This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

Towards Accurate Quality Estimation of Screen Content Pictures with Very Sparse Reference Information

Zhifang Xia, Ke Gu, Shiqi Wang, *Member, IEEE*, Hantao Liu, and Sam Kwong, *Fellow, IEEE*

***Abstract*—The screen content (SC) pictures, such as webpages, serve as a visible and convenient medium to well represent the Internet information, and therefore the visual quality of SC pictures is highly significant and has attained a growing amount of attention. Accurate quality evaluation of SC pictures not only provides the fidelity of the conveyed information, but also contributes to the improvement of the user experience. In practical applications, a reliable estimation of SC pictures plays a considerably critical role for the optimization of the processing systems as the guidance. Based on these motivations, we propose a novel method for precisely assessing the quality of SC pictures using very sparse reference information. Specifically, the proposed quality method separately extracts the macroscopic and microscopic structures, followed by comparing the differences of macroscopic and microscopic features between a pristine SC picture and its corrupted version to infer the overall quality score. By studying the feature histogram for dimensionality reduction, the proposed method merely requires two features as the reference information that can be exactly embedded in the file header with very few bits. Experiments manifest the superiority of our algorithm as compared with state-of-the-art relevant quality metrics when applied to the visual quality evaluation of SC pictures.**

***Index Terms*—Screen content picture, sparse reference, quality estimation, macroscopic/microscopic structure**

I. Introduction

DUE to the massive expansion of smart phones which have dramatically revolutionized many fields, exquisite screen content (SC) pictures have been extensively employed as the medium for webpage browsing, online gaming and advertising. Compared with the ever-popular natural scene (NS) pictures [1], SC pictures simultaneously contain pictorial, textual and graphic contents. Such type of pictures can be broadly applied in industrial safety monitor-control and early-warning systems, as shown in Fig. 1. During the last few years, picture quality assessment (QA) has long been one attractive research topic, on account of its crucial applications to compression [2], [3], enhancement [4], [5], restoration [6]–[8], transmission [9], [10], etc. Generally speaking, the quality of a picture can be assessed in two types of methods, containing subjective assessment and objective assessment. The former one is decisive and generally regarded as the benchmark because it accurately judges the overall picture quality through human viewers. Nonetheless, subjective assessment inevitably causes time-consuming, expensive and labor-intensive issues, and therefore is unable to be widely applied in the real-time application scenarios. To assess the picture quality efficiently and effectively, many objective picture QA models were proposed based on low-level vision [11]–[13], brain theory [14], [15], statistics [16], [17], etc, for better mimicking the process of human visual perception.

Fig. 1: Example of an industrial safety monitoring-control system.

Nevertheless, the methods mentioned above are primarily applied for NS pictures. As compared with the NS pictures, the SC pictures have the features of thin lines, limited colors and diversified shapes, which lead to a series of challenging issues. In [18], [19], it is illustrated that those QA models for NS pictures fail to evaluate SC picture quality. Hence, a growing amount of attention has been focused on this emerging research field. More recently, objective assessment models using a few reference information were proposed. These models, e.g. RWQMS [20], RQMSC [21], PBM [22],

This work was supported in part by the National Science Foundation of China under Grant 61703009, 61527804, the Young Elite Scientist Sponsorship Program by China Association for Science and Technology under Grant 2017QNRC001, and the Young Top-Notch Talents Team Program of Beijing Excellent Talents Funding under Grant 2017000026833ZK40. (Corresponding author: Ke Gu.)

Z. Xia is with Beijing Advanced Innovation Center for Future Internet Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China, and also with the State Information Center of P.R.China, Beijing, China (email: spidergirl21@163.com).

K. Gu is with Beijing Advanced Innovation Center for Future Internet Technology, Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (email: guke.doctor@gmail.com).

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

H. Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, U.K. (e-mail: liuh35@cardiff.ac.uk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

can feasibly evaluate the quality of SC pictures even compared with the QA models with complete reference information.

As compared with NS pictures, current research of quality evaluation of SC pictures lies still in the infancy stage. On one hand, abundant reference information is still required (e.g., SPQA [18] and GDI [23]). On the other hand, the prediction monotonicity is far from satisfactory (e.g., RWQMS [20], RQMSC [21], PBM [22]). In order to circumvent the deficiencies, we devise a novel method for accurately estimating the quality of SC pictures using little information of the original picture. Motivated by many researches which indicate that incorporating the gradient magnitude information in picture QA performs well particularly in extracting structures [11], [12], [24]–[27], in our work, the gradient magnitude is fused respectively with the anisotropy measurement and the uncertainty information to detect the macroscopic and microscopic structures, such that the integral and detailed structures of a picture can be captured. Subsequently, we compare and combine the differences of macroscopic and microscopic structures between a corrupted SC picture and its corresponding lossless version to infer the overall quality score. Finally, we apply the histogram establishment method to reduce the dimensionality of features extracted. As such, minimum additional information is required to be transmitted to infer the visual quality, which is very practical for real application scenarios.

The main contributions of this research are summarized as follows. First, to the best of our knowledge, we propose a new effective quality assessment framework of SC pictures, which appropriately incorporates the measurements of macroscopic and microscopic structural variations. Based on this framework, our paper deploys the proper measurements to capture the variations of macroscopic and microscopic structures, and then fuses these two measurements to infer the overall quality prediction of the input SC picture. Second, our QA approach has achieved superior performance when comparing with the current state-of-the-art QA metrics for SC pictures quality evaluation. Third, the proposed QA model merely adopts very sparse reference information, i.e. two features, which can be exactly conveyed in the header file with very few bits.

The rest of this paper is organized as follows. In Section II, the current QA metrics and models are introduced. Section III presents the design philosophy and methodology of the quality model. In Section IV, the experimental setup, results, analysis and application are described and discussed. Finally, the concluding remarks are given in Section V.

II. Related Work

In the last few decades, many works have been committed to picture QA, thus plenty of picture QA metrics and models have been developed. In this section, we will review typical QA models of NS pictures and SC pictures respectively.

QA models of NS pictures. Based on the gradient magnitude, quite a few QA models were proposed to estimate the quality of NS pictures. Typical quality models are illustrated as follows. In [11], considering the fact that the visual perception understands a picture mainly according to its low-level features, Zhang *et al.* proposed the feature similarity (FSIM) method. In [12], by using the gradient similarity to estimate the changes in contrast and structure of pictures, Liu *et al.* proposed the gradient similarity measurement (GSM) method. In [25], by introducing visual saliency-based index, Zhang *et al.* proposed the visual saliency induced index (VSI) to gauge the picture quality in consistent with subjective assessment. In [28], Li *et al.* put forward the divisive normalization domain reduced-reference quality model (DNT-RR) on the basis of a divisive normalization image representation. In [29], with consideration of visual information fidelity, Wu *et al.* came up with the visual information fidelity based reduced-reference model (VIF-RR). In [30], based on a natural image statistic model in the wavelet transform domain, Wang *et al.* proposed the wavelet domain natural image statistic model (WNISM). In [31], Narwaria *et al.* designed the fourier transform-based scalable image quality metric (FTQM) to use the magnitude and phase of the 2-D discrete Fourier transform. In [55], by degrading the distorted image in several ways and to certain degrees, Min *et al.* built the blind multiple pseudo reference images-based (BMPRI) measure to calculate the similarities between the distorted images and multiple pseudo reference images. In [56], by utilizing a 'reference' called pseudo-reference image (PRI) and a PRI-based blind picture QA frame, Min *et al.* devised the Blind PRI (BPRI) to measure the similarity between the distorted image's and the PRI's structures.

QA models of SC pictures. A number of QA models of SC pictures have been proposed during recent years. In [18], Yang *et al.* proposed the screen content perceptual quality assessment model (SPQA) by combining the perceptual differences of pictorial and textual regions between a corrupted SC picture and its associated uncorrupted counterpart to infer the overall quality. In [23], Ni *et al.* put forward the gradient direction-based index (GDI) method by gauging the gradient direction in term of local information and applying a deviation-inspired model for pooling. In [20], the reduced-reference wavelet-domain quality measure of SC pictures (RWQMS) was developed with the large-scale training data to learn a set of features in the wavelet domain, according to the information content, the fluctuations of energy and the generalized spectral behavior. In [21], Wang *et al.* devised the reduced-reference quality measure of screen content pictures (RQMSC) from the viewpoint of SC picture quality perception. In [22], Jakhetiya *et al.* established the prediction backed model (PBM) by simultaneously fusing a perceptually relevant predictive model and distortion classification, for evaluating the quality of SC pictures.

III. Methodology

The design philosophy of the proposed model is to incorporate the macroscopic and microscopic structures together to infer the SC picture quality. In particular, the structure-based methods mainly depend on the supposition that the human visual system (HVS) is used for extracting structural information from a picture [34], and thus computing the similarity of structures is a promising strategy to approximate the picture

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

(a) Natural Scene pictures (b) Screen Content picture

(c) Macroscopic structure of (a) (d) Macroscopic structure of (b)

Fig. 2: Comparisons of the macroscopic structures of the natural scene and screen content pictures. (a) and (b) are pristine natural scene and screen content pictures; (c) and (d) are the maps of macroscopic structures of (a) and (b).

quality. Several classical algorithms have been developed by gauging the structural variations to well infer the quality of pictures [11], [12], [24]–[26]. However, these algorithms may fail when they are applied to the SC pictures mainly due to the reason that the macroscopic structures of NS and SC pictures are different. In particular, the macroscopic structures of NS pictures can be regarded as image contours. By contrast, those of SC pictures are concentrated in the textual regions, which contain abundant amount of semantic information, as shown in Fig. 2. With the above considerations, it is highly required to propose a novel QA framework for the quality estimation of SC pictures. Specifically, the macroscopic and microscopic structures are first extracted, along with the histogram established to dramatically reduce the dimensionality of reference information to only two features. As such, those two features can be used to measure the differences of the distorted and pristine SC pictures. A diagram of the proposed quality model is illustrated in Fig. 3.

A. Macroscopic Structure Measurement

In this research, a macroscopic structure measurement (MASM) is developed to extract the vital structures of a SC picture. The MASM model, on the basis of anisotropy, can be regarded as a directional metric, which is constructed by fusing the gradient and anisotropy measurements.

(1) Gradient magnitude. As discussed, the variations in structures tend to attract the perception of HVS [34]. The gradient magnitude which has been commonly used in the applications of computer vision and image processing algorithms such as contour inspection, optical flow estimation and picture segmentation, is used in this research to detect the structures in a picture. Specifically, this work computes the gradient magnitude of a picture with the Scharr operator [10], [11], [19], [35], [36], which can be calculated by the following two convolution masks:

$$G(S) = \sqrt{G_h^2 + G_v^2} \tag{1}$$

where

$$G_h = H_h \otimes S = \frac{1}{16}\begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix} \otimes S \tag{2}$$

$$G_v = H_v \otimes S = \frac{1}{16}\begin{bmatrix} +3 & +10 & +3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \otimes S \tag{3}$$

where S indicates a SC picture signal; H_h and H_v stand for the Scharr convolution masks along the horizontal and vertical directions. Based on the above convolution operation, the gradient magnitude can be highlighted as the structural information in textual and pictorial regions.

(2) Anisotropy measurement. The unequal visual sensation of stimuli that has congruent yet various directions can be easily caused by the intensity-changing structures [37]. Accordingly, it is assumed that the strong visual perception is prone to be measured by the macroscopic structures that have intensity variation as well as a preferred direction. Such macroscopic structures are regarded as the large anisotropy. By contrast, the structures with homogeneous scattering, which usually lead to less perception of structural changes, can be regarded to be with small anisotropy. Based on the analysis, the anisotropy measurement can serve as a straightforward way of extracting the macroscopic structures. In practice, the anisotropy measurement has been broadly explored in several pioneering works to potentially capture local heterogeneity of intensity variance [38]–[40]. In this research, the anisotropy measurement not only implies the distribution of pixel intensity of a SC picture but also indicates the basic directional variation in the local vicinity of a pixel [41]–[43].

From the perspective of structure tensor, we can represent the anisotropy measurement as a matrix generated from the gradient magnitude of a SC picture. More specifically, the structure tensor is formulated as:

$$T(i) = \begin{pmatrix} \sum_j \langle \nabla_h S_j, \nabla_h S_j \rangle & \sum_j \langle \nabla_v S_j, \nabla_h S_j \rangle \\ \sum_j \langle \nabla_h S_j, \nabla_v S_j \rangle & \sum_j \langle \nabla_v S_j, \nabla_v S_j \rangle \end{pmatrix} \tag{4}$$

where $j \in R(i)$ belongs to the vicinity of a pixel i with a predefined radius; ∇_h and ∇_v indicate the partial differential operators along the horizontal and vertical directions; $\langle \cdot, \cdot \rangle$ is the inner product of a pair of vectors. In mathematical form, $T(i)$ is a semi-positive definite symmetric 2×2 matrix, with two eigenvectors $\boldsymbol{\eta}_i^\star$ and $\boldsymbol{\eta}_i^*$ and the corresponding two eigenvalues $\lambda_i^\star$ and λ_i^* $(\lambda_i^\star \geq \lambda_i^* \geq 0)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

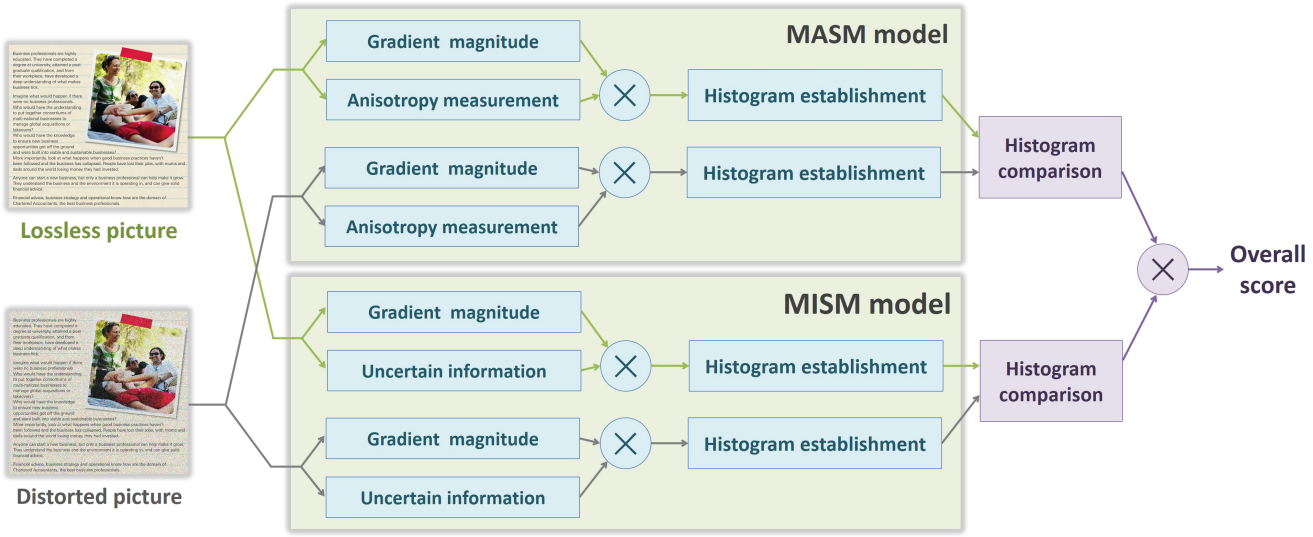


Fig. 3: Illustration of the workflow of the proposed QA method.

Accordingly, the anisotropy measurement can be defined as the relative distance between the two eigenvalues $\lambda_i^\star$ and λ_i^*,

$$A(i) = \frac{\lambda_i^\star - \lambda_i^* + \epsilon}{\lambda_i^\star + \lambda_i^* + \epsilon} \tag{5}$$

where ϵ is a small constant to avoid the division-by-zero case. Eqn. (5) points out that the minimum value and maximum value in A are zero and one. The anisotropy measurement A equals to or approximates the maximum value "one", when the structures have significant changes in the pixel intensity, namely, 1) $\lambda_i^\star \gg \lambda_i^*$; 2) $\lambda_i^\star > 0$ and $\lambda_i^* = 0$. By contrast, as for the structures that have the uniform direction, namely $\lambda_i^\star \approx \lambda_i^*$, A converges to the minimum value "zero". The inner product operation $\langle \boldsymbol{\eta}_i^\star, \boldsymbol{\eta}_i^* \rangle$, bounded from 0 to 1, serves to evaluate the difference of the two vectors in terms of the direction. In other words, the angle between the above two vectors ranges from 0 to π, and when a vector is orthogonal to the other, the value of inner product approaches 0.

The pixel of a SC picture with fine anisotropy is conveniently perceived in most cases, such that the anisotropy measurement supplies a better way to find the primary direction of a SC picture. In view of the above considerations, our MASM model can be defined as follows:

$$MASM(S) = G \cdot A \tag{6}$$

where G and A are yielded by Eqns. (1)-(3) and Eqns. (4)-(5), respectively.

To sum up, Fig. 2 describes the calculation of MASM by incorporating gradient magnitude and anisotropy measurement, as presented in Eqn. (6). As illustrated in Fig. 1, one can observe that the MASM model can be adequately applied in suppressing the minor textures, which are visually imperceptible. Meanwhile, it is also able to maintain the dominant structures with large anisotropy, as well as strengthen the edges of objects to detect the macroscopic structures from a SC picture.

B. Microscopic Structure Measurement

Recent studies, such as the Bayesian brain theory [44] and free-energy principle [45], have revealed that the HVS actively deduces the visual inputs using the internal generative mechanism. The primary visual information plays a particularly critical role in understanding and recognizing a picture, whereas the uncertain information, which can be caused by the gap between the real scene and its estimation from the brain, cannot be well explained by the HVS. Some early researches show that the picture QA is much related to the uncertain information, and toward qualifying the uncertain information, it is required to compare the reference with the orderly signals which are derived from the inference procedure of the brain. With these considerations, we introduce a microscopic structure measurement (MISM), which is mathematically defined by gauging the primary visual information and the uncertain information.

(1) Gradient magnitude. The computation of gradient magnitude can be found in the above subsection.

(2) Uncertain information. When a picture is perceived, the input visual signal enters the brain after passing through the HVS channel. During this process, the lens play the role of a strong low-pass filter which can eliminate specific high-frequency information [46], [47]. Essentially, the process of human visual perception can be approximately modeled as a low-pass filter [34]. By fully taking the particular features of the HVS and SC pictures into consideration, we can fuse the Gaussian and motion low-pass filters to measure the uncertainty information for effectively reflecting the artifacts around high-contrast edges [48], accounting for the viewing behavior of textual content, and validly measuring the uncertainty information produced during the phases of eye "fixation" and "saccade" [49], [50].

Specifically, a Gaussian filter can be computed by

$$H_g(p,q) = \frac{1}{2\pi\delta^2} \exp\left(-\frac{p^2+q^2}{2\delta^2}\right), \tag{7}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

(a) SC picture (b) Microscopic structure of (a)

Fig. 4: Illustration of microscopic structures of screen content pictures: (a) is a pristine screen content picture; (b) is the map of microscopic structures of (a).

where δ is the standard deviation that controls the smoothing strength. Subsequently, we can produce the Gaussian smoothed picture by convoluting it with the input SC picture S:

$$S_g = S \otimes H_g. \tag{8}$$

As such, the uncertain information is obtained by quantifying the gap between S and S_g. Owing to the properties of unique maximum, boundedness and symmetry, we utilize the normalized version of gradient similarity [26]:

$$GS_g = f(S, S_g) = \frac{(G(S) - G(S_g))^2}{G^2(S) + G^2(S_g)}. \tag{9}$$

The motion filter is expressed by

$$H_m(p,q) = \begin{cases} 1/t & \text{if } \Gamma(p,q,\phi) = 0, \Upsilon(p,q) \le t^2/4 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $\Gamma(p,q,\phi) = p\sin\phi + q\cos\phi$ with ϕ denoting a special direction of motion filter; $\Upsilon(p,q) = p^2 + q^2$; t indicates the amount of motion in pixels which is taken into account in the convolution procedure. The motion filtered picture can be produced by convoluting it with the input SC picture S:

$$S_m = S \otimes H_m. \tag{11}$$

In analogy to the Gaussian filters, the uncertain information derived from the motion blur is expressed as

$$GS_m = f(S, S_m) = \frac{(G(S) - G(S_m))^2}{G^2(S) + G^2(S_m)}. \tag{12}$$

Comprehensively, as for the SC picture S, we compute the amount of uncertain information via a simple direct average [19], as defined as follows:

$$U = \frac{1}{2}(GS_m + GS_g). \tag{13}$$

Finally, our MISM model is calculated by

$$MISM(S) = G \cdot U. \tag{14}$$

where G and U are derived by using Eqns. (1)-(3) and Eqns. (7)-(13) respectively. The maps derived by applying the MISM model are illustrated in Fig. 4. It can be found that the detailed structures are found and highlighted by comparison with the results of the MASM model shown in Fig. 2.

C. Overall Quality Measure

Based on the above-mentioned analysis, it is reasonable to properly incorporate the MASM and MISM models, in order to compensate the disadvantage of each component and achieve better prediction accuracy. More concretely, to noticeably promote the process of distortion comparison, we firstly distinguish the insignificant and significant structures by passing them through a non-linear mapping, the principle of which lies in that the insignificant structures corresponds to 0 and the significant ones are associated to 1. The significant and insignificant feature values can be achieved by using the psychometric function with the sigmoid shape [51]. This work adopts the Galton's ogive [52], which is in the formulation of a cumulative normal distribution function (CDF):

$$C(s) = \frac{1}{\sqrt{2\pi}\phi} \int_{-\infty}^{s} \exp\left[-\frac{(t-\kappa)^2}{2\phi^2}\right] dt \tag{15}$$

where $C(s)$ is the prediction probability density applied for distinguishing the insignificant and significant structures; κ is the modulation threshold; s is the stimulus amplitude; ϕ is the parameter controlling the slope of prediction probability variation. In this work we empirically assign ϕ as a fixed value of 0.05. By separately passing the maps of MASM and MISM models through the CDF, for an input SC picture, we can acquire its associated two significance maps.

Subsequently, in practice, we perform the above feature extraction on the pristine SC picture $\dot{S}$ and its corresponding corrupted version $\ddot{S}$. However, using the pristine picture as the reference information for quality evaluation leads to the significant burden of transmission, which is impractical. To overcome this difficulty, this work makes use of the histogram for signifying the distribution. Essentially, such solution aims to reach to a good tradeoff between the quality prediction performance and the reference information data rate. We first discuss how to apply the solution to the MASM model. In particular, we can divide its distribution range of $\Phi([d_{\min}, d_{\max}])$ into N equal-length gaps. The histogram bin depends on the number of elements by setting $\mathcal{W}_k$ as follows:

$$h_k = |\mathcal{W}_k|, \quad \mathcal{W}_k = \{w | \Phi(w) \in J_k\} \tag{16}$$

where

$$J_k = \left[d_{\min} + (k-1)\frac{\tilde{d}}{N}, d_{\min} + k\frac{\tilde{d}}{N}\right], \quad \tilde{d} = d_{\max} - d_{\min}. \tag{17}$$

The histogram of the lossless SC picture can be measured as

$$H_{\dot{S}}(k) = h_k \Big/ \sum_{l=1}^{N} h_l. \tag{18}$$

The identical operation is applied to the distorted picture $\ddot{S}$ to get $H_{\ddot{S}}(k)$. The value of each histogram bin is associated to the probability of the interval. The score of the MASM model is yielded by comparing the two histogram as follows:

$$Q_{\text{MASM}}(\dot{S}, \ddot{S}) = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{\min\{H_{\dot{S}}(k), H_{\ddot{S}}(k)\} + \varepsilon}{\max\{H_{\dot{S}}(k), H_{\ddot{S}}(k)\} + \varepsilon}\right) \tag{19}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

where $\min\{\cdot,\cdot\}$ and $\max\{\cdot,\cdot\}$ are used to find the minimum and maximum values from two values; ε is a tiny positive constant closed to zero, in order to prevent the denominator from zero when $\max\{H_{\dot{S}}(k), H_{\ddot{S}}(k)\}$ approaches zero; N is set to be 2 for simplicity and simultaneously for reducing the burden of transmission as much as possible. It is quite evident that the value of $Q_{\text{MASM}}(\dot{S},\ddot{S})$ ranges from $[0,1]$. The larger the $Q_{\text{MASM}}(\dot{S},\ddot{S})$ value, the better the quality of the input SC picture. Moreover, we employ the same process to the MISM map and derive its score $Q_{\text{MISM}}(\dot{S},\ddot{S})$.

Finally, the overall picture quality score can be attained by multiplying the scores of the MASM and MISM models, i.e. $Q_{\text{MASM}}(\dot{S},\ddot{S})$ and $Q_{\text{MISM}}(\dot{S},\ddot{S})$:

$$Q(\dot{S},\ddot{S}) = Q_{\text{MASM}}(\dot{S},\ddot{S}) \cdot Q_{\text{MISM}}(\dot{S},\ddot{S})^{\alpha} \tag{20}$$

where α is an exponential parameter for adjusting the significances of two terms. Here we suppose the two terms have the equivalent importance and thus simply set α as 1.

IV. Experiments and Analysis

In this section, we aim to validate the performance of the proposed method from the perspectives of prediction accuracy, monotonicity and consistency. We first set up the experimental protocols by introducing the quality measure, testing database and evaluation criteria, and then the experimental results including performance, statistical significance, visualized comparisons and application are provided.

A. Experimental Setting

(1) Quality Measures. In this work, we have collected 15 state-of-the-art quality metrics, which can be divided into the following five categories according to their application scenarios. The first category is composed of three typical algorithms based on gradient magnitude and the whole information of the lossless picture, including FSIM [11], GSM [12], and VSI [25]. The second category includes five algorithms using only partial reference information, i.e. DNT-RR [28], VIF-RR [29], WNISM [30], FTQM [31], and SDM [53]. The third category consists of two referenceless picture quality assessment methods, i.e. BMPRI [55], BPRI [56]. The fourth category is made up of two recently devised metrics specific for quality evaluation of SC pictures, i.e. SPQA [18] and GDI [23]. The last category is composed of three QA algorithms, including RWQMS [20], RQMSC [21] and PBM [22], which were devoted to the SC picture quality assessment under the condition of partial reference information available.

(2) Testing Database. In order to verify the performance of the proposed QA model, we use the screen image quality assessment database (SIQAD) [18], which includes the most diversified types of distortions. This database was established by Nanyang Technological University in 2015. It is composed of 20 reference SC pictures and 980 distorted counterparts. Those 980 distorted pictures are generated by applying seven distortion types with seven grades to the 20 reference SC pictures. The seven types of distortions are contrast change, Gaussian blur, motion blur, Gaussian noise, layer segmentation based coding, JPEG compression, and JPEG2000 compression, respectively. When building the SIQAD database, beyond 20 participants graded the 980 pictures at the observing distance of 2.25 time the screen height away from the display device. The quality score has the minimum and maximum values of 0 and 10 with the interval of 1. The differential mean opinion score (DMOS) of each picture in the SIQAD database is normalized to $[24.2, 90.1]$.

(3) Evaluation Criteria. Using the SIQAD database as a testing platform, we compare the proposed QA metric with the above-mentioned 15 models by calculating the similarities between the objective scores and the corresponding subjective scores of images. The subjective opinion scores were derived by conducting subjective assessment experiments to collect the opinions of human viewers. The subjective scores of human viewers are generally considered as the most accurate measures. Based on those subjective scores, the correlation in terms of five evaluation criteria can be calculated and used for the performance comparison. Three of the five indices are Kendall's rank correlation coefficient (KRC), Spearman rank correlation coefficient (SRC) and Pearson linear correlation coefficient (PLC), which have the monotonically increasing relationship with the prediction performance. More specifically, the PLC, as a commonly used index in picture QA, is measured to gauge the strength of linear association between two variables:

$$PLC = \frac{\sum_{l=1}^{M}(\mu_l - \bar{\mu})\cdot(\nu_l - \bar{\nu})}{\sqrt{\sum_{l=1}^{M}(\mu_l - \bar{\mu})^2 \cdot \sum_{l}(\nu_l - \bar{\nu})^2}} \tag{21}$$

where M stands for the total number of the SC pictures; μ_l is the mapped objective score for the l-th picture by using a logistic regression function [54] and ν_l is the corresponding subjective score. $\bar{\mu}$ and $\bar{\nu}$ represent the mean score of μ and ν over the testing set. The SRC, characterized by estimating the evaluation monotonicity, is a very useful tool for comparative data analysis in picture QA and is computed as follows:

$$SRC = 1 - \frac{6\sum_{l=1}^{N}(\mu_l - \nu_l)^2}{M(M^2-1)}. \tag{22}$$

The KRC, aiming to estimate the degree of similarity between two sets of ranks given to a same set of objects, is formulated as

$$KRC = \frac{2(M_c - M_d)}{M(M-1)} \tag{23}$$

where M_c and M_d respectively stand for the number of concordant and discordant pairs in the database. The other two criteria, i.e. mean absolute error (MAE) and root mean-squared error (RMS), have the monotonically decreasing relationship with the prediction accuracy, and they are defined by

$$MAE = \frac{1}{M}\sum_{l=1}^{M}\Big|\mu_l - \nu_l\Big|, \tag{24}$$

$$RMS = \sqrt{\frac{1}{M}\sum_{l=1}^{M}(\mu_l - \nu_l)^2}. \tag{25}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

TABLE I: Comparisons of 15 state-of-the-art QA algorithms from the five evaluation criteria on the SIQAD database.

Index	Type I			Type II					Type III		Type IV		Type V			Pro.
	FSIM	GSM	VSI	DNT-RR	**VIF-RR**	WNISM	FTQM	SDM	BMPRI	**BPRI**	SPQA	**GDI**	**RWQMS**	RQMSC	PBM	
	[11]	[12]	[25]	[28]	**[29]**	[30]	[31]	[53]	[32]	**[33]**	[18]	**[23]**	**[20]**	[21]	[22]	
KRC	**0.4253**	0.4054	0.3874	0.3615	**0.4431**	0.3540	0.3268	0.4322	0.2283	**0.4040**	0.6803	**0.6486**	**0.5835**	0.5756	0.5280	**0.6255**
SRC	**0.5824**	0.5483	0.5381	0.5054	**0.6082**	0.5188	0.4575	0.6020	0.3325	**0.5612**	0.8416	**0.8436**	**0.7815**	0.7655	0.7168	**0.8213**
PLC	**0.5906**	0.5686	0.5568	0.5291	**0.5758**	0.5857	0.4691	0.6034	0.3737	**0.5863**	0.8584	**0.8515**	**0.8103**	0.8014	0.7264	**0.8343**
MAE	**9.0066**	9.1660	9.2875	9.7913	**9.5197**	9.4566	10.132	9.0139	10.693	**8.9966**	5.7890	**5.9744**	**6.8200**	6.8021	7.7632	**6.3196**
RMS	**11.551**	11.775	11.890	12.147	**11.703**	11.602	12.641	11.414	13.277	**11.596**	7.3421	**7.5055**	**8.3892**	8.5620	9.8375	**7.8924**

TABLE II: Statistical significance comparison of the proposed metric and competing QA models with the f-test on SIQAD.

F-test	FSIM	GSM	VSI	DNT-RR	VIF-RR	WNISM	FTQM	SDM	BPRI	BMPRI	SPQA	GDI	RWQMS	RQMSC	PBM
Pro.	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	0	0	+1	+1	+1

B. Performance

(1) Performance Comparison. Table 1 tabulates the performance result of our proposed quality method as well as 15 popular or state-of-the-art picture QA metrics on the testing SIQAD database. The KRC, SRC, PLC, MAE and RMS of our model have attained very impressive performance, which are 0.6255, 0.8213, 0.8343, 6.3196 and 7.8924, respectively. The top method in each of the five types mentioned above is highlighted in boldface towards a straightforward comparison. It is obvious that our proposed quality model, in light of the five evaluation criteria, has delivered the decent performance compared to those 15 picture QA metrics.

More concretely, it was found that our model has obtained 47.07%, 41.02%, 41.26% relative performance improvements in terms of KRC, SRC and PLC beyond the best method in the first type. By comparison with the top method in the second type, our QA model has derived the relative performance gains of 41.16% in KRC, 35.04% in SRC and 44.89% in PLC. In a similar way, our model improves the performance by 54.83%, 46.35%, 42.30% in KRC, SRC and PLC on the basis of the best method in the third type. From Table I, one can see that our approach is slightly inferior to the SPQA and GDI models in the fourth type. However, our quality method employs very sparse reference information (only two features) while the SPQA and GDI models require the whole lossless SC picture. Moreover, we compare the proposed method with the highest-performance QA model in the fifth type, and the relative performance gain is 7.20%, 5.09% and 2.96%. It is worth mentioning that the last type of these methods is devoted to the quality assessment of SC pictures with partial information of the pristine SC picture.

Furthermore, the proposed picture QA model is also compared with each of its two components, which are the scores from the MASM and MISM models $Q_{\text{MASM}}(\dot{S}, \ddot{S})$ and $Q_{\text{MISM}}(\dot{S}, \ddot{S})$. For simplicity, we merely focus on the KRC, SRC and PLC indices as well. It is observed that the values of the MASM model are respectively 0.3442 for KRC, 0.4934 for SRC and 0.5114 for PLC, and those of the MISM model are 0.5603, 0.7532 and 0.7763. Apparently, the proposed QA metric has delivered better prediction accuracy, leading to 63% and 12% relative performance gains than the solely MASM model and MISM model.

(2) Statistical Significance. The f-test is a statistically meaningful way to examine the statistical significance of metrics. In the f-test, the ratio between two prediction residual variances denoted by V is employed to compare with the critical threshold denoted by V_{ct}. When $V > V_{ct}$, we can have "+1", which indicates that one metric is significantly superior to the other one. When $V = V_{ct}$, the number "0" is acquired, which indicates that one metric is no worse than the other. When $V < V_{ct}$ which corresponds to "-1", it is illustrated that one metric is significantly worse than the other one. With f-test, we find that our QA model is statistically better than any of metrics in the first, second, third and fifth types. When our quality model is compared with the two metrics in the fourth type, which requires the completed information of the reference SC picture, the statistical significance result is "0". This means that our model is statistically matchable with the SPQA and GDI models. As depicted in Table II, our QA algorithm is statistically superior to or comparable to all the QA models tested.

(3) Visualized Comparison. To further illustrate the differences of those methods tested, we provide the scatter plots between the human ratings versus the objective quality scores of 12 testing picture QA models on the SIQAD database, as illustrated in Fig. 5. The algorithms to be compared include FSIM, GSM, VSI, PSIM, FTQM, SDM, RWQMS, RQMSC, PBM, MASM, MISM and our QA algorithm. The consistency across distinct types of distortions is a well-established indicator to access the effectiveness of a picture QA method. As shown in Fig. 5, the scatter distributions of sample points derived from our model are more consistent among the seven distortion types. As for the other metrics, the scatter distributions associated to different distortion types are far away from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

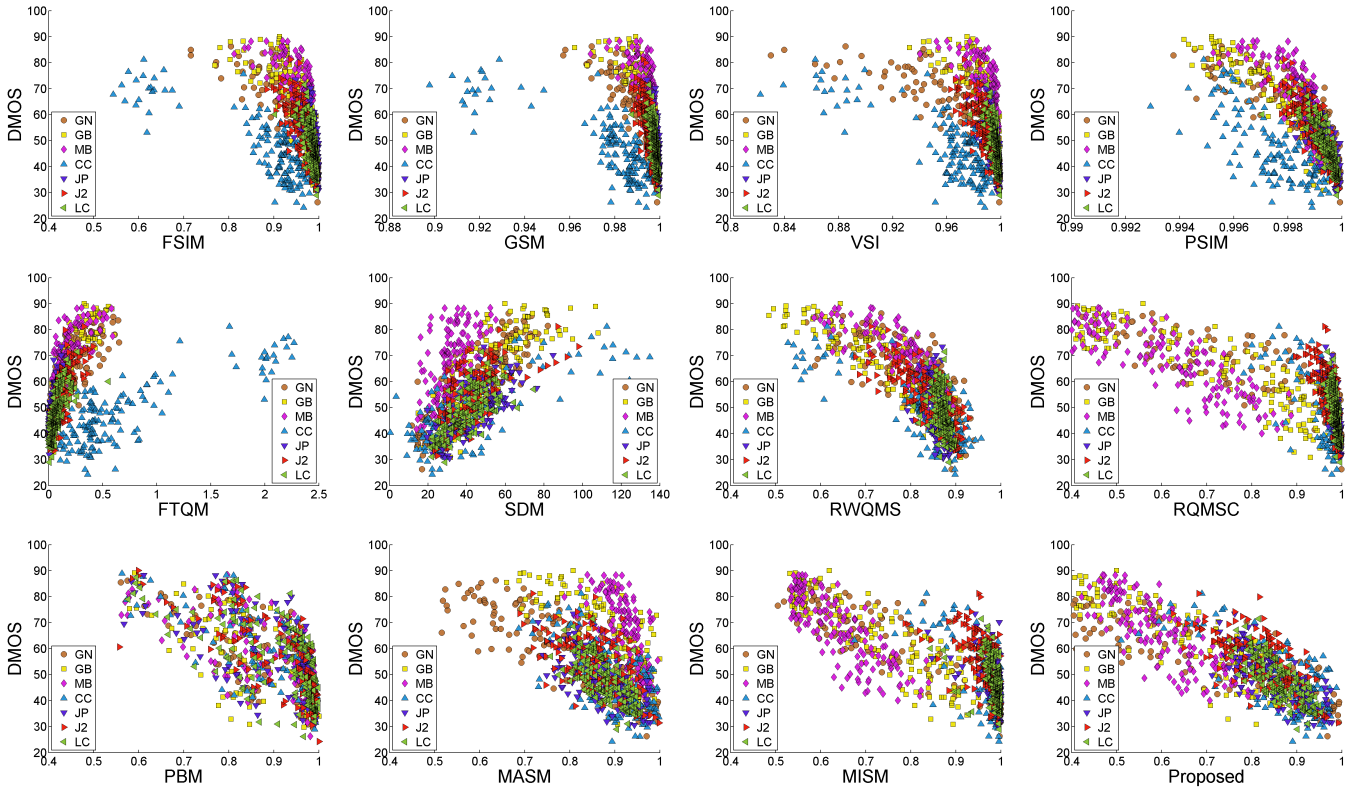


Fig. 5: Scatter plots of the DMOS values versus the predictions of FSIM, GSM, VSI, PSIM, FTQM, SDM, RWQMS, RQMSC, PBM, MASM, MISM and our proposed model on the SIQAD database. GN: Gaussian noise; GB: Gaussian blur; MB: motion blur; CC: contrast change; JP: JPEG compression; J2: JPEG2000 compression; LC: Layer segmentation-based coding.

TABLE III: Comparisons of our model with five types of typical QA models on the SCID database.

Index	Type I : FSIM [11]	Type II : VIF-RR [29]	Type III : BPRI [33]	Type IV : GDI [23]	Type V : RWQMS [20]	Proposed
KRC	0.5362	0.4596	0.3842	0.5475	0.5227	0.5633
SRC	0.7384	0.6546	0.5484	0.7546	0.7244	0.7627
PLC	0.7539	0.6873	0.5949	0.7660	0.7402	0.7737
MAE	7.7652	8.6364	9.1471	7.4845	7.6597	7.2478
RMS	9.2944	10.275	11.371	9.0932	9.5109	8.9629

TABLE IV: Statistical significance comparison on SCID.

F-test	FSIM	VIF-RR	BPRI	GDI	RWQMS
Proposed	+1	+1	+1	0	+1

each other. Based on the above comparisons, the scatter plot of our proposed metric is of good consistency and more robust among distinct types of distortions. As a consequence, we can draw the conclusion that in terms of consistency, our model is better than the compared QA methods as well as the two components (MASM and MISM models).

C. Cross Validation

We have also introduced one more database (named as SCID) to corroborate the effectiveness of our model [55]. The SCID database is, to the best of our knowledge, the largest-size screen content picture database. It is composed of up to 40 lossless SC pictures and multiple distortion types, including Gaussian noise, contrast change, color saturation change, HEVC-based screen content compression, JPEG compression, JPEG2000 compression, Gaussian blur and motion blur. There include five distortion intensities in each distortion type. The performance of our proposed quality method is tabulated in Table III. It can be found that our QA model has achieved impressive results. In particular, it attains 0.5633, 0.7627, 0.7737, 7.2478 and 8.9629 of KRC, SRC, PLC, MAE and RMS, respectively.

We then carry out the numerical performance comparisons between our algorithm with five types of typical QA models

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2905831, IEEE Transactions on Industrial Electronics

(a) QP: 30; prediction score: 43.3; bpp: 0.0727 (b) QP: 40; prediction score: 42.6; bpp: 0.0839 (c) QP: 50; prediction score: 41.0; bpp: 0.0945

(d) QP: 60; prediction score: 39.4; bpp: 0.1059 (e) QP: 70; prediction score: 37.8; bpp: 0.1240 (f) QP: 52; prediction score: 39.6; bpp: 0.0966

Fig. 6: Compressed screen content pictures with different QP values and their corresponding predicted quality scores and bbp values.

on the SCID database. Those five models are FSIM, VIF-RR, BPRI, GDI and RWQMS, respectively. One can easily see that the proposed QA model performs better than all the five models considered. It is worthy to emphasize the following two points. One is that the relative performance gain between our method and its same type of RWQMS are separately 7.77%, 5.29% and 4.53% in terms of KRC, SRC and PLC. The second point is that we can also find that the proposed QA model is even superior to the state-of-the-art GDI, which requires the whole information of lossless screen content picture. The f-test is further conducted to make the comparison between our proposed quality method with each of the above computing QA models, as shown in Table IV. As seen, our method is comparable to the GDI while outmatches the other four models. In summary, our QA model works fairly well not only on the popular SIQAD database, but also on the largest-scale SCID database.

D. Application

As suggested by ITU-R BT.500-13 [56], the visual perception can be divided into five levels in the subjective picture quality assessment: "excellent", "good", "fair", "poor" and "bad". In general, the quality of a satisfied picture should be no worse than "good"; otherwise, the lossy picture will make the human uncomfortable and even convey wrong semantic information. According to the range of DMOS used in the SIQAD database, a satisfied picture's DMOS should be equal or less than 40. In such case, the quantization parameter (QP) is expected to be as small as possible, in order to save network bandwidth and storage resources.

We can convert the above statement to be an optimization problem:

$$\begin{aligned} &\arg\min \quad QP \\ &\text{s.t.} \quad Q(\dot{S}, \ddot{S}) <= 40. \end{aligned} \tag{26}$$

Our proposed QA model can serve as an optimization criterion for the purpose of automatically finding the least QP under the condition of guaranteeing the picture of "excellent" or "good" quality. Taking the picture shown in Fig. 1 for example, we solve the above optimization problem to derive that the best QP value is 52. Its corresponding compressed picture is shown in Fig. 6(f), and the associated prediction score and bit per pixel (bpp) are 39.6 and 0.0966, respectively. One can easily see that the compressed picture optimized by our proposed QA model meets the requirement of guaranteeing the picture of "excellent" or "good" quality.

Furthermore, we use five typical QP values to compress the picture in Fig. 1 for comparison. Their compressed versions are shown in Figs. 6(a)-(e). Their prediction scores are 43.4, 42.6, 41.0, 39.4 and 37.8 in order. Their bpp values are 0.0727, 0.0839, 0.0945, 0.1059 and 0.1240 in order. Obviously, the quality scores of compressed pictures in Figs. 6(a)-(c) are unsatisfied. Next, comparing Figs. 6(d)-(f), the compressed picture optimized by our proposed QA model is the best since it has attained the smallest bbp values, or in other words, consumed the least network bandwidth and storage resources.

V. Conclusion

In this paper we have elaborately developed a new quality evaluation model for automatically assessing the quality of SC picture. From the complementary perspectives, macroscopic and microscopic structures are found to reliably capture the variations of SC pictures after quality degradation. As such, the overall quality score of a corrupted SC picture is derived by systematically incorporating the measurements of variations occurred in the macroscopic and microscopic structures. Based on the established histogram for dimensionality reduction with the proposed QA model, only two features serving as the very sparse reference information are required to be transmitted, which are ignorable when compared to the entire compressed

SC picture. In practical applications, we are able to accurately embed the two features in the file header during transmission. Experimental results on the SIQAD and SCID database demonstrate the effectiveness of the proposed QA model by comparing with the state-of-the-art picture quality assessment metrics.

REFERENCES

[1] Kodak Lossless True Color Image Suite: http://r0k.us/graphics/kodak/

[2] 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

[3] L. Xu, S. Kwong, Y. Zhang, and D. Zhao, "Low-complexity encoder framework for window-level rate control optimization," *IEEE Trans. Ind. Electron.*, vol. 60, no. 5, pp. 1850-1858, May 2013.

[4] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480-1494, Sep. 2015.

[5] 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. Learning Syst.*, vol. 29, no. 4, pp. 1301-1313, Apr. 2018.

[6] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116-3132, Dec. 2010.

[7] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, pp. 4695-4708, vol. 21, no. 12, Dec. 2012.

[8] B. H. Chen, S. C. Huang, and S. Y. Kuo, "Error-optimized sparse representation for single image rain removal," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6573-6581, Aug. 2017.

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

[10] M. H. Pinsonm, L.-K. Choi, and A. C. Bovik, "Temporal video quality model accounting for variable frame delay distortions," *IEEE Trans. Broadcasting*, vol. 60, no. 4, pp. 637-649, Dec. 2014.

[11] 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.

[12] 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.

[13] G. Yue, C. Hou, K. Gu, S. Mao, and W. Zhang, "Biologically inspired blind quality assessment of tone-mapped images," *IEEE Trans. Ind. Electron.*, pp. 2525-2536, vol. 65, no. 3, Mar. 2018.

[14] 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.

[15] 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.

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

[17] K. Gu, J. Zhou, J. Qiao, G. Zhai, W. Lin, and A. C. Bovik,"No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005-4018, Aug. 2017.

[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, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1-13, Jun. 2016.

[20] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma, and W. Gao, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. T. Circuits Syst.*, vol. 6, no. 4, pp. 532-543, Dec. 2016.

[21] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 1-14, Jan. 2018.

[22] V. Jakhetiya, K. Gu, W. Lin, Q. Li, S.-P. Jaiswal,"A Prediction Backed Model for Quality Assessment of Screen Content and 3-D Synthesized Images," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 652-660, Feb. 2018.

[23] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Process. Lett.*, vol.23, no. 10, pp. 1394-1398, Oct. 2016.

[24] 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.

[25] 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.

[26] 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.

[27] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-14, Oct. 2018.

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

[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, Jul. 2013.

[30] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *in Proc. SPIE Human Vision and Electronic Imaging X*, vol. 5666, pp. 149-159, Mar. 2005.

[31] 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.

[32] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind Image Quality Estimation via Distortion Aggravation," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508-517, Mar. 2018.

[33] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2049-2062, Aug. 2018.

[34] 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.

[35] B. Jähne, H. Haubecker, and P. Geibler, Handbook of Computer Vision and Applications. *New York: Academic*, 1999.

[36] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, and W. Lin, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5462-5474, Nov. 2017.

[37] G. Kanizsa and G. Kanizsa, Organization in Vision: *Essays on Gestalt Perception.* New York, NY, USA: Praeger, 1979.

[38] B. Jähne, *Spatio-Temporal Image Processing: Theory and Scientific Applications*, vol. 751. Berlin, Germany: Springer-Verlag, 1993.

[39] G. Medioni, M. S. Lee, and C. K. Tang, *A Computational Framework for Feature Extraction and Segmentation.* Amsterdam, The Netherlands: Elsevier, 2000.

[40] G. Z. Yang, P. Burger, D. N. Firmin, and S. R. Underwood, "Structure adaptive anisotropic image filtering," *Image Vis. Comput.*, vol. 14, no. 2, pp. 135-145, Mar. 1996.

[41] H. Kang, S. Lee, and C. K. Chui, "Flow-based image abstraction," *IEEE Trans. Vis. Comput. Graphics.*, vol. 15, no. 1, pp. 62-76, Jan. 2009.

[42] J. E. Kyprianidis and H. Kang,"Image and video abstraction by coherence-enhancing filtering," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 593-602, 2011.

[43] W. Föstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and circular features," *in Proc. ISPRS Intercommission Conf. Fast Process. Photogramm. Data*, pp. 281-305, 1987.

[44] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *Trends in Neurosciences*, vol. 27, no. 12, pp. 712-719, 2004.

[45] K. Friston, "The free-energy principle: a unified brain theory," *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127-138, 2010.

[46] R. G. Boothe, *Perception of the visual environment.* Springer Science & Business Media, 2001

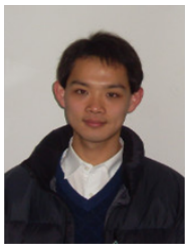
[47] D. Pamplona, J. Triesch, and C. Rothkopf, "Power spectra of the natural input to the visual system," *Vision research*, vol. 83, pp. 66-75, 2013.

[48] G. Zhai, A. Kaup, J. Wang, and X. Yang, "Retina model inspired image quality assessment," *in Proc. IEEE Vis. Commun. Image Process.*, pp. 1-6, Dec. 2013.

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

[50] A. Thiele, P. Henning, M. Kubischik, and K.-P. Hoffmann, "Neural mechanisms of saccadic suppression," *Science*, vol. 295, pp. 2460-2462, Nov. 2002.
[51] J. P. Guilford, "Psychometric methods." 1954.
[52] P. G. Barten, *Contrast sensitivity of the human eye and its effects on image quality.* SPIE press, vol. 72, 1999
[53] 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 and Syst.*, pp. 1095-1098, May 2013.
[54] H. R. Sheikh, M. F. Sabir, and A. C. 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.
[55] Z. Ni, L. Ma, H. Zeng, Y. Fu, L. Xing, and K.-K. Ma, "SCID: A database for screen content images quality assessment," *IEEE Intelligent Signal Processing and Communication Systems*, pp. 774-779, 2017. http://smartviplab.org/pubilcations/SCID/SCID.zip
[56] *Methodology for the Subjective Assessment of the Quality of Television Pictures.* ITU-R Recommendation Standard BT.500-13, 2012.

Zhifang Xia received the B.S. degree in measuring and control instrument from Anhui University, Hefei, China in 2008 and received the Master degree in control science and engineering from Tsinghua University, Beijing, China in 2012. She is currently an engineer and a registered consultant(investment) with State Information Center, Beijing, China. Her interests include image processing, quality assessment, machine learning and e-government. She won the second prize of National excellent engineering consultation award in 2016.



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 currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE Transactions on Multimedia (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017, and serves as a Guest Editor for the Digital Signal Processing Journal. He is currently an Associate Editor for the IEEE ACCESS and IET Image Processing (IET-IP), and an Area Editor for the Signal Processing Image Communication (SPIC). He is a Reviewer for 20 top SCI journals.

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from the Peking University, in 2014. From Mar. 2014 to Mar. 2016, he was a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From Apr. 2016 to Apr. 2017, he was with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed over 40 technical proposals to ISO/MPEG, ITU-T and AVS standards. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is serving for the IEEE MMTC, as the Chair of the Interest Group on Quality of Experience for Multimedia Communications, and he is an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Transactions on Multimedia.

Sam Kwong (F'13) received the B.Sc. and MA.Sc. degrees in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA and University of Waterloo, Waterloo, ON, Canada, in 1983 and 1985, respectively. In 1996, he received the Ph.D. degree from the University of Hagen, Hagen, Germany. From 1985 to 1987, he was a Diagnostic Engineer with the Control Data Canada, where he designed the diagnostic software to detect the manufacture faults of the VLSI chips in the Cyber 430 machine. He later joined the Bell Northern Research Canada as a member of the scientific staff. In 1990, he joined the City University of Hong Kong, Hong Kong, as a Lecturer with the Department of Electronic Engineering. He is currently a Professor with the Department of Computer Science. His research interests include pattern recognition, evolutionary algorithms, and video coding.