corner map $\mathbf{C} = (c_{ij})_{\tau \times \upsilon}$

and the pseudo corner map $\mathbf{P} = (p_{ij})_{\tau \times \upsilon}$, where

$$c_{ij} = \begin{cases} 1 & \text{if } s_{ij} \in \mathbb{C} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$p_{ij} = \begin{cases} 1 & \text{if } s_{ij} \in \mathbb{C}, mod(i,k) \leqslant 1, mod(j,k) \leqslant 1 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where $s_{ij} \in \mathbb{C}$ means that a corner was detected at location $(i, j)$, $mod$ retains the remainder, and $k$ denotes the size of the compression blocks (typically $8\times 8$ in JPEG). In Fig. 8, red dots indicate $\mathbf{C} = (c_{ij})_{\tau \times \upsilon}$ while red and blue dots together represent $\mathbf{P} = (p_{ij})_{\tau \times \upsilon}$. As compression distortion is increased, more pseudo corners appear due to blockiness, while genuine corners begin to disappear because of intra-block blurring. To combine these, compute the ratio of pseudo corners to all corners:

$$R = \sqrt{\xi_p / \xi_c} \quad (19)$$

where $\xi_p = \sum_{i,j} p_{ij}$ and $\xi_c = \sum_{i,j} c_{ij}$ are the number of pseudo corners and all corners, respectively. Hence the last features computed related to image sharpness and corners are $F_{sc} = \{L_2, L_3, R\}$.

Overall, there are a total of 15 features extracted, descriptive of image complexity (Index 1), screen content scene statistics (Index 2), global brightness and surface quality (Index 3), and compression-induced image sharpness loss and blocky corners (Index 4). We summarize these features in Table I.

### B. Module Regression

The 15 features must be combined to provide a single direct prediction of the visual quality of a screen content image. We therefore deploy a regression engine that can reliably

TABLE I
SUMMARY OF FEATURES FOR BLIND IQA OF SCREEN CONTENT

| Index | Feature type | Symbol | Feature description | Feature ID |
|---|---|---|---|---|
| 1 | Image complexity description | $E_r, E_d$ | Image complexity estimation at original and reduced resolution | $f_{01} - f_{02}$ |
| 2 | Screen content statistics | $T_{(a,b,c)}$ | Departure from scene statistics of undistorted screen content images | $f_{03} - f_{10}$ |
| 3 | Global brightness and surface quality | $O_1, O_3$ | Measurement of histogram mean and skewness | $f_{11} - f_{12}$ |
| 4 | Compression assessment | $L_2, L_3$ | Image sharpness loss assessment | $f_{13} - f_{14}$ |
| | | $R$ | Fraction of corners that are blocky | $f_{15}$ |

convert 15 features into a single quality index. We use an efficient support vector regression (SVR) [13], [14], [15] to transform the features to an overall quality score. Specifically, we used the LibSVM package to implement the SVR using the Radial Basis Function (RBF) kernel [48].

To test our model we computed the median performance across 1,000 trial splits into 80% data for training and 20% data for testing. Current image quality databases contain a limited number of different scenes and distortion levels, less than 1,500 in the case of screen content images. Hence, if a regression module is found using just a few thousand screen content images as training data, it is difficult to ensure that the derived regression module will succeed when applied to a broader scope of image scenes and distortion levels. To cope with this problem, a growing body of OU blind IQA metrics have been proposed [17], [18]. Broadly speaking, opinion-aware (OA) methods rely on human-labeled training images, while OU methods do not depend on training images labeled with subjective ratings. OU models are regarded as having greater potential to generalize on high-volumes of real world images.

One modern strategy for the design of OU-NR-IQA models relies on NSS constraints, as first exemplified by NIQE [17] and IL-NIQE [18]. These blind models predict image quality by measuring the distance between an input query image and a set of lossless natural images in accordance with NSS models. This design strategy is effective when constructing general-purpose NR-IQA algorithms that can handle various types of distortions. Another effective strategy for developing OU-NR-IQA models resorts to the general framework provided at the beginning of this section. A significant advantage of this framework is its flexibility in developing general-purpose (or distortion-specific) blind IQA models based on much larger datasets of training images corrupted by a wider array of distortion types. We use this general framework to train an SVR to learn a regression module using a very large body of training data. While the SVR is highly efficient, we plan to explore more sophisticated learning tools in the future.

*1) Training Samples:* Unlike camera-captured natural scene images, screen content images are usually generated or assembled by a computer. The aforementioned 800 screen content images we gathered were used to create the model of screen content images. We applied 11 types of distortions to corrupt 800 screen content images to create 100,000 distorted images as training samples. The 11 distortion types used were GN,

JC, J2C, HEVC, SCC, GB, MB and four CC-related distortions that include Gamma transfer, brightness intensity-shifting, etc, as used in the CCID2014 database that was designed to enable the analysis of contrast alterations [49]. The authors of [19] collected 1,000 "webpage" and "screen snap" images by downloading them from the "Google Images" website. However, those images were not examined to determine whether they were free from visible distortions. Further, the image content was somewhat limited and resolution of some of the images was quite low. Hence, we have manually collected 800 apparently distortion-free screen content images containing much richer content, as described earlier.

*2) Training Labels:* Training labels in IQA researches are generally derived from subjective experiments. This kind of experiment is quite time- and expense-consuming, and not suitable for labeling very large number of training images. Hence, we avoided the problem of large-scale human studies by following the method of [19], where scores produced by an objective quality algorithm were used as training labels to replace subjective opinion scores. Ideally, a high-performance FR-IQA model should be used to approximate human ratings. We deployed the FR SQMS metric, which achieves superior correlation performance when used to assess screen content pictures. We labeled about 100,000 training images (after outlier removal) with predicted quality scores delivered by SQMS. By training the SVR on such high-volume training data, we obtained a fixed regression module which converts 15 features extracted into a single quality prediction score. We call this model the Screen Image Quality Evaluator (SIQE).

*3) Data Cleaning:* An inevitable risk underlying any FR metric based learning framework is that incorrectly labeled training data may mislead the training process. This suggests that a mechanism to detect and eliminate noisy training data would be useful [52]. Our approach to this is to compare the quality predictions delivered by two high-performance FR metrics to detect potentially "noisy" quality predictions. Specifically, we deploy the SQMS and SIQM algorithms, both of which have been shown to have high prediction accuracy on the screen content IQA problem. To detect noisy instances, we measured the PLC between the SQMS and SIQM scores on each of the 800 image contents. Figure 9 plots the 800 PLC values, which shows that the vast majority of the PLC values was quite high, with just a few falling below 0.9, as indicated in red in Fig. 9. We assumed that these low PLC predictions were "noisy", and removed these image contents and their corresponding training images. Note that using an FR model based training framework can be used to introduce a very large number of training samples, thereby alleviating the overfitting problem. Using FR models based on complementary quality measurement techniques is a reasonable way to clean noisy training data, yet we believe that this new and tough problem merits further deep study.

### C. Complexity Reduction

The hybrid filter operates locally, which makes SIQE inefficient. For an image of size $2560 \times 1440$, the time consumed to
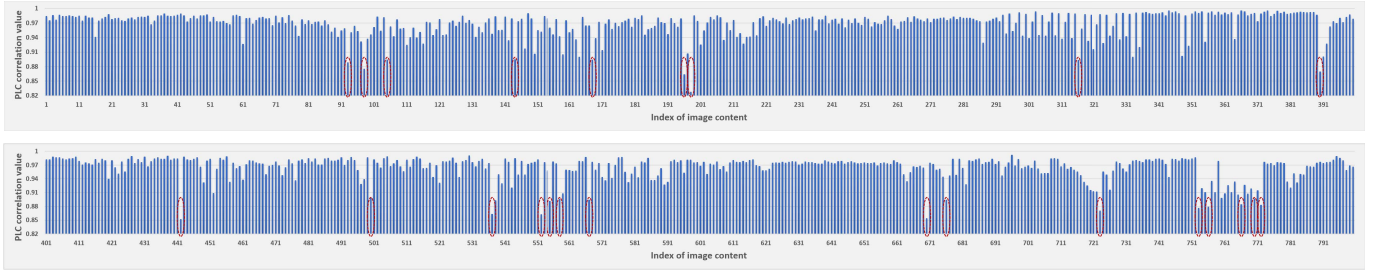
Fig. 9.    Illustration of PLC correlation values between SQMS scores and SIQM scores for each of 800 image contents.

TABLE II

COMPUTATION TIME OF EACH TYPE OF EXTRACTED FEATURES

| Feature type | Image complexity | | Screen content scene statistics | Brightness and surface quality | | Sharpness and corners | |
|---|---|---|---|---|---|---|---|
| Symbol | $E_r$ | $E_d$ | $T_{(a,b,c)}$ | $O_1$ | $O_3$ | $L_2$ , $L_3$ | $R$ |
| Feature ID | $f_{01}$ | $f_{02}$ | $f_{03}$ - $f_{10}$ | $f_{11}$ | $f_{12}$ | $f_{13}$ - $f_{14}$ | $f_{15}$ |
| Time (second) | 798.7 | 3.268 | 0.702 | 0.001 | 0.152 | 0.269 | 0.193 |
| | **802.0** | | **0.702** | **0.153** | | **0.462** | |

compute each of the four types of features is listed in Table II. Implementing SIQE on a high-definition image consumes considerable time: about 804 seconds. The cost of estimating image complexity is more than 600 times that of the other three feature types.

One way to simplify computation of the hybrid filter (especially the AR model contained therein) would be to remove the AR model while preserving the BL filter. We call this simplified version the Simplified Screen Image Quality Evaluator (SSIQE). SSIQE requires only 42.2 and 0.19 seconds of compute time on the above 2560 × 1440 image, resulting in computational efficiency gains of 19 and 17 times. The second way is to introduce highly efficient algorithms to simulate the output of the hybrid filter. Computing the entropy of the difference between an image and its prediction, namely $\Delta y_q$, is closely connected to predictive coding [50], [51]. Following this idea, we exploit the compressibility of an image to estimate complexity. We examined five compression methods: JPEG, JP2K, H.264, HEVC and SCC. As a good tradeoff between effectiveness and efficiency, we adopted JPEG compression in the "lossless" mode, and used the achieved bit per pixel (bpp) value as an alternate, but related method of image complexity estimation. Using the 100,000 training images, the scatter plot between the JPEG-based bpp values $B_r$ and the image complexity measure $E_r$ computed via the hybrid filter is shown in Fig. 10. There is a broadly linear relationship (linear correlation exceeds 95%). Similar to (12), establish this linear model and seek the two parameters found by least squares. Replacing the image complexity estimates $E_r$ and $E_d$ with compression-backed $B_r$ and $B_d$, results in an alternate, faster model, dubbed the Accelerated Screen Image Quality Evaluator (ASIQE). Using the same 2560 × 1440 image, the computation requires only 0.125 and 0.022 seconds when computing $B_r$ and $B_d$ respectively, or about 6400 and 150 times the computational efficiency relative to computing $E_r$ and $E_d$.
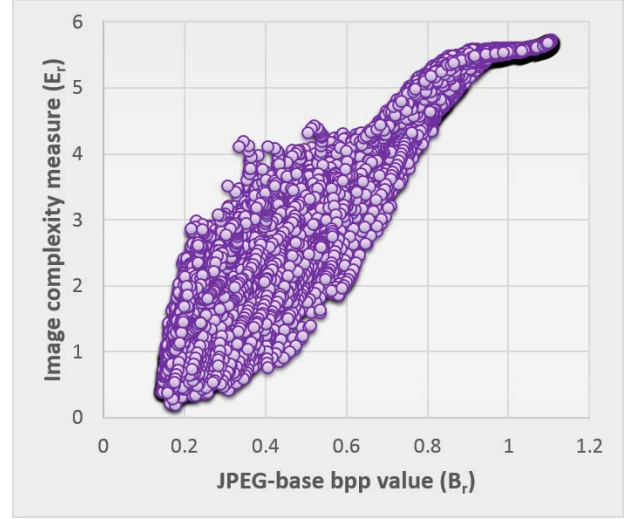


Fig. 10.    Scatter plot of image complexity measure ($E_r$) and JPEG-based bpp value ($B_r$) on 100,000 training images.

More comparisons between the SIQE and ASIQE models are given in the next section.

## IV. EXPERIMENTS AND DISCUSSIONS

We measured and compared the correlation performance of the blind SIQE, SSIQE and ASIQE models against 16 modern IQA models on the SIQAD and QACS databases.

### A. Testing Protocol

*1) Algorithms:* Numerous well-established IQA models have been proposed during the past decade. The majority of these have been demonstrated to be not only performance-effective but also time-efficient. In order to study the effectiveness of quality models proposed here, three classes of 16 approaches were selected for comparison. The first class is composed of five OA-NR-IQA algorithms: BLIINDS-II [13], BRISQUE [14], SSEQ [53], GMLF [54] and NFERM [55]. The second class involves eight recently explored FR-IQA models: FSIMc [6], GSI [7], IGM [56], VSI [10], PSIM [57], ADD-GSIM [58], SIQM [26] and SQMS [1]. The third class consists of three start-of-the-art blind OU-IQA methods, NIQE [17], IL-NIQE [18] and BQMS [19].

*2) Databases:* To the best of our knowledge, only two existing databases, SIQAD [20] and QACS [21], are relevant to screen content picture quality evaluation. Detailed illustrations

TABLE III

COMPARISON OF FEATURE EFFECTIVENESS OF EIGHT POPULAR BLIND IQA MODELS. WE BOLD THE TOP THREE METRICS

| SIQAD [20] | BLIINDS-II [13] | BRISQUE [14] | SSEQ [53] | GMLF [54] | NFERM [55] | SIQE (Pro.) | SSIQE (Pro.) | ASIQE (Pro.) |
|---|---|---|---|---|---|---|---|---|
| PLC | 0.5953 | 0.7882 | 0.7428 | 0.7520 | 0.8022 | **0.8384** | **0.8371** | **0.8201** |
| SRC | 0.5537 | 0.7608 | 0.6939 | 0.7257 | 0.7761 | **0.8235** | **0.8212** | **0.7931** |
| KRC | 0.3963 | 0.5717 | 0.5100 | 0.5356 | 0.5844 | **0.6320** | **0.6301** | **0.6016** |

| QACS [21] | BLIINDS-II [13] | BRISQUE [14] | SSEQ [53] | GMLF [54] | NFERM [55] | SIQE (Pro.) | SSIQE (Pro.) | ASIQE (Pro.) |
|---|---|---|---|---|---|---|---|---|
| PLC | 0.6319 | 0.5477 | 0.6146 | 0.5292 | 0.5348 | **0.7570** | **0.7549** | **0.7588** |
| SRC | 0.6265 | 0.5601 | 0.6092 | 0.5437 | 0.5489 | **0.7569** | **0.7530** | **0.7521** |
| KRC | 0.4609 | 0.4099 | 0.4493 | 0.3947 | 0.3997 | **0.5801** | **0.5797** | **0.5763** |

can be found in Section II-A. The SIQAD database includes 980 screen content images corrupted by conventional single distortion types, e.g. blur and noise, while the QACS database contains 492 screen content images distorted by two distortion types, i.e. HEVC and SCC compression technologies.

*3) Criteria:* In most cases, we use four typical performance evaluation criteria, the Pearson Linear Correlation coefficient (PLC), Spearman Rank order Correlation coefficient (SRC), Kendall's Rank-order Correlation coefficient (KRC) and Root Mean Square error (RMS). SRC and KRC are directly calculated between raw objective quality predictions and subjective scores, but the other two are computed after a regression process, following ITU Recommendation BT500 [59]. Here a logistic regression with five parameters was used:

$$q_c = v_1 \left[ \frac{1}{2} - \frac{1}{1 + e^{v_2(q_r - v_3)}} \right] + v_4 q_r + v_5 \qquad (20)$$

where $q_r$ and $q_c$ denote raw and converted quality predictions of an objective IQA metric. We used the MATLAB functions "nlinfit" and "nlpredci" to fit the curve and estimate those five model parameters' values. PLC gauges prediction accuracy between two input variable vectors, whereas RMS computes prediction consistency. SRC and KRC are non-parametric measures of monotonicity. A good model should lead to large values of PLC, SRC and KRC, and small values of RMS.

### B. Performance Evaluation

*1) Feature Comparison:* We first applied a popular and commonly employed test to examine the effectiveness of each of the selected features relative to five modern NR-IQA algorithms. Following the testing procedure conducted in [13], [14], and [55], we randomly divided the 980 SIQAD images into two sets. One set included 784 distorted screen content images associated with 16 reference images while the other set included 196 testing screen content images associated with the remaining 4 reference images. The regression module was then learned using the 80% of training data from the first set, then the performance indices were calculated using the other 20% of the test data from the second set. The above process was iterated 1,000 times and the median correlations across the 1,000 trials was recorded. In Table III the median performance indices of the BLIINDS-II, BRISQUE, SSEQ, GMLF, NFERM and the proposed SIQE, SSIQE and ASIQE are reported. The three best performing models are underlined and bolded. Clearly, our three proposed

blind quality models obtained highly competitive performance against the five compared models. It may be seen that the performance of ASIQE and SSIQE is a little inferior to that of SIQE in most cases.

*2) Metric Comparison:* We compared the performance of the proposed SIQE, SSIQE and ASIQE models with state-of-the-art FR methods: FSIMc, GSI, IGM, VSI, PSIM, ADD-GSIM, SIQM, SQMS, and the OU-NR models: NIQE, IL-NIQE, and BQMS. The results are shown in Table IV using a linearly weighted average performance comparison, where the relative weights were assigned in proportion to the number of images in the testing databases. The top three methods in each type are bolded and underlined in the table. We draw several main conclusions. First, the proposed SIQE, SSIQE and ASIQE models were clearly superior to the other blind quality models, especially on the QACS database. This is quite reasonable, since the features extracted were devised specifically for screen content IQA, and the training labels came from the top-performing FR SQMS model. Second, SIQE, SSIQE and ASIQE outperformed most FR IQA methods tested. Third, the ASIQE model achieved similar performance as SIQE, but with a vast reduction in cost, making it more applicable to real-time systems. Fourth, there was only a very small difference between SIQE and ASIQE on the SIQAD database, but a slightly larger on the QACS database, possibly because the linear correlation between $E_r$ and $B_r$ on compressed screen content images (about 96%) was much higher than that on other types of corrupted screen content images (about 90%).

*3) Feature Contribution:* The contribution of each feature is a critical aspect of any quality prediction model [60], [61]. Hence, we examined the variations of the SRC values for different combinations of feature groups (FGs). Each of four FGs, as provided in Table I, was compared and ranked in terms of SRC: FG2 (0.598) > FG4 (0.426) > FG1 (0.411) > FG3 (0.120). Next, we fixed the optimal FG2 and added each of the other three FGs individually, revealing the following rankings: FG2+4 (0.799) > FG1+2 (0.785) > FG2+3 (0.707). We then repeated the above process by fixing the best performing FG2+4, and separately adding FG1 and FG3. This yielded the rankings: FG1+2 + 4 (0.815) > FG2+3 + 4 (0.804). Compared with the SIQE's SRC value (0.824), we arrive at two conclusions: 1) all four FGs play crucial roles in the design of SIQE metric; 2) the rank of the feature contribution is FG2 > FG4 > FG1 > FG3. For the reviewers' conveniences, we present the above SRC values in Fig. 11. We furthermore checked the performance of SIQE without the first feature, due

TABLE IV

COMPARISON OF 14 MODERN FR- AND OU-NR-IQA METHODS. WE BOLD THE TOP THREE MODELS OF EACH TYPE

| Databases | SIQAD (980 images) [20] | | | | QACS (492 images) [21] | | | | Weighted average | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Indices | PLC | SRC | KRC | RMS | PLC | SRC | KRC | RMS | PLC | SRC | KRC | RMS |
| FSIMc [6] | 0.5920 | 0.5817 | 0.4253 | 11.537 | **0.9019** | **0.9039** | **0.7331** | **0.9585** | 0.6955 | 0.6894 | 0.5282 | 8.0013 |
| GSI [7] | 0.5686 | 0.5483 | 0.4054 | 11.775 | 0.8921 | 0.8947 | 0.7215 | 1.0025 | 0.6767 | 0.6641 | 0.5110 | 8.1744 |
| IGM [56] | 0.6439 | 0.6424 | 0.4762 | 10.952 | **0.9116** | **0.9126** | **0.7436** | **0.9117** | 0.7334 | 0.7327 | 0.5656 | 7.5961 |
| VSI [10] | 0.5568 | 0.5381 | 0.3874 | 11.890 | 0.8715 | 0.8719 | 0.6941 | 1.0879 | 0.6620 | 0.6497 | 0.4899 | 8.2795 |
| PSIM [57] | **0.7144** | **0.7056** | **0.5393** | **10.016** | 0.8680 | 0.8683 | 0.6897 | 1.1015 | **0.7658** | **0.7600** | **0.5896** | **7.0364** |
| ADD-GSIM [58] | 0.6844 | 0.6842 | 0.5237 | 10.436 | 0.8690 | 0.8713 | 0.6887 | 1.0977 | 0.7461 | 0.7467 | 0.5789 | 7.3148 |
| SIQM [26] | **0.8625** | **0.8582** | **0.6679** | **7.2433** | 0.8920 | 0.8953 | 0.7241 | 1.0030 | **0.8724** | **0.8706** | **0.6867** | **5.1575** |
| SQMS [1] | **0.8872** | **0.8803** | **0.6936** | **6.6039** | **0.9059** | **0.9096** | **0.7470** | **0.9396** | **0.8935** | **0.8901** | **0.7114** | **4.7107** |
| NIQE [17] | 0.3758 | 0.3742 | 0.2543 | 10.927 | 0.4240 | 0.3701 | 0.2563 | 1.7091 | 0.3919 | 0.3729 | 0.2550 | 7.8460 |
| IL-NIQE [18] | 0.1958 | 0.2021 | 0.1357 | 11.569 | 0.2374 | 0.2603 | 0.1792 | 1.8395 | 0.2097 | 0.2216 | 0.1503 | 8.3170 |
| BQMS [19] | 0.7575 | 0.7251 | 0.5326 | 7.4608 | 0.4264 | 0.3761 | 0.2640 | 1.6701 | 0.6468 | 0.6084 | 0.4428 | 5.5253 |
| SIQE (Pro.) | **0.7906** | **0.7625** | **0.5665** | **8.7650** | **0.7168** | **0.7012** | **0.5228** | **1.5470** | **0.7659** | **0.7420** | **0.5519** | **6.3525** |
| SSIQE (Pro.) | **0.7849** | **0.7537** | **0.5592** | **8.8689** | **0.7025** | **0.6898** | **0.5138** | **1.6067** | **0.7574** | **0.7323** | **0.5440** | **6.4416** |
| ASIQE (Pro.) | **0.7884** | **0.7570** | **0.5609** | **8.8064** | **0.7046** | **0.7005** | **0.5182** | **1.5743** | **0.7603** | **0.7381** | **0.5466** | **6.3891** |

TABLE V

PERFORMANCE EVALUATION OF USING DIFFERENT FR-IQA METRICS FOR LABELING THE LARGE-SCALE TRAINING SAMPLES

| Databases | SIQAD (980 images) [20] | | | | QACS (492 images) [21] | | | | Weighted average | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Indices | PLC | SRC | KRC | RMS | PLC | SRC | KRC | RMS | PLC | SRC | KRC | RMS |
| SIQE-II (Pro.) | 0.7904 | 0.7593 | 0.5632 | 8.7899 | 0.6598 | 0.6543 | 0.4787 | 1.6671 | 0.7467 | 0.7242 | 0.5350 | 6.4092 |
| SSIQE-II (Pro.) | 0.7838 | 0.7495 | 0.5536 | 8.8892 | 0.6452 | 0.6392 | 0.4685 | 1.6949 | 0.7375 | 0.7126 | 0.5252 | 6.4846 |
| ASIQE-II (Pro.) | 0.7869 | 0.7541 | 0.5589 | 8.8287 | 0.6544 | 0.6471 | 0.4709 | 1.6775 | 0.7426 | 0.7183 | 0.5295 | 6.4385 |
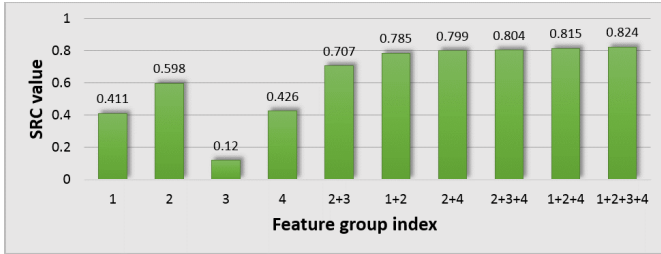


Fig. 11.    SRC values of different combinations of feature types.

to its high cost. The SRC and KRC values were respectively 0.8201 and 0.6296, which is a little inferior to SIQE. Despite little performance loss, we are unable to isolatedly remove the first feature since it is also used to compute the second type of features. We therefore introduce a very fast compression technique to approximate the first feature.

*4) Training Label Influence:* The impact of using different FR metrics to label the training data deserves exploration and discussion. Apart from the SQMS metric, we also used a high-performance SIQM model specific to screen content IQA. Similar to the proposed SQMS-trained SIQE, SSIQE and ASIQE models, SIQM was used to generate training labels and thus to generate new models that we denote as SIQE-II, SSIQE-II and ASIQE-II. We illustrate the performance evaluation results in Table V. As may be seen, SIQE-II, SSIQE-II and ASIQE-II achieved encouraging performance. As compared to the SQMS-trained models, the performance was a little lower, likely because SQMS performs somewhat better than SIQM on the screen content IQA problem.

TABLE VI

ROBUST OF PARAMETERS IN THE PROPOSED SIQE METRIC

| $\kappa$ | PLC | SRC | KRC | $\gamma$ | PLC | SRC | KRC |
|---|---|---|---|---|---|---|---|
| 1 | 0.8362 | 0.8213 | 0.6291 | 1 | 0.8121 | 0.7986 | 0.6038 |
| 2 | 0.8373 | 0.8222 | 0.6308 | 2 | 0.8219 | 0.8065 | 0.6126 |
| 3 | 0.8375 | 0.8222 | 0.6309 | 3 | 0.8278 | 0.8111 | 0.6177 |
| 4 | 0.8375 | 0.8222 | 0.6309 | 4 | 0.8312 | 0.8150 | 0.6215 |
| 5 | 0.8378 | 0.8224 | 0.6312 | 5 | 0.8340 | 0.8173 | 0.6243 |
| 6 | 0.8376 | 0.8221 | 0.6309 | 6 | 0.8356 | 0.8188 | 0.6263 |
| 7 | 0.8374 | 0.8219 | 0.6304 | 7 | 0.8362 | 0.8199 | 0.6275 |
| 8 | 0.8374 | 0.8217 | 0.6306 | **8** | **0.8384** | **0.8235** | **0.6320** |
| **9** | **0.8384** | **0.8235** | **0.6320** | 9 | 0.8377 | 0.8231 | 0.6312 |
| 10 | 0.8372 | 0.8214 | 0.6309 | 10 | 0.8375 | 0.8231 | 0.6310 |
| 11 | 0.8373 | 0.8215 | 0.6308 | 11 | 0.8376 | 0.8231 | 0.6305 |
| 12 | 0.8374 | 0.8218 | 0.6308 | 12 | 0.8373 | 0.8231 | 0.6303 |
| 13 | 0.8375 | 0.8216 | 0.6307 | 13 | 0.8373 | 0.8228 | 0.6306 |
| 14 | 0.8374 | 0.8216 | 0.6307 | 14 | 0.8372 | 0.8225 | 0.6304 |
| 15 | 0.8374 | 0.8215 | 0.6308 | 15 | 0.8369 | 0.8223 | 0.6304 |
| 16 | 0.8373 | 0.8216 | 0.6306 | 16 | 0.8364 | 0.8221 | 0.6304 |
| 17 | 0.8373 | 0.8217 | 0.6306 | 17 | 0.8361 | 0.8220 | 0.6303 |
| 18 | 0.8373 | 0.8219 | 0.6304 | 18 | 0.8362 | 0.8211 | 0.6301 |
| 19 | 0.8373 | 0.8218 | 0.6304 | 19 | 0.8362 | 0.8207 | 0.6301 |
| 20 | 0.8373 | 0.8219 | 0.6304 | 20 | 0.8361 | 0.8201 | 0.6291 |

*5) Parameters Robustness:* The two parameters, namely $\kappa$ in Eq. (4) and $\gamma$ in Eq. (16), were empirically assigned. Here we discuss the sensitivity of each parameter by enumerating 19 values in a proper interval around the determined value while settling the other one. The results are listed in Table VI.

TABLE VII

MEAN COMPUTATION COST (IN SECONDS/IMAGE) OF THE SIQE, SSIQE, ASIQE AND 16 OTHER MODELS ON THE SIQAD DATABASE

| Metrics | BLIINDS-II | BRISQUE | SSEQ | GMLF | NFERM | FSIM | GSI | IGM | VSI | PSIM | ADD-SSIM | ADD-GSIM | SQMS | NIQE | IL-NIQE | BQMS | SIQE | SSIQE | ASIQE |
|---------|-----------|---------|------|------|-------|------|-----|-----|-----|------|----------|----------|------|------|---------|------|------|-------|-------|
| Time (s) | 84.42 | 0.261 | 2.055 | 0.170 | 92.18 | 0.213 | 0.018 | 20.72 | 0.166 | 0.097 | 0.166 | 0.188 | 0.074 | 0.319 | 5.308 | 92.57 | **97.28** | **4.504** | **0.627** |
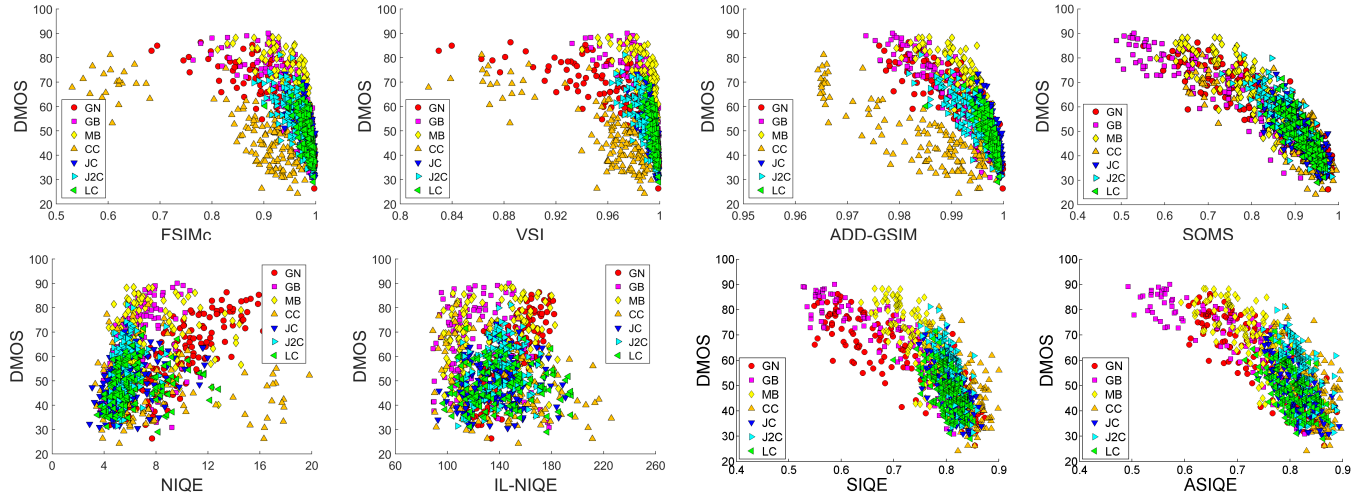


Fig. 12. Scatter plots of DMOS versus FR FSIMc, VSI, ADD-GSIM, SQMS, and NR NIQE, IL-NIQE, SIQE (proposed), ASIQE (proposed), all on the SIQAD database. GN: red; GB: magenta; MB: yellow; CC: orange; JC: blue; J2C: cyan; LC: green.

For convenience, we highlight the performance associated to the parameters used in the SIQE metric. From the results, we can derive two conclusions. First, the determined values for these two parameters lead to the optimal performance. Second, we can find that the performance is comparatively robust across those enumerated values. For varying $\kappa$ and fixed $\gamma = 8$, the worst performance corresponds to 0.8362, 0.8213 and 0.6291 in PLC, SRC and KRC. And for changing $\gamma$ and fixed $\kappa = 9$, the worst performance corresponds to 0.8121, 0.7986 and 0.6038 in PLC, SRC and KRC. The above-mentioned two worst performance results are both still superior to state-of-the-art competitors. Furthermore, we also notice that the PLC and SRC values grew close to 0.836 and 0.820 as $\gamma$ was increased, suggesting that we can remove $L_{LH,n}$ and $L_{HL,n}$, and only compute $L_{HH,n}$ in Eq. (16) for saving a small amount of computation.

*6) Runtime:* A graceful IQA technique is expected to not only deliver high correlation performance, but also to be computationally efficient. The runtimes of 19 competing quality measures were computed on the 980 distorted screen content images in the SIQAD database. The testing platform was based on MATLAB2015 running on a desktop computer with 16GB of internal memory and a 3.20GHz CPU processor. Table VII provides the average runtime of each IQA model. When assessing a corrupted screen content image of resolution about $700 \times 700$ by the proposed ASIQE metric, less than one second was required on average, which is an acceleration of more than 155 times over SIQE and 7.2 times over SSIQE.

*7) Scatter Plots:* We also examined the scatter plots of objective quality models against subjective opinion scores, as shown in Fig. 12. The algorithms studied include FR models: FSIMc, VSI, ADD-GSIM, SQMS, and the OU-NR models: NIQE, IL-NIQE, SIQE (proposed), and ASIQE (proposed). For each scatter plot, different colors distinguish

the sample points associated with different distortion types: red for GN, magenta for GB, yellow for MB, orange for CC, blue for JC, cyan for J2C, and green for LC. Generally, an effective general technique is able to accurately and uniformly predict image quality across different categories of distortions. As shown in Fig. 12, the scatter plots of the proposed SIQE and ASIQE models are quite consistent across distortion levels and types.

*8) Comparison With a CNN-Based Metric:* The convolutional neural network (CNN) model has been broadly used on many image processing and computer vision tasks, also including blind IQA [62], [63]. Hence we compared our model against the MultI-Channel CNN (MIC-CNN) model [63]. Since many samples are generally required to learn the CNN framework, we retrained the MIC-CNN model using the large-scale training data described earlier and derived a fixed CNN model to be tested on the SIQAD and QACS databases. The performance scores, in terms of PLC, SRC, KRC and RMS, of the MIC-CNN model were respectively 0.6794, 0.6558, 0.4632 and 12.767 on the SIQAD database and 0.3551, 0.3330, 0.2327 and 3.8704 on the QACS database. These results are inferior to those obtained using SIQE, SSIQE and ASIQE. The reason for this is likely that screen content images present different complexities than natural scene images, hence a deeper, better designed CNN network may be needed.

*9) Implementation:* Towards a more straightforward illustration of how to implement the proposed models, we used three screen content images from the SIQAD database as examples. These images were distorted from the same source image, as exhibited in Fig. 13. From Fig. 13(a) to 13(c), their DMOS values were respectively 43.88, 53.20 and 60.12, which means that their quality rank was (a) > (b) > (c). We implemented the proposed SIQE, SSIQE and ASIQE models on these sample images and derived the quality scores,

(a)        (b)        (c)

| Image index | DMOS | SIQE | SSIQE | ASIQE |
|:---:|:---:|:---:|:---:|:---:|
| (a) | 43.88 | 0.7892 | 0.7931 | 0.7838 |
| (b) | 53.20 | 0.7466 | 0.7544 | 0.7525 |
| (c) | 60.12 | 0.6625 | 0.6709 | 0.6867 |

(d)
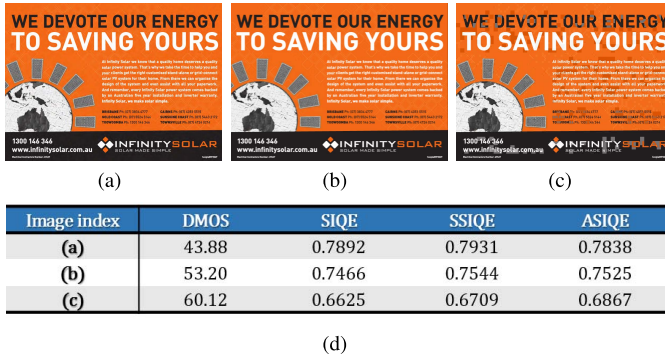
Fig. 13. Sample screen content images and their associated DMOS, SIQE, SSIQE and ASIQE scores.

as provided in Fig. 13(d). One can see that all three models generated faithful quality predictions consistent with subjective DMOS values.

## V. CONCLUSION

We have investigated an important and timely emerging research topic - quality evaluation of screen content images. Images of screen content typically involve virtual desktop applications and remote processing systems, which access remote computational resources as well as acquiring and managing remote data through the network. Unlike natural scene images, screen content images arise by a process of algorithmic generation and/or assembly. While natural scene images are generally rich in color, shape complexity and detail, screen content images often have limited color variation, and contain simple shapes and fine lines. Such differences render the design theories developed for assessing the quality of natural scene images less reliable. This paper provides some practical solutions to the screen content IQA problem.

To do so, we deploy four types of factors relevant to the quality assessment of screen content pictures. We use these to establish a general IQA framework based on big data training samples, and propose high-performance NR models to automatically evaluate the quality of screen content images. We made three main contributions:

1) We proposed a unified framework for blind IQA model design, which we used for screen content scene IQA but could easily be used to design IQA models for other types of images and distortion types. As depicted in Fig. 2, 'completely blind' quality models can be designed for, e.g., hybrid distortions or graphical images.
2) Based on the framework above, we developed an OU-NR-IQA method called SIQE. The features comprising SIQE have four aspects - image complexity, screen content statistics, global brightness and surface quality, and image sharpness and corners. The experimental results demonstrated the superiority of the model relative to state-of-the-art competitors. The training images were gathered by ourselves, with a wide range of image content. We hope that these pristine images help promote future screen content IQA studies. Training labels were generated using the FR SQMS model, which delivers excellent performance on existing quality assessment databases related to screen content pictures.

3) We introduced a method to accelerate the SIQE algorithm. By sacrificing a little performance, the implementation speed was greatly improved by a factor of more than 150.

In the future, we plan to focus on four research directions: 1) the reliable segmentation of pictorial and textual regions from a distorted screen content image; 2) subjective and objective assessment of enhanced screen content images generated by contrast improvement, brightness adjustment, interpolation, and more; 3) the development of universal IQA models that can faithfully evaluate the visual quality of natural scenes and screen content images simultaneously; 4) collection of more testing data, e.g. via online crowdsourcing subjective image quality assessment [30], [64], for the purpose of better designing and verifying the robustness of the objective screen content IQA models.

## REFERENCES

[1] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
[2] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
[3] 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.
[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.
[5] 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.
[6] 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.
[7] 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.
[8] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[9] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 506–510.
[10] 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.
[11] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[12] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.
[13] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[15] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Process., Image Commun.*, vol. 29, no. 7, pp. 725–747, Aug. 2014.
[16] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.
[17] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[18] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[19] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, pp. 140–149, Jul. 2016.

[20] 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.

[21] S. Wang *et al.*, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.

[22] 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.

[23] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[24] 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. 2014.

[25] 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.

[26] K. Gu, S. Wang, G. Zhai, S. Ma, and W. Lin, "Screen image quality assessment incorporating structural degradation measurement," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2015, pp. 125–128.

[27] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1693–1697.

[28] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sep. 2014.

[29] T. Goodall, A. C. Bovik, and N. G. Paulter, "Tasking on natural statistics of infrared images," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 65–79, Jan. 2016.

[30] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.

[31] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[32] 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.

[33] K. Gu *et al.*, "Blind quality assessment of tone-mapped images via analysis of information, naturalness and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.

[34] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3600–3604.

[35] 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.

[36] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4892–4904, Dec. 2013.

[37] X. Wu, G. Zhai, X. Yang, and W. Zhang, "Adaptive sequential prediction of multidimensional signals with applications to lossless image coding," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 36–42, Jan. 2011.

[38] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 836–846.

[39] C. F. Stromeyer and B. Julesz, "Spatial-frequency masking in vision: Critical bands and spread of masking," *J. Opt. Soc. Amer.*, vol. 62, no. 10, pp. 1221–1232, Oct. 1972.

[40] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vis. Res.*, vol. 22, no. 5, pp. 545–559, 1982.

[41] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, "No-reference image quality assessment metric by combining free energy theory and structural degradation model," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.

[42] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature*, vol. 447, pp. 206–209, May 2007.

[43] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

[44] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.

[45] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.

[46] X. Min *et al.*, "Blind quality assessment of compressed images via pseudo structural similarity," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2016, pp. 1–6.

[47] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 593–600.

[48] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. no. 27.

[49] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

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

[51] H. Attias, "A variational Bayesian framework for graphical models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12. 2000, pp. 209–215.

[52] J. Han, J. Pei, and M. Kamber. *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.

[53] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, Sep. 2014.

[54] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[55] 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.

[56] 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.

[57] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[58] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.

[59] ITU, "Methodology for the subjective assessment of the quality of television pictures," Int. Telecommun. Union Recommendation, Geneva, Switzerland, Tech. Rep. ITU-R BT.500-13, 2012.

[60] Q. Wu, H. Li, F. Meng, K. N. Ngan, and S. Zhu, "No reference image quality assessment metric via multi-domain structural information and piecewise regression," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 205–216, Oct. 2015.

[61] Q. Wu *et al.*, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.

[62] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE CVPR*, Jun. 2014, pp. 1733–1740.

[63] L. Zuo, H. Wang, and J. Fu, "Screen content image quality assessment via convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 2082–2086.

[64] Q. Xu, Q. Huang, and Y. Yao, "Online crowdsourcing subjective image quality assessment," in *Proc. 20th ACM Int. Conf. Multimedia*, Oct. 2012, pp. 359–368.

**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. His research interests include image quality assessment and air quality prediction. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He is the leading Organizer of special session in VCIP2016 and ICIP2017. He is currently the Associate Editor of the IEEE ACCESS and a Reviewer of the IEEE T-IP, T-NNLS, T-MM, T-CYB, T-IE, T-CSVT, T-BC, J-STSP, SPL, Access, *Information Sciences*, *Neurocomputi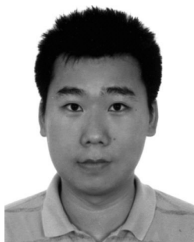ng*, MTAP, SPIC, JVCI, DSP, and ELL.