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/TBDATA.2019.2895605, IEEE Transactions on Big Data

Learning a Unified Blind Image Quality Metric via On-Line and Off-Line Big Training Instances

Ke Gu, Xin Xu, Junfei Qiao, Qiuping Jiang, Weisi Lin, *Fellow, IEEE*, and Daniel Thalmann

Abstract—In this work, we resolve a big challenge that most current image quality metrics (IQMs) are unavailable across different image contents, especially simultaneously coping with natural scene (NS) images or screen content (SC) images. By comparison with existing works, this paper deploys on-line and off-line data for proposing a unified no-reference (NR) IQM, not only applied to different distortion types and intensities but also to various image contents including classical NS images and prevailing SC images. Our proposed NR IQM is developed with two data-driven learning processes following feature extraction, which is based on scene statistic models, free-energy brain principle, and human visual system (HVS) characteristics. In the first process, the scene statistic models and an image retrieve technique are combined, based on on-line and off-line training instances, to derive a novel loose classifier for retrieving clean images and helping to infer the image content. In the second process, the features extracted by incorporating the inferred image content, free-energy and low-level perceptual characteristics of the HVS are learned by utilizing off-line training samples to analyze the distortion types and intensities and thereby to predict the image quality. The two processes mentioned above depend on a gigantic quantity of training data, much exceeding the number of images applied to performance validation, and thus make our model's performance more reliable. Through extensive experiments, it has been validated that the proposed blind IQM is capable of simultaneously inferring the quality of NS and SC images, and it has attained superior performance as compared with popular and state-of-the-art IQMs on the subjective NS and SC image quality databases. The source code of our model will be released with the publication of the paper at https://kegu.netlify.com.

Index Terms—Image quality metric (IQM), natural scene (NS) image, screen content (SC) image, no-reference (NR), data-driven process, big data learning, on-line, off-line

1 Introduction

Through wire transmission or wireless transmission, ubiquitous networks have largely shortened the interpersonal distances. For example, due to the explosive growth of science and technology, our communication with each other becomes more prompt, convenient and comfortable with the equipments developed from wire phones to wireless cellulars, and further to the video telephones. On the other hand, with the emergence of mighty search engines, people are capable of effectively and efficiently finding and obtaining the required information by text search or even by image search based on popular image retrieval technologies [1]. In this paper, we propose a novel solution to resolve a difficulty in the research of blind image quality assessment by turning to the networks and search engines.

The study of image quality metrics (IQMs), especially no-reference (NR)/blind IQMs, has long attracted a wide range of attention owing to its two aspects of pivotal functions: 1) monitoring the quality degradation caused by image acquisition, compression, transmission and exhibition [2]-[5]; 2) optimizing the parameters utilized in image processing systems such as denoising, deblurring and enhancement [6]. Most of present blind IQMs were applied to two types of application scenarios. The first type of models is the distortion-specific quality metrics of natural scene (NS) images, typically applied for compression [7]-[8], blurriness [9]-[10], tone mapping [11]-[12], 3D synthesization [13]-[14], etc, and the second type of models is the general-purpose NS IQMs [15]-[25]. Mathematically, the above two types of NR IQMs concentrate on solving the subsequent optimization problem by finding the optimal parameter vectors:

$$\boldsymbol{\theta}_t^*, \boldsymbol{\theta}_l^* = \arg\min_{\boldsymbol{\theta}_t, \boldsymbol{\theta}_l} \left\| \mathcal{Q}(\mathbf{x}) - \mathcal{P}(\mathbf{x}; \boldsymbol{\theta}_t, \boldsymbol{\theta}_l) \right\|_{\mathcal{D}} \quad (1)$$

where $\mathbf{x}$ is an image; $\mathcal{Q}$ and $\mathcal{P}$ are truth and predicted quality scores; $\|\cdot\|_{\mathcal{D}}$ computes the distance of two terms contained based on the $\mathcal{D}$ measure; $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_l$ are two parameter vectors to be determined, one each for differentiating the distortion types and levels. In general, a general-purpose NR IQM with high efficacy and efficiency is much more desirable.

- *Manuscript received xxxx, 2018; revised xxxx, 2018; accepted xxxx, 2018. This work was supported in part by the National Science Foundation of China under Grant 61527804, 61703009, the Beijing Advanced Innovation Center for Future Internet Technology under Grant 110000546619001, the Young Elite Scientist Sponsorship Program by China Association for Science and Technology under Grant 2017QNRC001, and the Nova Programme Interdisciplinary Cooperation Project under Grant Z161100004916041. (Corresponding author: Ke Gu.)*

- *Ke Gu, Xin Xu, and Junfei Qiao are with Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing Advanced Innovation Center for Future Internet Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (email: guke.doctor@gmail.com; 18813188655@emails.bjut.edu.cn; junfeiq@bjut.edu.cn).*
- *Qiuping Jiang is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (email: jqp910707@126.com).*
- *Weisi Lin is with School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798 (email: wslin@ntu.edu.sg).*
- *Daniel Thalmann is with EPFL, CH 1015 Lausanne, Switzerland, (e-mail: Daniel.Thalmann@epfl.ch).*

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

However, it is not hard to find that both these two types of NR IQMs were developed particularly for blindly predicting the quality of NS images, overlooking the 'generality' across different image contents, namely, the hotly researched screen content (SC) images. During the last several years, a growing number of attentions have been shifted to the exploration of the SC images, such as compression, quality evaluation, and segmentation. As compared to the NS images that have been long studied for decades, limited endeavors have been devoted to the SC images. There include obvious differences between the NS and SC images. By contrast, the former ones usually contain rich and complicated distributions of luminance and color, while the latter ones are generally composed of fewer luminance and color variations and simpler structures. To deal with this issue, Gu *et al.* proposed a blind quality measure for SC images (BQMS) through firstly extracting features in terms of a statistical model of SC images before learning the features to predict the image quality based on a great number of training samples [26].

Transparently, *a more unified NR IQM, which is not only applied to different types and levels of distortions but also to both NS and SC images, is higher desired*. So it is reasonable to modify Eq. (1) to be

$$\hat{\boldsymbol{\theta}}^{*} = \arg\min_{\hat{\boldsymbol{\theta}}} \left\| \mathcal{Q}(\mathbf{x}) - \mathcal{P}(\mathbf{x}; \hat{\boldsymbol{\theta}}) \right\|_{\mathcal{D}} \tag{2}$$

where $\hat{\boldsymbol{\theta}} = \{\boldsymbol{\theta}_t, \boldsymbol{\theta}_l, \boldsymbol{\theta}_c\}$ with $\boldsymbol{\theta}_c$ denoting a parameter vector for distinguishing the content type of a distorted image. To specify, we figure out this problem via a two-step framework. The first step is to propose a new loose classifier for judging which type (NS or SC) of an input image belongs to. We can straightforwardly use the natural scene statistics (NSS) model [27] to infer the type of a lossless image. *However, as for a lossy image, this classification is not easy since the NSS regulation was shown to be broken when introducing unnatural distortions (e.g. blur and noise), or texts and/or graphics [16].* That is to say, in the blind IQM, we need to devise a reliable technique which can accurately distinguish corrupted NS images and SC images. This paper attempts to cope with this problem based on real-time on-line training instances and the above-mentioned NSS model. Briefly speaking, we firstly apply the image search technology to retrieve several high-quality images that are akin to the input image from cloud. Then a pre-trained NSS-based classifier is deployed to identify the types of those similar high-quality images and thus to deduce the input image's content type. Experiments validate that combining image search technique and NSS model is capable of precisely classifying distorted SC images and NS images. Note that the image retrieving relies on real-time on-line training data downloaded by networks, while, the pre-trained NSS model is built based on a large class of off-line training samples gathered via networks beforehand.

The second step of our framework is to assess the quality score under the condition that the image type is known. One direct way is to deploy two blind IQMs, one each for NS images and SC images, such as C-DIIVINE [17] and BQMS [26]. Apart from the ad-hoc problem, this solution also brings a problem that, unlike the BQMS that was proposed using a large quantity of training data, the C-DIIVINE model was not developed based on big-data samples but a small-size image database. It is lucky that a recently designed statistic model was found to simultaneously well applied to NS images and SC images. On the basis of this statistic model, we extract 13 features from the input image, regardless of distortion types, distortion levels and content types. Afterwards, in accordance to the type of the input image inferred using the first step, we incorporate the above 13 features with a pre-trained regression module to generate the quality prediction. In this research, we separately take advantage of 100,000 NS images and 100,000 SC images, which were off-line training data collected using networks, to learn the two regression modules beforehand, followed by employing the proper one in light of the inferred content type for fusing the extracted features and predicting the image quality score. It deserves to stress that the proposed UBQI is the first blind IQM which can simultaneously assess the quality of both NS and SC images.

In the remainder of this paper, we arrange the structure as follows. Section 2 reviews representative and relevant works during recent years. Section 3 describes the proposed unified NR IQM, including how to distinguish distorted NS images and distorted SC images and learn the fixed regression module using on-line and off-line training instances downloaded via networks. Section 4 demonstrates the superiority of our proposed quality model by comparison with mainstream and state-of-the-art full-reference (FR) and NR IQMs. Section 5 concludes the whole paper.

2 Related Work

During the past years, hundreds of NR/blind IQMs were elaborately developed. In this section, we first review some of popular NR quality models of NS images. Next, we introduce prevailing explorations of SC image quality evaluation from two aspects: subjective assessment and objective assessment.

For blind IQMs devoted to NS images, they can be basically categorized into distortion-specific models and general-purpose models. The distortion-specific models are heavily suitable for one certain type of distortion. For example, in [10], Li *et al.* proposed a blind image blur evaluation index to characterize blur with discrete moments by considering that noticeable blur affects the moment magnitudes of images.

By contrast, the general-purpose models which can tackle different distortion types attract much attention due to their wider application scopes compared with the distortion-specific models. In [15], Mittal *et al.* designed the blind/referenceless image spatial quality evaluator (BRISQUE) based on the observation that there exist some statistical properties in natural images and they are influenced by corruptions. In [18], Gu *et al.* devised the NR free energy based robust metric (NFERM) that primarily relies on a structural computational model of HVS that incorporates the free-energy based brain principle and low-level perceptual characteristics of HVS. The aforementioned models are opinion-aware (OA), requiring human opinion scores. This inevitably leads to the performance instability of blind IQMs models since subjective experiments are always high-expense and time-consuming and therefore a

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

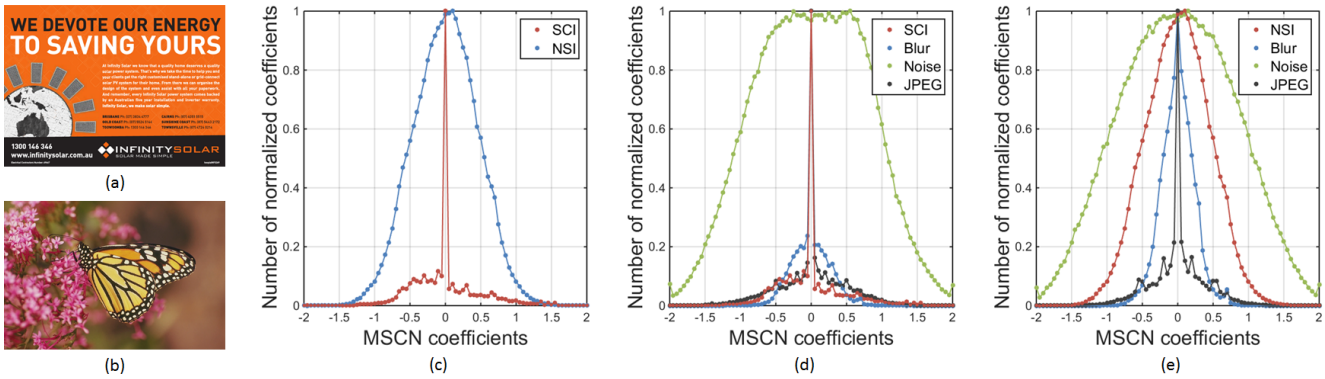


Fig. 1: Illustration of image statistics: (a) SCI (SC image); (b) NSI (NS image); (c) the histogram of MSCN coefficients of (a)-(b); (d)-(e) histograms of MSCN coefficients of images (a)-(b) under the distortions of blur, noise and JPEG compression.

large number of human-labeled training samples are difficult to collect.

To that end, another type of opinion-unaware (OU) blind IQMs independent of real human ratings are greatly popular in recent years. A pair of representative OU models are the natural image quality evaluator (NIQE) [16] and its modified version of the integrated local NIQE (IL-NIQE) [19]. The NIQE provides a NSS-based regularity which was established upon hundreds of unlabeled natural images, and then it takes the distance between a corrupted image and the above regularity for the quality prediction. Along the same research line, the IL-NIQE introduces three more statistical features from the viewpoints of image gradient, Log-Gabor filter responses, as well as colorfulness.

Notwithstanding the great successfulness achieved by the NR IQMs illustrated above, they are merely fit for assessing the quality of NS images, yet not good at the quality assessment of SC images [26]. That is, there does not exist a NR quality model which is simultaneously good at NS images and SC images. To fill in this research gap, recent years have witnessed a few ground-breaking researches. On the one hand, in subjective assessment, Yang *et al.* established the first screen image quality assessment database (SIQAD) [28], which is composed of 20 lossless SC images and their associated 980 corrupted images under seven typical types of degradations, i.e., JPEG2000 compression, JPEG compression, layer segmentation-backed coding, motion blur, Gaussian blur, Gaussian noise, and contrast change. On the other hand, the objective BQMS [26] was developed based on the screen content statistics (SCS) model. It was found that the BQMS model established based on both real human ratings and FR IQM based labels has achieved better performance than the existing NR IQMs.

3 Proposed Model

The majority of present NR IQMs have the same limitation, merely proper for quality prediction of NS or SC images. To resolve this problem, we in this work contrive a unified blind quality index (UBQI) for assessing the image quality without any prior knowledge regarding distortion categories and image contents (namely tackling both NS and SC images). Actually, many studies have pointed out that existing models dedicated to the NS images are improper for SC images [29]-[31]. On one hand, from the intuitive perspective, obvious differences exist between NS images and SC images. Fewer and simpler luminance and color changes are mostly included in the SC images, yet rich and complex luminance and color variations are often encompassed in the NS images. On the other hand, from the viewpoint of research experience, detailed structures in SC images receive more attention whereas basic structures in NS images arouse more considerations. According to these observations, we propose the UBQI by firstly using a loose image classifier[1] to judge the input image's type and then selecting the appropriate model for quality evaluation, and we thus rewrite Eq. (2) as

$$\hat{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta}_c} \arg\min_{\boldsymbol{\theta}_t,\boldsymbol{\theta}_l} \left\| \mathcal{Q}(\mathbf{x}) - \mathcal{P}(\mathbf{x};\hat{\boldsymbol{\theta}}) \right\|_{\mathcal{D}}. \quad (3)$$

3.1 Loose Image Classifier

We illustrate the proposed loose image classification under the subsequent two conditions: one refers to classifying the lossless NS images and SC images and the other refers to classifying the lossy NS images and SC images. Obviously, the latter loose image classifier, whose basis is the former classifier, is our terminal target.

3.1.1 *Loose classifier of lossless images*

The NSS regulation reveals that there exist certain statistical characteristics of the lossless NS images. But the characteristics will be destroyed when texts and/or graphics are added to the images; in other words, the characteristics do not take effects in the lossless SC images [16], [27]. For illustration consider the example in Fig. 1. Figs. 1(a)-(b) provide two representative lossless images, separately associated to SC and NS images chosen from the SIQAD and LIVE databases [28], [32]. Fig. 1(c) illustrates two histograms of mean subtracted contrast normalized (MSCN) coefficients of images shown in Figs. 1(a) and (b). The MSCN coefficients are generated by imposing local mean removal and divisive normalization on the input image $\mathbf{x}$:

$$\hat{x}(i,j) = \frac{x(i,j) - \mu_{\mathbf{x}}(i,j)}{\sigma_{\mathbf{x}}(i,j) + \epsilon} \quad (4)$$

1. Loose image classifier means a kind of classifier which can discriminate NS images and SC images.

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

TABLE 1: Primary principle of the used image search technology.

Step	Description
1	Rescaling the input images to the resolution of 8×8, 64 pixels in total.
2	Converting the rescaled colorful images into grayscale images.
3	Calculating the mean value of all the 64 image pixels.
4	Thresholding each of 64 pixels as 1 if no less than the mean value; otherwise 0.
5	Computing the Hash code of the above 64-bit integer.

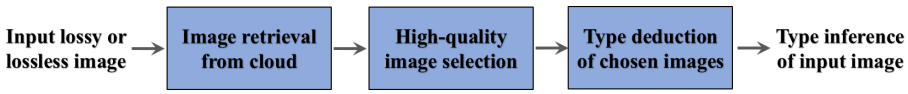


Fig. 2: A basic flow diagram of the proposed loose image classifier.

where i and j means the horizontal and vertical indices; ϵ is a tiny fixed number to eliminate the division-by-zero; $\mu_{\mathbf{x}}(i,j)$ and $\sigma_{\mathbf{x}}(i,j)$ are local mean and variance maps defined by

$$\mu_{\mathbf{x}}(i,j) = \sum_{s}\sum_{t} w(s,t)x(i+s,j+t) \tag{5}$$

$$\sigma_{\mathbf{x}}(i,j) = \left(\sum_{s}\sum_{t} w(s,t)[x(i+s,j+t)-\mu_{\mathbf{x}}(i,j)]^2\right)^{\frac{1}{2}} \tag{6}$$

where $\mathbf{w} = \{w(s,t)|s = -3,-2,...,3; t = -3,-2,...,3\}$ is one 2D circularly-symmetric Gaussian weighting function. From Fig. 2(c), we can see that the lossless SC image exhibits a Laplacian-like MSCN distribution, whereas, the lossless NS image produces a Gaussian-like MSCN distribution. Based on this finding, we propose an efficient loose classifier with the following steps: 1) computing the MSCN coefficients of an input image; 2) sampling the histogram of MSCN coefficients to form a n-dimensional vector from -2 to 2 with an interval of $\frac{4}{n-1}$; 3) learning the n-dimensional vector to judge the input image's type. The reason that we use the range of $[-2,2]$ is because, in most images, the MSCN values out of this range are usually too sparse to be ignored [15]. Here we assign n to be 81 and exploit the support vector machine (SVM) [33] for learning. Comparison with other machine learners such as random forests [34] will be compared in the next section.

3.1.2 Loose classifier of lossy images

Nevertheless, in many application scopes, images suffer from different categories of distortions, such as compression, blur, noise, and transmission error. Distortions and texts and/or graphics all reshape the histogram of MSCN coefficients, and thereby the NSS-based loose classifier provided above does not work anymore. From Fig. 2, we find that, the JPEG compressed NS image has a Laplacian-like MSCN distribution as given in Fig. 2(e), which is quite similar to that of the lossless SC image as shown in Fig. 2(d). Therefore, the loose classifier between lossy SC images and NS images poses a big challenge. A simple and straightforward solution is to restore the input lossy images with, e.g., denoising, deblurring, etc. But one main target of loose classifier is to find more suitable models or parameters for processing the input image and thus getting better-quality output images. That is, this solution seems meaningless, or even put upside down. Note that the loose classifier of lossless images mentioned above only uses the statistical information, rather than the pixel-level image information. Based on this consideration, a new solution will be proposed in this paper to restore the statistical information of a given lossy image.

Our new solution implements relying on the image search technique and cloud data. More concretely, as for an image (no matter whether it is lossless or not), a large quantity of images which have the similar content to the given image are retrieved from cloud using an image search algorithm. This is akin to recent cloud-based image processing technologies to some extent [35]-[36]. Images of poor quality are inevitably retrieved from cloud. To this end, we then pick high-quality images out from the images retrieved. It can be observed that those high-quality retrieved images, in most cases, have the same content type (NS or SC) with the input image; that is, they have the similar statistical information to the input image. Presently, we have found some replacements, which can be roughly regarded to be lossless and have the same content with the input image, and thereby we can apply the loose classifier for lossless images proposed above to those replacements to faithfully infer the content type of the input image.

According to the primary idea mentioned above, this work introduces a new network-based loose image classifier which implements with the following steps. First, about 100 images similar to the input image are retrieved from cloud using an image search technology [37], whose elementary principle is composed of five steps as shown in Table 1. Superior image search technologies can be introduced in the future work to improve the implementation efficacy and efficiency. More comparisons concerning the performance of image search technologies can be directed to [38]. In this work we use some specific visual features, consisting of shapes, colors, geometric configuration, textures, etc, to assist the computers to understand what the image looks like. Second, we rank the retrieved 100 images according to their quality and pick the nine best-quality images out, because we are unable to guarantee the images of high quality. To the best of our knowledge, there is no general blind IQM that works effectively for quality evaluation of both NS images and SC images. Therefore, this

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

TABLE 2: Main steps for type deduction of chosen images.

Step	Description
1	Building a pre-trained classifier based on a large number of lossless NS images and SC images.
2	Extracting the 81-dimensional vector of MSCN parameters from each of the nine chosen images.
3	Deducing the content type of chosen images by inputting the 81-dimensional vectors into the pre-trained classifier.

paper simply compares the bits per pixel (bpp) of each image for ranking. In reality, the proposed loose classifier of lossless images is of substantially high accuracy when an image has relatively large bbp value. Trials for validating this will be provided in the next section. Third, for each chosen image, we estimate its MSCN coefficients and sample the histogram of MSCN coefficients to constitute an 81-dimensional vector from -2 to 2 with an interval of 0.05. Fourth, we infer the content type of each chosen image by inputting the 81-dimensional vector into a pre-trained classifier, which was learned by conducting SVM on a large number of lossless NS images and SC images. We summarize the main steps of type deduction of chosen images in Table 2. Fifth, the content type of the input image is judged based on the types of the aforesaid nine chosen images. Note that, as illustrated above, '0' and '1' indicate NS images and SC images. So we can form a 9-dimensional vector according to the type results of nine retrieved images. It is clear that the input image is closer to NS images when the vector includes more zeros. On this basis, we infer the input image to be NS image when the sum of the vector is less than 4; otherwise, SC image. For the readers' conveniences, as given in Fig. 2, we further draw a basic flow diagram to help explain how to implement the proposed loose image classifier.

3.2 Blind Quality Prediction

In the second stage, in terms of the classification results of content type inferred using the above loose image classifier, we are able to straightforwardly deploy two different quality methods, involving a well-designed SC IQM and an elegant NS IQM, to evaluate the visual quality of a given image. Fig. 3 shows the framework. In practical applications, the module of 'SCI IQM' might be replaced with many image post-processing algorithms (such as denoising, deblurring, contrast enhancement, super-resolution, etc) devoted to SC images and that of 'NS IQM' is possibly replaced with image post-processing algorithms devoted to NS images, toward more unified image post-processing applications. We are able to further upgrade the above framework to image patches for identifying the type of each patch in a given image. Nonetheless, such type of combination based on two non-handpicked IQMs is not good due to the lack of consistency in quality assessment. Fortunately, recent works provide that there is one statistical model simultaneously applicable to both NS images and SC images, and based on this, we propose the unified blind UBQI model. Specifically, the UBQI model is established based on

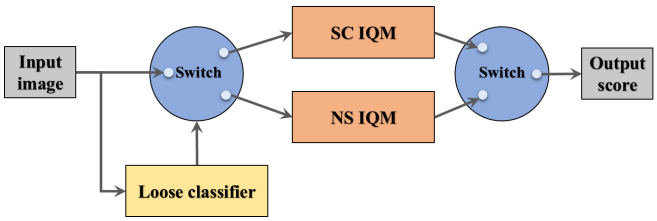


Fig. 3: The two-phase framework of the proposed UBQI model.

an image quality assessment model with two groups of fixed parameters. More details will be illustrated as follows.

From the perspective of neuroscience, the Friston's team reveals that some brain theories regarding human perception, action, and learning can be unified by the free energy theory [40]-[42]. Specifically, we can assume that the human brain performs the cognitive process by an internal generation mechanism, for instance, when perceiving visual signals. The above process can be described using a probabilistic model, which is composed of a prior term and a possibility term. The visual perception infers the posterior likelihood of a given scene by reversing this possibility term. Albeit the brain structure is more complex than the level we can recognize to date, a gap between the cognitive outcome of brain and the actual input visual signal still exists. It has been verified that this gap is highly correlated with the quality of human perceptions and thus can be used for the quality measurement [43]-[45].

Towards operational amenability, it can be supposed that the internal generation mechanism of the brain is parametric. Based on this, the internal generation mechanism can adjust the vector of model parameters $\boldsymbol{\xi}$ to well explain the external input visual signal. Given a visual signal $\mathbf{x}$, we integrate the joint distribution $\mathcal{J}(\mathbf{x}, \boldsymbol{\xi})$ on the space of model parameters to define its 'surprise':

$$-\log \mathcal{J}(\mathbf{x}) = -\log \int \mathcal{J}(\mathbf{x}, \boldsymbol{\xi})\, d\boldsymbol{\xi}. \tag{7}$$

A model of stronger expressive ability is better at simulating the brain, but it results in higher computational complexity and tends to use more parameters. This inevitably leads to higher model costs and makes it harder to estimate through observation [46]. In this work, considering the characteristics of simplicity and strong description ability [45], we choose the linear AR model as the generative model. As for an input signal $\mathbf{x}$, the model is defined as:

$$x_k = \Lambda^r(x_k)\, \mathbf{a} + b_k \tag{8}$$

where k is the index of pixel; x_k is a pixel in question; b_k is the error term; $\Lambda^r(x_k)$ is a row-vector of indices of x_k's r nearest pixels; $\mathbf{a} = (a_1, ..., a_r)^T$ is a vector of AR model parameters. We apply the least square method to find the optimal approximate of $\mathbf{a}$: $\hat{\mathbf{a}} = (X^T X)^{-1} X^T X$, where $\mathbf{x} = (x_1, ..., x_r)^T$; $X(t, :) = \Lambda^r(x_t)$. In real applications, the predicted $\hat{\mathbf{x}}$ can be finally estimated according to the method used in [43]:

$$\hat{x}_k = \Lambda^r(x_k)\, \hat{\mathbf{a}}. \tag{9}$$

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

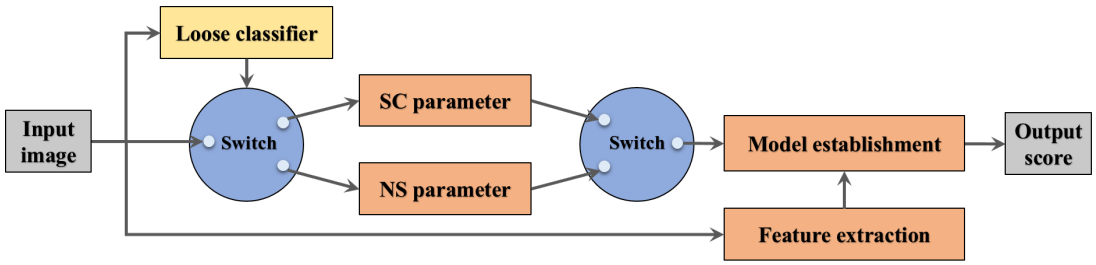


Fig. 4: The renewed framework of the proposed UBQI model.

The entropy of residual 'gap' is then used to approximate the free energy of $\mathbf{x}$:

$$\mathcal{E}_t(\mathbf{x}) = -\sum\nolimits_h \mathcal{H}_h(\bar{\mathbf{x}}) \log \mathcal{H}_h(\bar{\mathbf{x}}) \tag{10}$$

where $\bar{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ meaning the estimation difference between a given image and its estimated one; $\mathcal{H}_h(\bar{\mathbf{x}})$ is the possibility denseness of grayscale h in $\bar{\mathbf{x}}$.

From the viewpoint that the HVS is highly sensitive to the variations occurred in image structures [47], we introduce the structural degradation information defined by

$$\mathcal{S}_\alpha(\mathbf{x}) = \mathcal{G}\Big(\frac{\sigma_{(\dot{\alpha}_\mathbf{x}, \ddot{\alpha}_\mathbf{x})} + \gamma}{\sigma_{(\dot{\alpha}_\mathbf{x})}\sigma_{(\ddot{\alpha}_\mathbf{x})} + \gamma}\Big) \tag{11}$$

$$\mathcal{S}_\beta(\mathbf{x}) = \mathcal{G}\Big(\frac{\sigma_{(\dot{\beta}_\mathbf{x}, \ddot{\beta}_\mathbf{x})} + \gamma}{\sigma_{(\dot{\beta}_\mathbf{x})}\sigma_{(\ddot{\beta}_\mathbf{x})} + \gamma}\Big) \tag{12}$$

where $\mathcal{G}(\cdot)$ has the function of computing the global average; $\dot{\alpha}_\mathbf{x}$ and $\dot{\beta}_\mathbf{x}$ stand for the topical mean and variance of $\mathbf{x}$, which is calculated by utilizing a normalized Gaussian kernel $\mathbf{w}' = \{w'(u,v)|u = -U, ..., U, v = -V, ..., V\}$ with standard deviation of 1.5; $\ddot{\alpha}_\mathbf{x}$ and $\ddot{\beta}_\mathbf{x}$ have similar definitions but use an impulse function instead; $\sigma_{(\circ,\bullet)}$ denotes the partial covariance between $\circ$ and $\bullet$; γ is a small positive constant for removing the situation that the denominator value is zero. Furthermore we consider improving the structural degradation information from the subsequent three angles. First, in order to introduce diverse amount of adjacent information, we separately assign select (U, V) as $(1, 1)$, $(3, 3)$ and $(5, 5)$ in the kernels. Second, for distinguishing distinct frequency decrease between inside and outside areas caused by JPEG compression, we compute $\mathcal{S}_\alpha(\mathbf{x})$ and $\mathcal{S}_\beta(\mathbf{x})$ in the exterior block-edge part and interior 6×6 part. The last one is that $\mathcal{S}_\alpha(\mathbf{x})$ and $\mathcal{S}_\beta(\mathbf{x})$ are reversed when $\mathcal{E}_t$ is smaller than a predefined threshold, to make the structural degradation information has the consistent change with image quality. More details can be found in [26].

Early studies show that the structural degradation features and free energy feature have high linear correlation both for lossless NS images and SC images [18], [26]. This observation supplies an inspiration to build new image descriptive features for image quality evaluation. In particular, we fit the linear regression model as following:

$$\mathcal{E}_t(\mathbf{x}_0) = l_\tau \cdot \mathcal{S}^\tau_\alpha(\mathbf{x}_0) + m_\tau \tag{13}$$

$$\mathcal{E}_t(\mathbf{x}_0) = n_\tau \cdot \mathcal{S}^\tau_\beta(\mathbf{x}_0) + o_\tau \tag{14}$$

where $\mathbf{x}_0$ is a lossless image; $\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$; the parameters l_τ, m_τ, n_τ and o_τ are obtained using the least square method. We further define

$$\mathcal{T}^\tau_\alpha(\mathbf{x}) = \mathcal{S}_e(\mathbf{x}) - (l_\tau \cdot \mathcal{S}^\tau_\alpha(\mathbf{x}) + m_\tau) \tag{15}$$

$$\mathcal{T}^\tau_\beta(\mathbf{x}) = \mathcal{S}_e(\mathbf{x}) - (n_\tau \cdot \mathcal{S}^\tau_\beta(\mathbf{x}) + o_\tau). \tag{16}$$

When the quality of images are high, $\mathcal{T}^\tau_\alpha$ and $\mathcal{T}^\tau_\beta$ approach zero, while their absolute values will be far away from zero when distortions appear and grow. In such way, we yield 12 structural degradation features. Considering the fact that the free energy feature has good capability in quality prediction, the free energy feature is added to finally form 13 features.

After the extraction of features, a large number of training data are used to find the relationship between the features and the quality predictions. In order to acquire the data, we apply more than two million lossy images as training samples, one half for SC images and the other half for NS images. First, we picked up larger than a thousand high-quality SC images, which were downloaded from the website of 'Google Images' and selected by human eyes for eliminating low-quality lossy images, as lossless SC images. Likewise, we collected beyond one thousand high-quality NS images from Berkeley database [48] and PQD database [49], instead of from website, to be lossless NS images. Second, two million lossy images were generated by corrupting the above stated lossless SC and NS images with six classical distortion types as deployed in [26]. Having prepared the training samples, we label them by getting help from FR IQMs. In particular, we utilized the structure-induced quality metric (SIQM) [50] and perceptual similarity (PSIM) measure [51], due to their effectiveness in predicting the quality of SC and NS images, to respectively label the SC and NS images in training samples.

Eventually, one machine learning tool is required to build the connection between the 13 features of two million lossy images to the quality scores predicted using FR IQMs. In this paper, we apply the libSVM package to conduct the radial basis kernel backed supporting vector regression (SVR) [33]. To specify, for a training dataset $\mathcal{R} = \{(c_1, d_1), ..., (c_r, d_r)\}$, where c_i and d_i are respectively a vector of 13 features and the predicted quality score of the i-th training image. We can

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

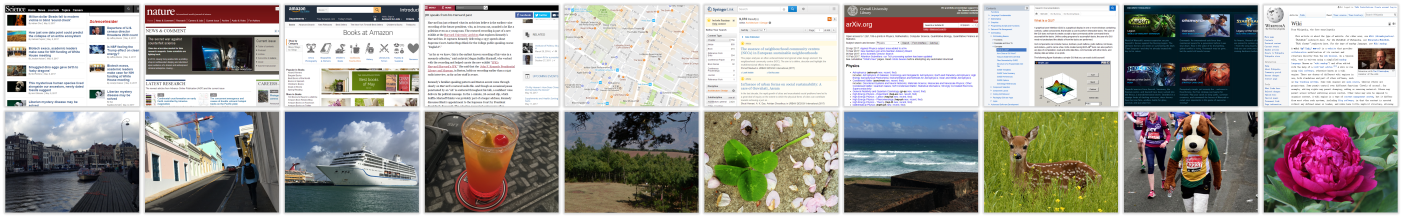


Fig. 5: Twenty lossless source screen content and natural scene images used in our new HSNID database.

express the standard SVR form by

$$\begin{aligned} \underset{\boldsymbol{\omega},\delta,\boldsymbol{\nu},\boldsymbol{\nu'}}{\arg\min} \quad & \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + \eta(\sum_{i=1}^{r}\nu_i + \sum_{i=1}^{r}\nu_i') \\ \text{subject to} \quad & \boldsymbol{\omega}^T\phi(c_i) + \delta - d_i \leq \rho + \nu_i, \\ & d_i - \boldsymbol{\omega}^T\phi(c_i) - \delta \leq \rho + \nu_i', \\ & \eta, \rho > 0,\ \nu_i, \nu_i' \geq 0,\ i = 1, ..., r. \end{aligned} \tag{17}$$

where $\mathcal{K}(c_i, c_j) = \exp(-k\,||c_i - c_j||^2)$ uses the radial basis kernel. We determine the parameters η, ρ and k based on a large quantity of training data mentioned above. In practice, several parts such as features are shared and thus we are able to simultaneously extract features and judge the content type of an input image, followed by selecting the proper parameters and regressor in light of the image content type. The renewed framework of our UBQI model is displayed in Fig. 4.

4 Results and Discussions

This section focuses on validating the performance of the proposed model from the following two aspects: classification performance of the proposed loose classifier and prediction performance of the proposed UBQI model.

4.1 Testing Databases

Before the performance comparison, we illustrate the three image databases which are considerably proper for this work. The first one is the LIVE database [32], which were constructed at the University of Texas at Austin in 2006. The database contains 29 sources and 779 distorted images corrupted by five classical distortion types. The second database we used in this paper is the SIQAD database [28], which was established by Nanyang Technological University (NTU) in the year of 2015. The database is composed of 20 reference SC images and 980 degraded images, which were produced by applying seven popular distortion types with seven intensities to the 20 reference images.

However, notice that subjective evaluations in building the LIVE and SIQAD databases are separately implemented and thus their quality scores cannot have a sufficient consistency. This inevitably makes the results unconvincing, hence a new specific database is highly desired and we therefore built the first hybrid screen content and natural scene image database (HSNID), which includes 600 distorted NS images and SC images generated by corrupting 20 sources at five distortion levels with six commonly used distortion types (i.e., contrast change, Gaussian blur, Gaussian noise, JPEG compression, JPEG2000 compression and motion blur). These 20 reference images are presented in Fig. 5, in which the top row refers to SC images while the bottom one refers to NS images. For scoring the images, the subjective test uses the single stimulation method based on a 5-point discrete scale from the worst '1' to the best '5' with the interval of '1'. We invited 50 participants to join our experiment to provide their visual opinion scores of each test image. They are undergraduate or graduate students at the university, but none of them has any experience or knowledge of quality evaluation, and thus they are pre-trained via dozens of images to accommodate them in advance. The distance between their eyes and the display monitor is approximately three times the picture height. After obtaining each test image's opinion score, the consistency is checked to ensure the usability of subjective quality grading. Also, the outlier detection is applied to the results and remove the data of three participants as outliers. Eventually, the mean opinion score (MOS) of each image is obtained by averaging all of its corresponding 47 scores.

4.2 Classification Performance

Based on the above databases, we check the classification performance of the proposed loose classifiers, respectively for lossless images and lossy images. First, we consider the lossless image's loose classification problem. If the proportion of samples involved in the classification is unbalanced, the accuracy of classifiers will be higher in the category which contains a larger size of samples, and vice versa. Therefore, in order to ensure a balanced classification ratio, we select a total of 2,000 lossless images, including 1,000 SC images and 1,000 NS images, for verification. How to collect these images has been illustrated in the context (seeing Section III-B).

In this paper, we adopt three indicators, namely precision, recall and f-measure, to evaluate the performance of classifiers. The first two indices are precision and recall:

$$\text{precision} = \frac{\mathcal{T}_p}{\mathcal{T}_p + \mathcal{F}_p} \tag{18}$$

$$\text{recall} = \frac{\mathcal{T}_p}{\mathcal{T}_p + \mathcal{F}_n} \tag{19}$$

where $\mathcal{T}_p$ (true positive) represents the number of samples that the actual situation is positive and the prediction is also positive, $\mathcal{F}_p$ (false positive) represents the number of samples that the actual situation is negative but the prediction is positive, and $\mathcal{F}_n$ (false negative) represents the number of

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

TABLE 3: Performance on the lossless images.

Classifier type	Precision	Recall	F-measure
Classifier-I	96.44%	95.08%	95.77%
MSCN+RF [34]	96.23%	91.62%	93.87%
LBP [52]	99.06%	97.19%	98.05%
HOG [53]	99.04%	98.83%	98.90%

TABLE 4: Performance accuracy on the lossy images.

F-measure	LIVE (779)	SIQAD (980)	HSNID (600)	Overall (2359)
Classifier-I	91.35%	71.22%	79.09%	80.68%
Classifier-II	97.90%	94.74%	93.10%	95.72%
MSCN+RF [34]	96.68%	70.46%	79.09%	82.84%
LBP [52]	99.68%	69.42%	80.87%	84.25%
HOG [53]	99.16%	87.55%	85.87%	92.00%

F-measure	Gain-I	Gain-II
Classifier-I	15.04%	18.64%
MSCN+RF [34]	12.88%	15.55%
LBP [52]	11.47%	13.61%
HOG [53]	3.720%	4.043%

samples that the actual situation is positive but the prediction is negative. The third one, f-measure, is expressed by

$$\text{f-measure} = \frac{(1+\kappa) * \text{precision} * \text{recall}}{\kappa * \text{precision} + \text{recall}} \tag{20}$$

where κ is a positive variable to adjust the importance of precision over recall. Via a harmonic mean of precision and recall, f-measure plays a good role in combining precision and recall. We assign κ as one in Eq. (20) for simplicity.

We employ the three indices to check the performance of our loose classifier. Towards clear discrimination, we denote the loose classifier of lossless images as Classifier-I and that of lossy images as Classifier-II. The first experiment utilizes the above-mentioned 2,000 lossless NS and SC images. We apply the proposed Classifier-I to the 2,000 images and record the median classification performance results across 1,000 iterations of random 80% train-20% test procedure. The results are tabulated in Table 3. As seen, the precision, recall and f-measure values are all greater than 95%. Furthermore, we also introduce the MSCN+random forests (RF) [34], local binary patterns (LBP) [52] and histogram of oriented gradient (HOG) [53] for classification comparison. The results can be found in Table 3 as well. The MSCN+RF's precision value is quite similar to the proposed Classifier-I, but in terms of other two indices (recall and f-measure), our classifier remarkably performs better. By comparison with LBP and HOG, which are specifically devoted to texture classification, the proposed simple classifier is of a little worse performance. In summary, our Classifier-I is capable of accurately deducing the content type of an input lossless image.

The second experiment is to examine the performance for lossy image's loose classification. In this experiment, the three databases mentioned in the previous section are employed. We apply the proposed Classifier-II to the totally 2,359 lossy images (779 from LIVE, 980 from SIQAD, and 600 from HSNID) and list the f-measure values in Table 4. The index of the overall images is obtained akin to the first experiment. We find that the proposed Classifier-II has lead to high accuracy, with its f-measure values beyond 97% for LIVE, 93% for SIQAD and HSNID, and 95% for the overall 2359 images. For highlighting the necessity of introducing image search techniques and cloud data, we further testify the performance of the proposed Classifier-I, RF, LBP and HOG. The associated results can be also found in Table 4. By contrast, we can deduce that the Classifier-II performs much better than all the four competing classifiers on overall average. We also calculate two performance gains on overall average, as listed in Table 4, as defined as follows:

$$\text{Gain-I} = \mathcal{V}_2 - \mathcal{V}_1 \tag{21}$$

$$\text{Gain-II} = \frac{\mathcal{V}_2 - \mathcal{V}_1}{\mathcal{V}_1} \times 100\% \tag{22}$$

where $\mathcal{V}_2$ and $\mathcal{V}_1$ are the performance values of each of four testing classifiers and our Classifier-II. We are able to yield the same results, namely the Classifier-II noticeably outperforms the whole four classifiers in terms of both Gain-I and Gain-II. Based on the two experiments above, introducing image retrieve methods and cloud data can lead to a substantially positive effect on loose image classification.

Furthermore, we also testify the classification performance of Classifier-I on those images whose MOS values exceed 3 in the HSNID database. In other words, we only consider the 'excellent', 'good' and 'fair'-quality images, overlooking the 'poor' and 'bad'-quality images. The precision of Classifier-I on those images is up to 96.77%. This implies that, as for the images of comparatively high quality (i.e. high-bbp images), Classifier-I can accurately identify their content types.

4.3 Quality Prediction Performance

Depending on the superior loose classifier, we propose the UBQI model. Afterwards, we further examine its prediction correlation performance in accordance of three typically used indices, which include Kendall's rank-order correlation coefficient (KROCC), Spearman's rank ordered correlation coefficient (SROCC), and Pearson's linear correlation coefficient (PLCC). The above three indices are computed to evaluate the correlation between the objective quality predictions and subjective opinion values. The former two indices estimate the monotonicity of prediction, while the last one estimates the accuracy of prediction. Note that, when computing PLCC, the nonlinearity existed in the objective quality estimations should be removed based on officially provided logistic function in advance. Here we utilize the commonly used five-parameter function expressed by

$$f(q) = \zeta_1 \left(0.5 - \frac{1}{1 + e^{\zeta_2 (q - \zeta_3)}}\right) + \zeta_4 q + \zeta_5 \tag{23}$$

where q and $f(q)$ are the objective quality scores and its associated mapped scores; ζ_i $(i \in 1, 2, 3, 4, 5)$ are five free-parameters to be fitted based on the Gauss-Newton method.

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

TABLE 5: Comparison among 13 IQMs on LIVE, SIQAD and HSNID databases. We bold the best model in each type.

LIVE (779)	Type	KROCC	SROCC	PLCC
SSIM [47]	FR	0.7963	0.9479	0.9449
VSNR [54]	FR	0.7624	0.9279	0.9236
VSI [55]	FR	0.8058	0.9524	0.9482
PSIM [51]	FR	**0.8294**	**0.9622**	**0.9584**
RRED [56]	RR	**0.7888**	**0.9429**	**0.9385**
OSVP [57]	RR	0.6275	0.8218	0.8201
RWQMS [58]	RR	0.4864	0.6693	0.6948
RQMSC [59]	RR	0.5825	0.7786	0.7815
IL-NIQE [19]	NR	**0.7115**	**0.8970**	**0.9020**
BQMS [26]	NR	0.5548	0.7496	0.7478
ASIQE [60]	NR	0.5439	0.7474	0.7602
LPSI [20]	NR	0.6175	0.8181	0.8280
UBQI (Pro.)	NR	**0.7103**	**0.8866**	**0.8951**

SIQAD (980)	Type	KROCC	SROCC	PLCC
SSIM [47]	FR	0.4235	0.5836	0.5912
VSNR [54]	FR	0.4374	0.5693	0.5966
VSI [55]	FR	0.3874	0.5381	0.5568
PSIM [51]	FR	**0.5393**	**0.7056**	**0.7144**
RRED [56]	RR	0.3984	0.5358	0.5557
OSVP [57]	RR	0.4171	0.5870	0.6341
RWQMS [58]	RR	**0.5835**	**0.7815**	**0.8103**
RQMSC [59]	RR	0.5756	0.7655	0.8014
IL-NIQE [19]	NR	0.1357	0.2021	0.1958
BQMS [26]	NR	0.5326	0.7251	0.7575
ASIQE [60]	NR	**0.5609**	**0.7570**	**0.7884**
LPSI [20]	NR	0.2254	0.3206	0.3516
UBQI (Pro.)	NR	**0.5085**	**0.7001**	**0.7322**

HSNID (600)	Type	KROCC	SROCC	PLCC
SSIM [47]	FR	0.4590	0.6425	0.6409
VSNR [54]	FR	0.3611	0.4990	0.5088
VSI [55]	FR	0.5698	0.7654	0.7788
PSIM [51]	FR	**0.6278**	**0.8263**	**0.8261**
RRED [56]	RR	**0.5328**	**0.7413**	**0.7279**
OSVP [57]	RR	0.4265	0.6167	0.6064
RWQMS [58]	RR	0.5091	0.7107	0.7154
RQMSC [59]	RR	0.3788	0.5446	0.5339
IL-NIQE [19]	NR	0.4098	0.5788	0.7065
BQMS [26]	NR	0.3553	0.5335	0.5330
ASIQE [60]	NR	0.3906	0.5734	0.5912
LPSI [20]	NR	**0.4252**	**0.6014**	**0.6064**
UBQI (Pro.)	NR	**0.5778**	**0.7816**	**0.7687**

Given an image quality model, a value close to 1 for KROCC, SROCC and PLCC illustrates superior performance.

The performance correlation comparison is conducted with FR, reduced-reference (RR) and NR IQMs. Apart from the proposed UBQI model, we dominantly consider the following three types of 12 state-of-the-art IQMs. The first type is FR IQMs, which include structural similarity (SSIM) [47], visual signal-to-noise ratio (VSNR) [54], visual saliency induced index (VSI) [55], and PSIM [51]. They supposed that the lossless images are completely known and were proposed to measure the distance between the lossless and lossy images.

TABLE 6: Average performance comparison among 13 IQMs. We bold the optimal quality model in each type.

Average-II (2359)	Type	KROCC	SROCC	PLCC
SSIM [47]	FR	0.5557	0.7189	0.7206
VSNR [54]	FR	0.5253	0.6698	0.6823
VSI [55]	FR	0.5719	0.7327	0.7425
PSIM [51]	FR	**0.6576**	**0.8210**	**0.8234**
RRED [56]	RR	0.5615	0.7225	0.7259
OSVP [57]	RR	0.4890	0.6721	0.6885
RWQMS [58]	RR	**0.5325**	**0.7264**	**0.7480**
RQMSC [59]	RR	0.5278	0.7136	0.7268
IL-NIQE [19]	NR	0.3956	0.5274	0.5589
BQMS [26]	NR	0.4948	0.6844	0.6972
ASIQE [60]	NR	**0.5120**	**0.7071**	**0.7289**
LPSI [20]	NR	0.4057	0.5563	0.5738
UBQI (Pro.)	NR	**0.5928**	**0.7824**	**0.7953**

The second type is RR IQMs, which are composed of reduced-reference entropic differencing (RRED) [56], orientation selectivity based visual pattern (OSVP) [57], reduced-reference wavelet-domain quality measure of SC pictures (RWQMS) [58], and reduced-reference quality measure of screen content pictures (RQMSC) [59]. They supposed that the partial information of the lossless images are accessible and were developed to calculate the difference of the lossy images from the associated lossless versions. The last type is NR IQMs, which include IL-NIQE [19], BQMS [26], accelerated screen image quality evaluator (ASIQE) [60], and local pattern statistics index (LPSI) [20]. They supposed that none of lossless images can be obtained and were devised to assess the visual quality based on statistics such as intrinsic characteristics of undistorted images.

We illustrate the results of performance measure and comparison in Tables 5-7. As seen, four conclusions are derived.

1) The proposed UBQI model has attained very encouraging performance. In terms of SROCC, its performance approximates 0.9 on the LIVE database, surpasses 0.7 on the SIQAD database, and is close to 0.8 on the new HSNID database.
2) By comparison with relevant IQMs, our UBQI method also exhibits high performance. For conveniences, we highlight the optimal image quality model based on the SROCC index in each type. It can be found that, on the LIVE and SIQAD databases, the performance of the proposed model is near to the optimal NR IQM. On the novel HSNID database, our UBQI model is remarkably superior to the best performing FR and RR IQMs, and even close to the state-of-the-art FR PSIM method.
3) A comprehensive comparison is carried out as well to validate the superiority of our proposed UBQI method. Here we introduce the data size-weighted average, defined as follows:

$$\text{Average-II} = \frac{\sum_{i=1}^{3} \varrho_i \cdot \pi_i}{\sum_{i=1}^{3} \pi_i} \tag{24}$$

where ϱ_1, ϱ_2, ϱ_3 are the performance indices for each of

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

TABLE 7: Performance comparison of the UBQI model with different classifiers.

LIVE (779)	KROCC	SROCC	PLCC
UBQI-NS	0.7164	0.8916	0.9007
UBQI-SC	0.5548	0.7496	0.7478
UBQI-Classifier-I	0.6907	0.8722	0.8729
UBQI-Classifier-II	0.7103	0.8866	0.8951
UBQI-RF	0.7113	0.8898	0.8959
UBQI-LBP	0.7154	0.8909	0.8998
UBQI-HOG	0.7095	0.8855	0.8915

SIQAD (980)	KROCC	SROCC	PLCC
UBQI-NS	0.0944	0.1368	0.1532
UBQI-SC	0.5326	0.7251	0.7575
UBQI-Classifier-I	0.3418	0.4930	0.5390
UBQI-Classifier-II	0.5085	0.7001	0.7322
UBQI-RF	0.3633	0.5212	0.5706
UBQI-LBP	0.3340	0.4766	0.5486
UBQI-HOG	0.4556	0.6372	0.6758

HSNID (600)	KROCC	SROCC	PLCC
UBQI-NS	0.3862	0.5484	0.6199
UBQI-SC	0.3553	0.5335	0.5330
UBQI-Classifier-I	0.5477	0.7521	0.7429
UBQI-Classifier-II	0.5778	0.7816	0.7687
UBQI-RF	0.5477	0.7521	0.6577
UBQI-LBP	0.5535	0.7506	0.7456
UBQI-HOG	0.5607	0.7639	0.7545

Average-II (2359)	KROCC	SROCC	PLCC
UBQI-NS	0.3740	0.4908	0.5187
UBQI-SC	0.4948	0.6844	0.6972
UBQI-Classifier-I	0.5094	0.6841	0.7011
UBQI-Classifier-II	0.5928	0.7824	0.7953
UBQI-RF	0.5251	0.7017	0.7002
UBQI-LBP	0.5158	0.6831	0.7147
UBQI-HOG	0.5662	0.7514	0.7671

three databases; π_1, π_2, π_3 are assigned as the number of images in each database, i.e. 779 for LIVE, 980 for SIQAD and 600 for HSNID. One can readily observe from Table 6 that, in accordance to this average index, the proposed blind UBQI model has achieved the much higher performance than the same type of state-of-the-art NR IQMs and prevailing RR IQMs. Particularly, even compared with the recently devised FR PSIM model, our quality model shows the comparable performance.

4) Notice that the proposed UBQI model depends on the Classifier-II, which can well work under the conditions of lossless and lossy image content inference. It is natural to compare the performance of inserting difference classifiers into the UBQI model. As illustrated in Table 7, we report the correlation performance among four IQMs: UBQI-NS supposes that all the images are natural scene type; UBQI-SC supposes that all the images are screen content type; UBQI-Classifier-I applies the loose classifier of lossless images; UBQI-Classifier-II applies the loose classifier of lossy images; UBQI-RF applies the RF-based classifier; UBQI-LBP applies the LBP-based classifier; UBQI-HOG applies the HOG-based classifier. We can find that the UBQI-Classifier-II, namely the proposed UBQI model, has lead to considerably good performance on the three testing databases, particularly exceeding the others on the new HSNID database and average performance. In addition, it can be viewed that the loose classifier of lossless images also leads to a large performance boost as compared to UBQI-NS and UBQI-SC, even though it does not take the influence of distortions on classification of content type into account. That is to say, an effective classifier is highly required for building a unified blind image quality metric.

Further, the scatter plots of MOS versus eight testing IQMs on the HSNID database are displayed in Fig. 6 for a visual comparison. RR IQMs (including OSVP, RWQMS, RQMSC), and NR IQMs (including IL-NIQE, BQMS, ASIQE, LPSI) are used for the comparison with the proposed UBQI. We use various colors to label the sampling points associated with different image types in each scatter plot: Red sampling points refer to natural scene images, and blue sampling points refer to screen content images. As shown in Fig. 6, the scatter plot illustrate that, as compared to other metrics, both the red and blue sampling points of UBQI are more densely closed to each other. That is to say, the proposed UBQI metric has a greater performance in the evaluation of both NS images and SC images. Moreover, it is more stable than other algorithms involved in the comparison.

Along the research line of this work, the concept of big data can promote the study of quality prediction from the following two aspects. First, the high-performance full-reference quality methods can be used via weak supervision learning to label a huge number of distorted image samples for training more robust NR IQMs [6]. Second, we can establish some statistics models based on a large quantity of instances as the 'standard', followed by measuring the deviation of a new image with the 'standard' to be the quality index [61].

5 Conclusions

In this paper, we have specifically developed a unified NR IQM, not only applied to various distortion types and intensities but also to different image contents including classical NS images and prevailing SC images. Previous image quality models are only devoted to the study of one of the natural scene images or screen content images. Our proposed NR UBQI model implements by: 1) extracting features based on the scene statistical model, free energy brain principle and HVS characteristics; 2) inferring the content type of the input image; 3) predicting the visual quality of the image. In the last two steps, we introduce two data-driven learning processes based on on-line and off-line training instances. Via sufficient experiments, it is proved that the proposed blind IQM can simultaneously infer the quality of different types of images, and has acquired high performance and good stability in the quality assessment of both NS images and SC images.

References

[1] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image

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/TBDATA.2019.2895605, IEEE Transactions on Big Data

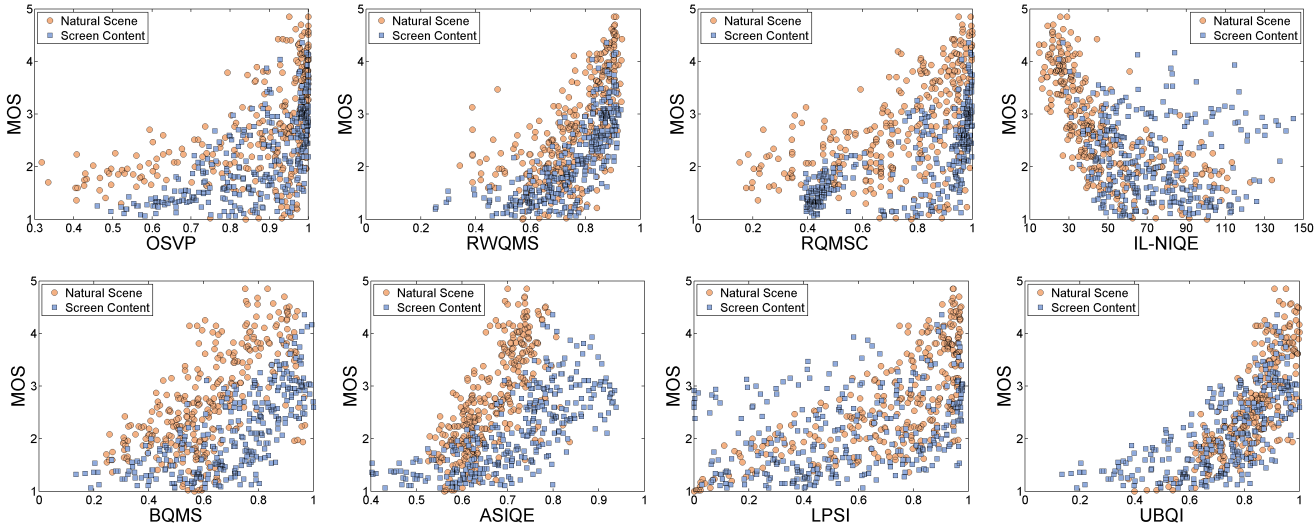


Fig. 6: Scatter plots of MOS versus RR IQMs (including OSVP, RWQMS, RQMSC), NR IQMs (including IL-NIQE, BQMS, ASIQE, LPSI) and our UBQI model on the HSNID database. Red and blue sample points refer to natural scene images and screen content images.

retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088-1099, Jul. 2006.

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

[4] S. Wang, X. Zhang, X. Liu, J. Zhang, S. Ma, and W. Gao, "Utility-driven adaptive preprocessing for screen content video compression," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 660-667, Mar. 2017.

[5] 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.*, vol. 38, no. 1, pp. 47-58, Jan. 2018.

[6] 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.

[7] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529-539, Apr. 2010.

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

[9] 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.

[10] 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.

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

[12] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure," *IEEE Trans. Ind. Electron.*, 2018.

[13] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914-926, Apr. 2018.

[14] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for DIBR-synthesized image quality evaluation," *IEEE Trans. Image Process.*, 2018.

[15] 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.

[16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Sig. Process. Lett.*, vol. 22, no. 3, pp. 209-212, Mar. 2013.

[17] 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," *Sig. Process.: Image Commun.*, vol. 29, no. 7, pp. 725-747, Aug. 2014.

[18] 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.

[19] 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.

[20] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," *in Proc. IEEE Int. Conf. Image Process.*, pp. 339-343, Sep. 2015.

[21] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 26, no. 6, pp. 1275-1286, Jun. 2015.

[22] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 26, no. 10, pp. 2275-2290, Oct. 2015.

[23] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444-4457, Sep. 2016.

[24] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multi-stage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2035-2048, Aug. 2018.

[25] G. Yue, C. Hou, and K. Gu, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2722-2732, Oct. 2018.

[26] 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.

[27] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517-548, 1994.

[28] 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.

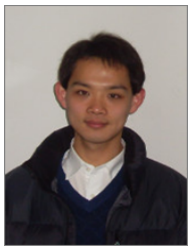
[29] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838-3851, Aug. 2016.

[30] Z. Pan, H. Shen, 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.

[31] S. Wang, K. Gu, S. Ma, and W. Gao, "Joint chroma downsampling and upsampling for screen content image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1595-1609, Sep. 2016.

[32] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment Database Release 2," 2006, Online at: http://live.ece.utexas.edu/research/quality

[33] 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.
[34] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
[35] H. Yue, X. Sun, J. Yang, and F. Wu, "Landmark image super-resolution by retrieving web images," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4865-4878, Dec. 2013.
[36] H. Yue, X. Sun, J. Yang, and F. Wu, "Image denoising by exploring external and internal correlations," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1967-1982, Jun. 2015.
[37] G. S. Pass and F. Wood, "Image searching techniques," *U.S. Patent* No. 6556710, Apr. 2003.
[38] X. Tian, Y. Lu, N. Stender, L. Yang, and D. Tao, "Exploration of image search results quality assessment," *IEEE Trans. Big Data*, vol. 1, no. 3, pp. 95-108, Sep. 2015.
[39] H. Neven Sr and H. Neven, "Image-based search engine for mobile phones with camera," *U.S. Patent*, No. 7565139, Jul. 2009.
[40] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology Paris*, vol. 100, pp. 70-87, Jul. 2006.
[41] K. Friston, "The free-energy principle: A rough guide to the brain?" *Trends in cognitive sciences*, vol. 13, no. 7, pp. 293-301, Jul. 2009.
[42] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, pp. 127-138, Feb. 2010.
[43] 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.
[44] L. Xu, W. Lin, L. Ma, Y. Zhang, Y. Fang, K. N. Ngan, S. Li, and Y. Yan, "Free-energy principle inspired video quality metric and its use in video coding," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 590-602, Apr. 2016.
[45] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379-391, Feb. 2018.
[46] J. Rissanen, Jr. and G. G. Langdon, "Universal modeling and coding," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 12C23, Jan. 1981.
[47] 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.
[48] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *in Proc. IEEE Int. Conf. Comput. Vis.*, pp. 416-423, 2001.
[49] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930-1943, Dec. 2013.
[50] 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 and Syst.*, pp. 125-128, May 2015.
[51] 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.
[52] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, Jul. 2002.
[53] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *in Proc. IEEE Conf. Comput. Vision & Patt. Recogn.*, vol. 1, pp. 886-893, Jun. 2005.
[54] 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, Sept. 2007.
[55] 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.
[56] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517-526, Feb. 2012.
[57] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Science*, vol. 351, pp. 18-29, Jul. 2016.
[58] 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.
[59] 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.
[60] 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.
[61] K. Gu, J. Qiao, and X. Li, "Highly efficient picture-based prediction of PM2.5 concentration," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3176-3184, Apr. 2019.

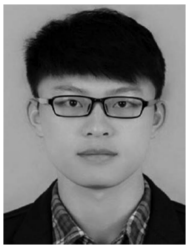


Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE TRANSACTIONS ON MULTIMEDIA, the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo 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 of the IEEE ACCESS, Signal Processing Image Communication (SPIC), and IET Image Processing. He is a Reviewer for 20 top SCI journals.

Xin Xu is currently working toward the B.S. degree at Beijing University of Technology, Beijing, China. Her interests include image processing and machine learning.

Junfei Qiao received the B.E. and M.E. degrees in control engineer from Liaoning Technical University, Fuxin, China, in 1992 and 1995, respectively, and the Ph.D. degree from Northeast University, Shenyang, China, in 1998. He was a Post-Doctoral Fellow with the School of Automatics, Tianjin University, Tianjin, China, from 1998 to 2000. He joined the Beijing University of Technology, Beijing, China, where he is currently a Professor. He is the Director of the Intelligence Systems Laboratory. His current research interests include neural networks, intelligent systems, self-adaptive/learning systems, and process control systems. Prof. Qiao is a member of the IEEE Computational Intelligence Society. He is a Reviewer for more than 20 international journals, such as the IEEE Transactions on Fuzzy Systems and the IEEE Transactions on Neural Networks and Learning Systems.

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/TBDATA.2019.2895605, IEEE Transactions on Big Data



Qiuping Jiang is currently working toward the Ph.D. degree in signal and information processing at Ningbo University, Ningbo, China. Since January 2017, he has been a joint Ph.D. student in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has authored and co-authored more than ten academic papers in refereed journals and conferences in the areas of three-dimensional image/video processing, visual quality assessment, and saliency detection.

Weisi Lin (F'16) received the Ph.D. degree from Kings College London. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with more than 340 refereed papers published in international journals and conferences. He has been on the Editorial Board of the IEEE T-IP, T-MM (2011-2013), SPL, and JVCI. He has been elected as an APSIPA Distinguished Lecturer (2012/13). He served as a Technical-Program Chair for Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo 2013, and the International Workshop on Quality of Multimedia Experience 2014. He is a fellow of Institution of Engineering Technology, an Honorary Fellow of the Singapore Institute of Engineering Technologists, and a Fellow of IEEE.

Daniel Thalmann is a Swiss and Canadian Computer Scientist. He is one of the most highly cited scientists in Computer Graphics. He is currently Honorary Professor at EPFL, Switzerland, and Director of Research Development at MIRALab Sarl. Pioneer in research on Virtual Humans, his current research interests also include social robots, crowd simulation and Virtual Reality. Daniel Thalmann has been the Founder of The Virtual Reality Lab (VRlab) at EPFL, Switzerland, Professor at The University of Montreal and Visiting Professor/Researcher at CERN, University of Nebraska, University of Tokyo, and National University of Singapore. From 2009 to 2017, he was Visiting Professor at the Institute for Media Innovation, Nanyang Technological University, Singapore. He is coeditor-in-chief of the Journal of Computer Animation and Virtual Worlds, and member of the editorial board of 12 other journals. Daniel Thalmann was Program Chair and CoChair of several conferences including IEEE-VR, ACM-VRST, and ACM-VRCAI. Daniel Thalmann has published more than 600 papers in Graphics, Animation, and Virtual Reality. He is coeditor of 30 books, and coauthor of several books including 'Crowd Simulation' (second edition 2012) and 'Stepping Into Virtual Reality' (2007), published by Springer. He received his PhD in Computer Science in 1977 from the University of Geneva and an Honorary Doctorate from University Paul-Sabatier in Toulouse, France, in 2003. He also received the Eurographics Distinguished Career Award in 2010, the 2012 Canadian Human Computer Communications Society Achievement Award, and the CGI 2015 Career Achievement. More details on http://en.wikipedia.org/wiki/Daniel_Thalmann