Local and Global Feature Learning for Blind Quality Evaluation of Screen Content and Natural Scene Images

Wujie Zhou, Lu Yu, *Member, IEEE*, Yang Zhou, Weiwei Qiu, Ming-Wei Wu, *Member, IEEE*, and Ting Luo

***Abstract*— The blind quality evaluation of screen content images (SCIs) and natural scene images (NSIs) has become an important, yet very challenging issue. In this paper, we present an effective blind quality evaluation technique for SCIs and NSIs based on a dictionary of learned local and global quality features. First, a local dictionary is constructed using local normalized image patches and conventional *K*-means clustering. With this local dictionary, the learned local quality features can be obtained using a locality-constrained linear coding with max pooling. To extract the learned global quality features, the histogram representations of binary patterns are concatenated to form a global dictionary. The collaborative representation algorithm is used to efficiently code the learned global quality features of the distorted images using this dictionary. Finally, kernel-based support vector regression is used to integrate these features into an overall quality score. Extensive experiments involving the proposed evaluation technique demonstrate that in comparison with most relevant metrics, the proposed blind metric yields significantly higher consistency in line with subjective fidelity ratings.**

***Index Terms*— Screen content images (SCIs), natural scene images (NSIs), image quality assessment, local and global features, locality-constrained linear coding, collaborative representation.**

I. Introduction

With the broad dissemination of the Internet in recent years, massive content composed of screen content images (SCIs) and natural scene images (NSIs), has become more frequently and closely employed in daily life. In addition, the use of SCIs and NSIs in various multimedia applications for computers and other electronic devices is increasing, e.g., for visual screen sharing, remote education, remote computing, cloud gaming, immediate communication, and snapshots [1]–[3]. During acquisition, compression, storage, and transmission in multiclient communication systems, SCIs and NSIs are inevitably degraded by various distortion artifacts, such as blurring, noising, compression artifacts, contrast change, quantization, and transmission loss [4]–[7]. Therefore, accurate techniques for evaluating the perceptual quality of SCIs and NSIs are required in order to design, monitor, and optimize the performance of each processing stage. That is, a quality evaluation model for SCIs and NSIs can be used as a performance index for further improving compression efficiency. Moreover, it can be used by the sender to control SCI and NSI quality according to user requirements. Thus, this topic is attracting considerable research interest [8]–[16].

In general, image quality assessment (IQA) methods can be divided into two types: subjective and objective methods [8]–[10]. Subjective methods are based on subjective judgments made by human beings, whereas objective methods provide an objective metric to quantify the perceptual quality of distorted images. Although subjective IQA using human subjects is the most natural and reliable method of quantifying perceptual quality, this approach has a number of limitations; it is time consuming, complex, and cumbersome. Most notably, subjective IQA cannot be implemented in real-time or automated systems [8]–[10]. Therefore, objective IQA metrics that can accurately and automatically predict image quality are highly desirable.

In recent years, various objective IQA metrics for SCIs or NSIs have been designed [11]–[16], [26], [27]. Depending on the availability of reference information, objective IQA metrics are typically classified into three categories: full-reference (FR), reduced-reference (RR), and blind/no-reference (NR) metrics. At present, the most commonly used IQA metrics for SCIs or NSIs are FR metrics, which assume that the reference information is completely known. The well-known structural similarity index (SSIM) proposed by Wang *et al*. [11] can be considered as a milestone study of FR-IQA metrics. Other relevant studies can be found in [12]–[17].

The FR-IQA metric provides a useful and effective means of predicting visual quality differences. However, the practical

Manuscript received May 26, 2017; revised September 17, 2017 and November 19, 2017; accepted January 9, 2018. Date of publication January 15, 2018; date of current version February 9, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61502429, Grant 61505176, Grant 61501270, and Grant 61431015, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY18F020012 and Grant LQ15F020010, and in part by the China Postdoctoral Science Foundation under Grant 2015M581932. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wen Gao. *(Corresponding author: Wujie Zhou.)*

W. Zhou is with the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China, and also with the Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: wujiezhou@163.com).

L. Yu is with the Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: yul@zju.edu.cn).

Y. Zhou, W. Qiu, and M.-W. Wu are with the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China (e-mail: zhouyang@zust.edu.cn; qiuweiwei@zust.edu.cn; wumingwei@zust.edu.cn).

T. Luo is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China.

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

Digital Object Identifier 10.1109/TIP.2018.2794207

application of FR-IQA metric is very limited, as reference information is rarely accessible on the client side. In such cases, if some reference information can be made available, RR techniques can then be implemented; this approach has attracted considerable interest and yielded reasonably good performance for different distortion types [8], [18]. However, the RR-IQA metric still requires reference information in practice; therefore, this approach is incompatible with most relevant image processing systems, which do not provide additional reference information. In comparison, blind metrics can predict the visual quality of distorted images with no prior knowledge of the reference information. This study focuses on blind quality metrics.

In recent years, numerous blind IQA metrics for NSIs have been studied extensively [19]–[28]. For instance, Ye and Doermann [19] proposed a visual codebook-based blind IQA metric to measure NSIs quality using histograms of codeword occurrences, but the codebook size is very large. Xue *et al*. [20] proposed a quality clustering (QAC) metric that learns a set of quality-predictive centroids. These centroids are then used as a codebook to calculate the quality of an NSI patch; hence, the final quality value of the overall NSI is inferred. Further, Mittal *et al*. [21] presented the NSI quality evaluator (NIQE), which does not require training with human-scored distorted NSIs. Inspired by NIQE, Zhang *et al*. [22] developed an integrated-local NIQE (IL_NIQE) by integrating more NSI feature statistics. Further, Xue *et al.* [23] presented a blind natural NSI IQA metric that uses joint statistics of the gradient magnitude and Laplacian of Gaussian features (GM_LOG). In addition, Zhou *et al.* [24] developed a natural IQA metric by analyzing the usefulness and effectiveness of two complementary image components, i.e., the gradient phase and magnitude. Xu *et al.* [25] presented a blind IQA metric by aggregating soft-weighted high-order statistical differences between a small codebook and normalized NSI patches. From the image content perspective, SCIs can be regarded as a mixture of natural images, computer graphics, document images, and other components. Obviously, these conventional blind metrics may not perform well on SCIs, as the statistical properties of the pictorial and textual regions in such images are distinct from those of natural images. Furthermore, the same levels of distortion in distinct regions may yield differences in perceptual quality [10]. In other words, the application of blind IQA to SCIs is significantly more complex than it is for NSIs, because it is affected by the quality of the natural images, computer graphics, document images, and other content [12]–[17], [29], [30].

Blind quality assessment of SCIs has received little research attention; thus, only a small number of studies have focused on this area [29]–[32]. For instance, Gu *et al.* [29] presented a blind quality metric (BQMS) for SCIs based on a new SCI statistical model. Further, Qian *et al.* [30] devised a blind SCI quality IQA metric utilizing an edge-preserving filter-based free energy and structural degradation model. Gu *et al.* [32] proposed a blind SCI IQA metric through bag data learning. However, the performance improvement yielded by these methods is limited by insufficient consideration of the statistical properties of the SCIs. Thus, the efficacy of blind IQA metrics for SCIs can be improved significantly.

In the current state of the research described above, the existing blind IQA metrics are either developed for NSIs or designed for SCIs. Only a very few quality metrics work for both simultaneously [25], [31]. In practical multimedia application systems, we may encounter cross-content-type images (e.g., NSIs, SCIs, and other image types). Efficient general blind IQA metrics that do not depend on image types are required in such circumstances.

To further advance the development of blind IQA metrics for SCIs and NSIs, in this study, we proposed an effective blind quality metric for SCIs and NSIs that fuses learned local and global quality features to efficiently represent both the local fine details and global statistical structures of the images. The key contributions of our work are summed up as follows:

(1) The local and global dictionaries are preconstructed based on an image statistical model. With these local dictionaries, the learned local and global quality features can comprehensively characterize the features of the images (e.g., spots, lines, and corners, which are the basic elements of SCIs and NSIs).

(2) The learned local quality features can be obtained using a strategy of locality-constrained linear coding (LLC) with max pooling. Meanwhile, the collaborative representation (CR) algorithm is used to efficiently code the learned global quality features of the distorted images using global dictionary.

(3) To the best of our knowledge, this is the first attempt to combine LLC-based quality local features and CR-based global quality features to achieve a fused representation for distorted SCIs and NSIs. The combined use of two types of features can effectively mitigate the respective shortcomings of the individual local and global features.

The remainder of this study is organized as follows. Section II describes in detail the proposed blind quality assessment metric, including the local and global feature learning process and the support vector regression (SVR)-based quality feature mapping process. The experimental results are presented and discussed in Section III. Finally, we conclude this study in Section IV.

II. Methodology

A flow chart of the proposed blind metric is depicted in Fig. 1. The process is composed of two stages: i) local and global feature learning and ii) perceived quality prediction. In the feature learning stage, the local and global dictionaries are preconstructed in advance based on an image statistical model. The local and global quality features are learned using these dictionaries and able to express the micro- and macro-structures of the distorted images. In the perceptual quality prediction stage, SVR is implemented to determine the overall quality score.

The final human perception of visual signals is a synthesis of local and global visual perception [32]–[35]. The average distortion in an image influences the overall human visual perception to some degree, whereas an extremely distorted local region lowers the overall visual quality severely. Because of

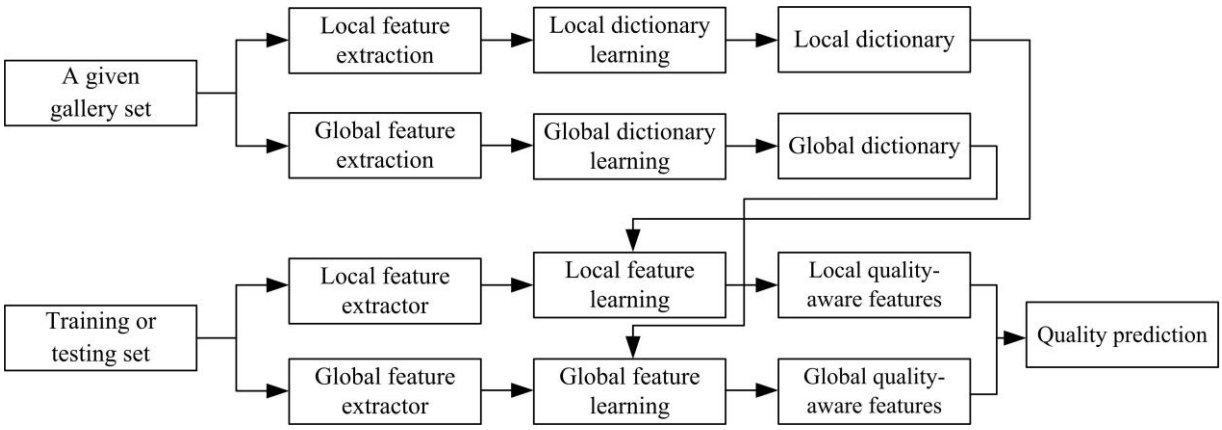


Fig. 1. A flow chart of the proposed blind metric.

the fundamental difference in the computational methods for the local and global quality features, we expect the two feature representations to provide different types of information. That is, most local quality features convey texture information for a given image patch. However, global quality features include contour representations, texture features, and shape descriptors. Global and local texture features provide different types of information about an image because the support over which the texture is calculated varies. Note that it is necessary to consider both the local and global statistical properties of the images when designing the proposed metric. Therefore, in this work, we combine metrics for patch-based local artifacts and global quality degradation to obtain the overall prediction score.

A. Local Quality Feature Learning

The local quality feature learning methods implemented in this work incorporate local feature extraction, local dictionary learning, and local feature encoding. These processes are individually described in more detail below.

1) Local Feature Extraction: For a given image, local features $X = [x_1, x_2, \ldots, x_N] \in R^d (d = B \times B)$ are extracted from a given gallery set of $B \times B$ image patches [22], [36], [37], where N is the total number of local features. To reduce redundancies in the local features and to simulate early visual nonlinear processing (the masking phenomenon) in the human visual system (HVS), each patch is normalized by subtracting the local mean value from the patch and dividing the result by the standard deviation of its elements [38], [39]. In addition, we perform a zero-phase component analysis whitening process on the normalized patches to further reduce the linear dependence relations between the local features [40].

2) Local Dictionary Learning: Local dictionary learning is based on the concept of mapping the patch-based local feature space into more meaningful informative structures. The resultant structures are usually called "visual codewords" [19], [41]. In this work, we follow the dictionary learning protocol presented in [25]. In particular, a 100-codeword local dictionary is constructed using K-means clustering [42]. The final generated local dictionary is described using vectors of the form $D_{d\times K}^{Local} = [D_1, D_2, D_3, \ldots, D_K]$, which captures various distortion characteristics. Hence, the statistical differences between the local dictionary and local features indicate the perceptual quality of the examined images.

3) Local Feature Encoding: In this approach, for each individual local feature $\widehat{x}_i$, the k nearest codewords $kNN(\widehat{x}_i)$ are selected based on their Euclidean distances. The soft-weighted mean of the patch based local features assigned to cluster c is expressed as:

$$\begin{cases} \widehat{X}_c = \sum_{i:c\in kNN(\widehat{x}_i)} [w_{ic}\widehat{x}_i], \quad c = 1, 2, 3 \ldots C, \\ w_{ic} = e^{-\beta\left\|\widehat{x}_i - D_c\right\|} \end{cases} \tag{1}$$

where w_{ic} is the Gaussian kernel similarity between feature $\widehat{x}_i$ and one of its k nearest codewords D_c. Then, all the soft-weighted means of the patch based local features are concatenated into the patch-based local feature vector $\widehat{X}_{d\times C}^{Local} = [\widehat{X}_1, \widehat{X}_2, \widehat{X}_3, \ldots, \widehat{X}_C] \in R^{d\times C}$.

Here, we do not utilize sparse coding but LLC coding instead, because sparse coding requires a greater computational effort. Furthermore, LLC coding not only ensures sparsity, but also obtains the feature representation by solving a constrained least-squares data-fitting problem [19], [43]. Thus, both efficiency and accuracy are obtained. The objective function of LLC coding is formulated as

$$\arg\min_{\alpha} \left\| \widehat{X}_i - D_{d\times C}^{Local} \cdot \alpha_i \right\|^2 + \lambda \left\| l_i \,\Theta\, \alpha_i \right\|^2 \quad s.t.\ l\alpha_i = 1, \tag{2}$$

where $\alpha = [\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_C]$ is a coding coefficient vector for $\widehat{X}_{d\times C}^{Local}$, Θ denotes element-wise multiplication, λ is a small constant that acts as a regularization parameter to adjust the weight decay speed, and l_i denotes the ith locality adaptor that assigns weights to each element according to its distance from the local feature vector. We define $l_i = f(\widehat{X}_i, D_{d\times C}^{Local})$, where $f(\cdot)$ measures the distance between the local dictionary and local feature vector [40]. Finally, the learned local quality features can be obtained from

$$\mathbf{F}^{local} = [\max(\alpha_1), \max(\alpha_2), \ldots, \max(\alpha_C)]. \tag{3}$$

B. Global Quality Feature Learning

Semantic structural alteration can generally reflect degradations in visual quality. Recent work indicated that structure descriptors (e.g., local binary patterns (LBPs) [44]) can effectively and efficiently represent the semantic structural information of visual signals and can be considered as the binary approximation of the semantic structural information primitives in the primary visual cortex [24].

The global quality feature learning method implemented in this work incorporates global dictionary learning and global feature representation. These processes are discussed in detail in the following subsections.

1) Magnitude and Phase of the Log-Gabor: The HVS is highly sensitive to the edge profile representation that is often encountered in images [17], [18]. Because cortical and retinal neurons in the primary stage of vision respond to stimulus frequency and orientation, multiscale orientation filter responses are similar to the orientation and scale sensitivities of the visual receptive fields of HVS [45]. In this work, perceptually similar log-Gabor filters (kernels) are utilized for global feature extraction [46]. The frequency and orientation responses of the log-Gabor filters with orientation j and scale n can be obtained by

$$G_{n,j}(\omega,\theta)=\exp\left(-\frac{(\log(\omega/\omega_0))^2}{2\sigma_r^2}\right)\cdot\exp\left(-\frac{(\theta-\theta_j)^2}{2\sigma_\theta^2}\right), \tag{4}$$

where J denotes the number of orientations, $\theta_j=j\pi/J$ ($j=\{0,1,\ldots,J-1\}$) denotes the orientation angle, σ_r controls the filter radial bandwidth, σ_θ determines the angular bandwidth of the filter, ω is the frequency, and parameter ω_0 denotes the center frequency.

For every log-Gabor kernel, at each point $\mathbf{x}$ at a particular scale and orientation, a complex number containing two log-Gabor components (i.e., the real component $\eta_{n,j}(\mathbf{x})$ and imaginary component $\zeta_{n,j}(\mathbf{x})$) is generated. Using these two components, the magnitude ρ at $\mathbf{x}$ at scale n for orientation j is calculated from

$$\rho(n,j,\mathbf{x})=\left(\eta_{n,j}(\mathbf{x})^2+\zeta_{n,j}(\mathbf{x})^2\right)^{1/2}. \tag{5}$$

Similarly, the phase φ at location $\mathbf{x}$ at scale n for orientation j is calculated from

$$\varphi(n,j,\mathbf{x})=\arg\tan\left(\frac{\zeta_{n,j}(\mathbf{x})}{\eta_{n,j}(\mathbf{x})}\right). \tag{6}$$

2) Binarization of $\rho(n,j,\mathbf{x})$: Given a magnitude value $\rho(n,j,\mathbf{x})$ of the central pixel z_c in a 3×3 pattern, the binarization of $\rho(n,j,\mathbf{x})$ is calculated by comparing its magnitude with its P-neighbor values. Then, the binary quantification of $\rho(n,j,\mathbf{x})$ can be expressed as

$$C^{\rho}(z_c-z_i)=\begin{cases}1, & \text{if } \rho(n,j,z_i)\geq\rho(n,j,z_c),\\ 0, & \text{else}.\end{cases} \tag{7}$$

3) Binarization of $\varphi(n,j,\mathbf{x})$: A phase change generally induces significant visual distortion. In this study, we first readjust the phase range to [0, 360). Motivated by the local binary pattern (LBP) strategy [44], we consider two cases when comparing feature types: different feature types or identical/similar feature types. To improve the stability of the binary quantification, the normalized phase range is divided into K intervals (here, we set $K=4$). The local feature types are considered similar when they belong to the same interval; otherwise, they are considered different. The binarization of $\varphi(n,j,\mathbf{x})$ can be expressed as

$$\xi^{\varphi(n,j,z_c)}(z_c-z_i)=\begin{cases}1, & \text{if } Q(\varphi(n,j,z_c))=Q(\varphi(n,j,z_i)),\\ 0, & \text{else},\end{cases} \tag{8}$$

where $Q(\theta)$ is the phase quantification operator, defined as

$$Q(\varphi)=q \quad \text{if } \frac{360\cdot(q-1)}{K}\leq\varphi<\frac{360\cdot q}{K}. \tag{9}$$

4) Global Dictionary Learning: Similar to the principle of the LBP operator [44], the rotational invariance of the magnitude pattern can be expressed as

$$MP_{P,R}^{riu2}=\begin{cases}\sum_{i=0}^{P-1}C^{\rho}(z_c-z_i), & \text{if } U(MP_{P,R})\leq2,\\ P+1, & \text{else},\end{cases} \tag{10}$$

$$U(MP_{P,R})=\left|C^{\rho}(z_{P-1}-z_c)-C^{\rho}(z_0-z_c)\right|+\sum_{i=1}^{P-1}\left|C^{\rho}(z_i-z_c)-C^{\rho}(z_{i-1}-z_c)\right|, \tag{11}$$

where P, R, and U denote the total number of neighbors for each location, neighborhood radius, and frequency of the 1-0 and 0-1 transformations in a circular representation of the magnitude pattern. There are $P+2$ different values of uniform magnitude patterns representing the structural distribution of the magnitude pattern features.

Then, the global statistical normalized histograms from $PP_{P,R}^{riu2}$ are represented as:

$$H_{MP_{P,R}^{riu2}}(k)=\frac{1}{\Omega}\sum_{\mathrm{x}}f(MP_{P,R}^{riu2}(\mathrm{x}),k), \tag{12}$$

where $f(i,j)=\begin{cases}1, & \text{if } i=j\\ 0, & \text{otherwise},\end{cases}$ and $k\in[0,K]$, where K denotes the maximum value of $PP_{P,R}^{riu2}$. Thus, $K=P+1$. Further, Ω is the position of $PP_{P,R}^{riu2}$.

For the phase response, we can similarly obtain the global statistical normalized histograms of the phase pattern, i.e., $H_{PP_{P,R}^{riu2}}(k)$, using Eqs. (10)–(12).

Finally, by concatenating the above global statistical normalized histograms, the quality feature vectors can be described as

$$\mathbf{b}=[H_{MP_{P,R}^{riu2}},H_{PP_{P,R}^{riu2}}]. \tag{13}$$

We calculate the global quality features for each sample in a given image gallery set [22], [36], [37] using the proposed

global quality features extraction method. We then construct a global dictionary $\mathbf{B}$ for the entire image gallery set by combining all the global quality features that represent the images, such that

$$\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_n] \in R^{\alpha \times n}, \tag{14}$$

where n denotes the total number of samples in the image gallery set and α denotes the feature dimension.

5) CR-Based Global Feature Representation: Given a query image, let $\mathbf{y} \in R^{m \times 3}$ denote its global quality features. Then, $\mathbf{y}$ can be expressed by a linear combination of all the global quality features in the gallery, and can be formulated as

$$\mathbf{y} = \mathbf{B}\mathbf{F}^{global}. \tag{15}$$

Note that because $\mathbf{B}$ is a redundant dictionary, the solution given by Eq. (14) is not unique. To render the sparse coefficient matrix $\mathbf{F}^{global}$ informative and suitable for IQA, some regularization terms should be added to this matrix. In [46], the l_2-minimization-based algorithm is shown to be less complex than the l_1-minimization-based algorithm. Hence, in this work, we utilize the l_2-norm regularization term to solve the optimization problem, leading to the following regularized least-squares instantiation of CR. The resultant model is expressed as

$$\mathbf{F}^{global} = \arg\min_{\tilde{\mathbf{x}}} \{\|\mathbf{y} - \mathbf{B}\tilde{\mathbf{x}}\|_2^2 + \lambda_1 \|\tilde{\mathbf{x}}\|_2^2\}, \tag{16}$$

where λ_1 is a scalar weight. Fortunately, it is analytically and easily derived that Eq. (15) has the closed-form solution

$$\mathbf{F}^{global} = (\mathbf{B}^T\mathbf{B} + \lambda_2\mathbf{I})^{-1}\mathbf{B}^T\mathbf{y}, \tag{17}$$

where $\boldsymbol{I}$ is the identity matrix. Let $\mathbf{P} = (\mathbf{B}^T\mathbf{B} + \lambda_2\mathbf{I})^{-1}\mathbf{B}^T$. Clearly, matrix $\mathbf{P}$ is independent of matrix $\mathbf{y}$, allowing it to be precalculated based on the image gallery set alone. Once a query image matrix $\mathbf{y}$ is considered, matrix $\mathbf{y}$ can be projected onto$\mathbf{P}$ via $\mathbf{Py}$.

C. Prediction Model Based on SVR

As discussed in [48] and [49], visual perception and visual working memory retrieval are highly correlated because they share a certain brain cognitive mechanism. From a practical perspective, we imitate this visual working memory module using the kernel-based SVR. The final objective quality score Q of an SCI can be predicted using $\{\mathbf{F}^{local}, \mathbf{F}^{global}\}$ and a nonlinear prediction function f. That is, the objective quality value representing the objective quality of the SCI is given by $Q = f(\mathbf{F}^{local}, \mathbf{F}^{global})$, where f is pre-estimated using kernel-based SVR. Here, f: $R^2 \rightarrow R$ takes the learned local and global quality features $\{\mathbf{F}^{local}, \mathbf{F}^{global}\}$ as input and outputs a difference mean opinion score. In kernel-based SVR, the unknown f can be learned as

$$f(\mathrm{x}) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(\mathrm{x}_i, \mathrm{x}) + b, \tag{18}$$

where α and α^* are Lagrange multipliers, b is a bias parameter, and $K(\mathrm{x}_i, \mathrm{x})$ is a radial basis kernel function [50].

TABLE I
PARAMETERS SELECTION EXPERIMENT RESULTS

Parameters		Criteria	
a	*b*	PLCC	SROCC
4	-7	0.816	0.776
4	-6	0.828	0.787
4	-5	0.828	0.788
4	-4	0.824	0.784
3	-4	0.822	0.783
2	-4	0.820	0.777

III. EXPERIMENTAL RESULTS AND ANALYSES

A. Experimental Protocol

We conducted several experiments using three IQA datasets: the newly released SIQAD dataset [11], categorical subjective image quality (CSIQ) dataset [51], and cross-content-type (CCT) dataset [31] to validate the effectiveness of the proposed blind metric.

The SIQAD dataset contains 20 reference SCIs altered by seven types of distortion, each with seven levels of degradation. Thus, this dataset consists of 140 distorted SCIs, comprising 20 reference SCIs distorted by Gaussian noise (GN), motion blur (MB), Gaussian blur (GB), contrast change (CC), JPEG2000 compression (JP2K), JPEG compression (JPEG), and layer-segmentation-based coding (LSC). For further details on this dataset, see [10].

The CSIQ dataset consists of 866 NISs and six types of distortion: white noise (WN), JPEG, JP2K, additive Gaussian pinknoise (PN), GB, and global contrast decrements (GCD). The difference mean opinion score (DMOS) of each NSI is included.

The CCT dataset, contains 1,320 distorted images (e.g., SCIs, NSIs, and others) and associated DMOS values generated from 72 pristine images distorted with two distortion types, at various levels of distortion.

Two typical performance criteria were employed to evaluate the IQA metrics tested in these experiments: i) Pearson's linear correlation coefficient (PLCC), which evaluates prediction accuracy, and ii) Spearman's rank order correlation coefficient (SROCC), which reflects the prediction monotonic of IQA metrics. Higher PLCC and SROCC values values indicate superior correlation performance. Thus, a high-performance objective model has the PLCC and SROCC values are close to 1.

As the proposed blind metric requires training and testing, a cross-validation test was implemented by randomly splitting each dataset into two non-overlapping subsets: training (80%) and test (20%). To offset the performance bias as much as possible, this 80:20 split of the data was iterated 1,000 times and the median PLCC and SROCC values of the 1,000 trials are reported. When deploying SVR [49] to train the prediction module, the parameters ($\mathrm{C} = 2^a$, $\gamma = 2^b$) of SVR must be set. To select the values of (a, b), a grid search was used. Some of the experimental results on the SIQAD dataset are listed in Table I. The parameters ($a = 4$, $b = -5$) obtained the best results and were used in the following experiments.

The parameters of the log-Gabor filter were set as: $\sigma_\theta = 0.71$, $\sigma_r = 0.61$, $\omega_0^1 = 0.417$, $\omega_0^2 = 0.318$, and $\omega_0^3 = 0.243$,

TABLE II

PERFORMANCE OF THE PROPOSED BLIND METRIC AND THE OTHER TWELVE METRICS USING THE SIQAD DATASET

Metrics	Criteria	
	PLCC	SROCC
PSNR	0.587	0.560
SSIM	0.591	0.582
SPQA	0.858	0.842
SFUW	0.891	0.880
ESIM	0.879	0.863
QAC	0.375	0.301
NIQE	0.342	0.370
IL_NIQE	0.388	0.322
BQMS	0.755	0.722
GM_LOG	0.720	0.663
DIIVINE	0.691	0.659
LGP	0.784	0.754
Proposed	**0.828**	**0.788**

TABLE III

MEAN AND STANDARD DEVIATION OF THE SROCC VALUES ACROSS 1,000 TRIALS

Criteria	GM_LOG	LGP	Proposed
Mean	0.663	0.753	0.787
Std	0.0715	0.0666	0.0359

where ω_0^1, ω_0^2, and ω_0^3 denote the three center frequencies on three scales. In this work, the scale and orientation of the filter were set to $n = 3$ and $j = 4$, respectively.

B. Performance Comparison on the SIQAD Dataset

To comprehensively evaluate the improvement in prediction of the proposed blind metric, the performance in terms of the SROCC and PLCC was compared with that of the following previous objective metrics: five FR IQA metrics (peak signal to noise ratio (PSNR), structural similarity index (SSIM) [11], SPQA [10], SFUW [12], and ESIM [17]) and six blind IQA metrics (QAC [20], NIQE [21], IL_NIQE [22], GM_LOG [23], local gradient patterns (LGP) [24], DIIVINE [52], and BQMS [29]). The PLCC and SROCC values for the given SIQAD are summarized in Table II. The best results in each case are highlighted in bold font. Hence, these experimental results indicate that the proposed blind metric can achieved comparative and reasonably encouraging quality predictions for distorted SCIs compared to other blind metrics. Specifically, it is well known that GM_LOG and LGP perform well for natural images. However, these metrics do not perform well when applied to SCIs. Moreover, compared with FR metrics, the proposed blind metric achieves competitive performance with the top-performing FR metric because it considers the local and global statistical properties of the SCIs. Furthermore, Table III shows the mean and standard deviation of the SROCC values across the 1,000 trials; higher means with lower standard deviations indicate outstanding prediction performance. In summary, the proposed blind metric quantifies and predicts perceptual distortions in SCIs stably.

To more comprehensively evaluate an IQA metric's ability to predict the perceptual quality of SCIs caused by different types of distortions, we examined the prediction performance of our blind metric against the competing metrics (PSNR, SSIM, SPQA, SFUW, ESIM, QAC, NIQE, IL_NIQE, GM_LOG, DIIVINE, and LGP) on specific types of distortions. The PLCC and SROCC results are presented (Tables IV and V, respectively). From the tables, it can be see that, compared with existing IQA metrics, the proposed blind metric can better handle GN, GB, MB, JPEG, and JP2k distortions. Possibly because of the quality features, which cannot efficiently and effectively represent the visual quality distortion of contrast or shape changes, the proposed blind metric is not good at handling CC and LSC distortions. However, although some blind metrics are good at handling certain individual distortions (CC and LSC), the proposed blind metric is clearly competitive with the blind metric that obtain very promising performance on CC and LSC distortions. In general, the proposed blind metric performs better than, or has comparable prediction performance to, the classical FR metrics.

C. Performance Comparison on the CSIQ and CCT Datasets

In this section, we tested our proposed blind metric on the CSIQ and CCT datasets to demonstrate that the proposed model has a strong ability to handle cross-content-type images. Table VI summarizes the test results on these two datasets, which lead to several useful findings. First, on the CSIQ dataset, which contains NISs only, the results indicate that the learned local and global quality features are efficient for NSIs. Second, on the CCT dataset, which contains cross-content-type images (e.g., NSIs, SCIs, and others), it is interesting to note that the conventional handcrafted features cannot represent cross-content-type images, which contain not only natural scenes but also text, tables, icons, and graphics. However, the learning-based features are able to catch the specific information with various degrees of distortion for text, tables, icons, and graphics. It is efficient for the analysis of NSIs and SCIs under these circumstance not to include any prior knowledge about the image type. In summary, the results show that the proposed blind metric has a generalization ability that can be applied to cross-content-type images, including SCIs, NSIs, and other types.

D. Contribution of Learned Local and Global Features to the Proposed Blind Metric

To better understanding the individual contributions of the learned local and global features in the proposed blind metric, we designed two test metrics, A and B, in which only the learned local or global features, respectively, were used to estimate the visual quality. Tables VII, VIII, and IX show the performance of metrics A and B, along with that of the proposed blind metric on the SIQAD dataset. For all distortions, it can be see that the performance can be promoted by correctly and properly integrating the learned local and global features. For each type of distortion, we were surprised to see that metric B obtained promising performances for

TABLE IV
OVERALL PERFORMANCE OF ELEVEN METRICS FOR EACH TYPE OF DISTORTION (PLCC)

Distortion	FR					Blind						
	PSNR	SSIM	SPQA	SFUW	ESIM	QAC	DIIVINE	NIQE	IL_NIQE	GM_LOG	LGP	Proposed
GN	0.905	0.866	0.892	0.887	0.899	0.853	0.872	0.805	0.815	0.898	0.892	**0.903**
GB	0.860	0.901	0.906	0.923	0.923	0.559	0.853	0.480	0.540	0.886	0.856	**0.911**
MB	0.704	0.806	0.832	0.878	0.889	0.378	0.803	0.201	0.359	0.678	0.789	**0.837**
CC	0.753	0.756	0.799	0.829	0.764	0.074	0.672	0.118	0.138	0.664	**0.731**	0.660
JPEG	0.770	0.749	0.770	0.757	0.800	0.302	0.631	0.427	0.268	0.689	0.628	**0.762**
JP2K	0.789	0.775	0.825	0.815	0.789	0.189	0.636	0.185	0.418	0.571	0.619	**0.668**
LSC	0.781	0.731	0.796	0.759	0.792	0.337	0.519	0.418	0.196	0.565	**0.748**	0.683

TABLE V
PERFORMANCE OF ELEVEN METRICS FOR EACH TYPE OF DISTORTION (SROCC)

Distortion	FR					Blind						
	PSNR	SSIM	SPQA	SFUW	ESIM	QAC	DIIVINE	NIQE	IL_NIQE	GM_LOG	LGP	Proposed
GN	0.879	0.869	0.886	0.869	0.876	0.842	0.862	0.824	0.816	0.878	0.868	**0.879**
GB	0.858	0.893	0.912	0.917	0.924	0.624	0.849	0.567	0.456	0.879	0.841	**0.894**
MB	0.713	0.804	0.844	0.874	0.894	0.338	0.799	0.375	0.446	0.665	0.779	**0.832**
CC	0.683	0.641	0.638	0.722	0.611	0.075	0.498	0.069	0.044	0.551	**0.615**	0.487
JPEG	0.757	0.758	0.771	0.750	0.799	0.145	0.626	0.447	0.287	0.691	0.611	**0.744**
JP2K	0.775	0.760	0.844	0.812	0.783	0.134	0.628	0.247	0.381	0.576	0.602	**0.645**
LSC	0.793	0.737	0.859	0.754	0.796	0.187	0.514	0.344	0.168	0.561	**0.723**	0.666

TABLE VI
PERFORMANCE COMPARISON ON THE CSIQ AND CCT DATASETS

Dataset	Criteria	DIIVINE	CORNIA	NIQE	IL_NIQE	LGP	Proposed
CSIQ	PLCC	0.898	0.918	0.716	0.917	0.921	**0.922**
	SROCC	0.876	0.893	0.627	0.889	0.893	**0.906**
CCT	PLCC	0.555	0.581	0.548	0.501	0.659	**0.674**
	SROCC	0.508	0.487	0.234	0.166	0.621	**0.641**

TABLE VII
PERFORMANCE OF EACH FEATURE IN THE PROPOSED BLIND METRIC FOR ALL DISTORTIONS ON THE SIQAD DATASET

Criteria	Metric A	Metric B	Proposed
PLCC	0.821	0.809	0.828
SROCC	0.783	0.769	0.788

TABLE VIII
PERFORMANCE OF EACH FEATURE IN THE PROPOSED BLIND METRIC FOR EACH DISTORTION TYPES (PLCC) ON THE SIQAD DATASET

Distortion	Metric A	Metric B	Proposed
GN	0.892	**0.914**	0.903
GB	0.902	0.887	**0.911**
MB	0.826	**0.875**	0.837
CC	**0.671**	0.335	0.660
JPEG	0.740	**0.819**	0.762
JP2K	0.678	0.666	**0.668**
LSC	0.682	**0.784**	0.683

the GN, MB JPEG, and LSC distortions, while metric A performed very poorly. Therefore, the learned local and global features are complementary, reflecting different attributes of the HVS for quality prediction.

E. Statistical Significance Comparison

To evaluate the statistical significance of the proposed blind metric's advantages over other blind IQA metrics, F-tests

TABLE IX
PERFORMANCE OF EACH FEATURE IN THE PROPOSED METRIC FOR EACH DISTORTION TYPES (SROCC) ON THE SIQAD DATASET

Distortion	Metric A	Metric B	Proposed
GN	0.866	**0.900**	0.879
GB	0.885	0.869	**0.894**
MB	0.824	**0.870**	0.832
CC	**0.492**	0.312	0.487
JPEG	0.720	**0.809**	0.744
JP2K	0.661	0.642	**0.645**
LSC	0.660	**0.790**	0.666

TABLE X
RESULTS OF F-TEST COMPARING SROCC VALUES OF VARIOUS METRICS ON THE SIQAD DATASET

	IL_NIQE	BQMS	GM_LOG	LGP	Proposed
IL_NIQE	0	-1	-1	-1	-1
BQMS	1	0	1	-1	-1
GM_LOG	1	-1	0	-1	-1
LGP	1	1	1	0	-1
Proposed	1	1	1	1	0

were conducted at the 5% significance level. The validation results on the SIQAD dataset are summarized in Table X. Symbol "1" denotes that the row metric is significantly better than the column metric, "–1" means that the row metric is significantly worse than the column metric, and "0" indicates that the row metric is significantly indistinguishable with the column metric. It is clear that outstanding performance was achieved by our proposed metric, whose results are "1" for all comparisons.

F. Complexity Analysis

The computational complexity is another important consideration when evaluating the computational performance of

TABLE XI
RUNNING TIMES OF THE COMPARED BLIND IQA METRICS ON THE SIQAD DATASET

Metric	Running times (s)
QAC	0.16
NIQE	0.56
IL_NIQE	10.84
BQMS	93.55
GM_LOG	0.14
LGP	1.89
Proposed	10.77

our proposed blind metric. The testing environment consisted of a 2.70 GHz Intel Core i5 CPU processor, 8 GB RAM, and the MATLAB R2010 platform. The experimental results on the SIQAD dataset are listed in Table XI, where we list the average computation time per SCI image. As Table XI clearly demonstrates, our proposed blind metric has medium computational complexity.

G. Discussions

1) The current computational complexity of the proposed metric is not very low. The most time-consuming part of the metric is the extraction of the log-Gabor base features. We could replace these features with less computationally expensive features (e.g., gradient based features) in future work. Furthermore, because the local and global feature methods can be executed simultaneously, and parallel computing could be used to improve the speed of the proposed metric.

2) Blind quality evaluation for NSIs and SCIs is still in a preliminary stage. Hence, we have sufficient space for improvement. In this work, only local and global visual characteristics for NSIs and SCIs were investigated, while the visual characteristics of NSIs and SCIs such as visual saliency and the visual statistical model are different from those of natural images. Further exploration of human visual physiology and psychology may bring more inspiration for establishing a model with more specific characteristics.

3) Table VII shows that the adding global features only improves the performance a little for all distortions; however, Tables VIII and IX show that the main contribution of the performances on GN, MB JPEG, and LSC distortions may come from the global feature. Metric B is marginally inferior to metric A for all distortions, perhaps because the global feature cannot efficiently and effectively represent the visual quality distortion for contrast change. Further, how to construct a more effective global feature for CC distortion should be considered.

IV. CONCLUSION

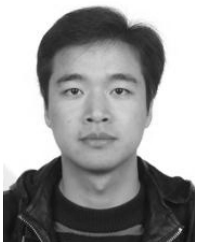
In this paper, we have put forward an effective blind quality predictor for distorted images by incorporating the learning of both local and global quality features. The novelty of our research resides in combining the complementary behaviors of the locality-constrained linear coding (LLC)-based local quality features and collaborative representation (CR)-based global quality features to achieve a fused representation for images. To extract the learned local features, we use a strategy of LLC with max pooling. Further, the learned global features of the distorted images can be obtained using the CR method. Compared with competing IQA metrics, experimental results show that the proposed blind metric can obtain significantly higher statistical consistency with evaluations from human subjects, confirming that the devised metric is a very robust blind IQA metric for NSIs and SCIs.

Various aspects of the present research merit further investigation and will, therefore, be considered in future work. In the feature extraction stage, we will focus on mining the special characteristics of the NSIs and SCIs more deeply. In the perceptual-quality prediction stage, we intend to design blind IQA metrics for NSIs and SCIs based on deep learning approaches.

REFERENCES

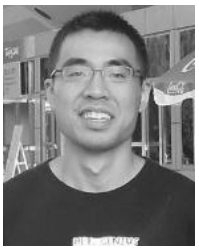
[1] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838–3851, Aug. 2016.
[2] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 50–62, Jan. 2016.
[3] 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.
[4] Z. Ma, W. Wang, M. Xu, and H. Yu, "Advanced screen content coding using color table and index map," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4399–4412, Oct. 2014.
[5] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Hash-based block matching for screen content coding," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 935–944, Jul. 2015.
[6] 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.
[7] C.-C. Chen and W.-H. Peng, "Intra line copy for HEVC screen content intra-picture prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 7, pp. 1568–1579, Jul. 2017.
[8] S. Wang *et al.*, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.
[9] H. Yang, Y. Fang, Y. Yuan, and W. Lin, "Subjective quality evaluation of compressed digital compound images," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 105–114, Jan. 2015.
[10] 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.
[11] 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.
[12] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, Apr. 2017.
[13] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
[14] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Comput. Graph. Appl.*, to be published.
[15] 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.
[16] X. Guo, L. Huang, K. Gu, L. Li, Z. Zhou, and L. Tang, "Naturalization of screen content images for enhanced quality evaluation," *IEICE Trans. Inf. Syst.*, vol. 100, no. 3, pp. 574–577, 2017.
[17] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge similarity for screen content image quality assessment," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4818–4831, Oct. 2017.
[18] 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.

[19] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.
[20] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 995–1002.
[21] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[22] 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.
[23] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
[24] W. Zhou, L. Yu, W. Qiu, Y. Zhou, and M.-W. Wu, "Local gradient patterns (LGP): An effective local-statistical-feature extraction scheme for no-reference image quality assessment," *Inf. Sci.*, vols. 397–398, pp. 1–14, Aug. 2017.
[25] 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.
[26] A. Saha and Q. M. J. Wu, "Utilizing image scales towards totally training free blind image quality assessment," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1879–1892, Jun. 2015.
[27] D. Lee and N. K. Plataniotis, "Toward a no-reference image quality assessment using statistics of perceptual color descriptors," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3875–3889, Aug. 2016.
[28] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "CID2013: A database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.
[29] 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.
[30] J. Qian, L. Tang, V. Jakhetiya, Z. Xia, K. Gu, and H. Lu, "Towards efficient blind quality evaluation of screen content images based on edge-preserving filter," *Electron. Lett.*, vol. 53, no. 9, pp. 592–594, 2017.
[31] 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 Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.
[32] K. Gu, J. Zhou, J.-F. 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.
[33] H. Gu, G. Zhai, M. Liu, and K. Gu, "Exploiting global and local information for image quality assessment with contrast change," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2015, pp. 1–5.
[34] M. Liu, G. Zhai, K. Gu, and X. Yang, "Learning to integrate local and global features for a blind image quality measure," in *Proc. IEEE Int. Conf. Smart Comput. (ICSC)*, Nov. 2014, pp. 49–53.
[35] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 506–510.
[36] S. Shi, X. Zhang, S. Wang, R. Xiong, and S. Ma, "Study on subjective quality assessment of screen content images," in *Proc. Picture Coding Symp. (PCS)*, May/Jun. 2015, pp. 75–79.
[37] C. Shen and Q. Zhao, "Webpage saliency," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 33–46.
[38] 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.
[39] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 1098–1105.
[40] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.
[41] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: An evaluation of recent feature encoding methods," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2011, pp. 1–12.
[42] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[43] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3360–3367.
[44] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[45] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: Energy models, position shifts and phase shifts," *Vis. Res.*, vol. 36, no. 12, pp. 1839–1857, 1996.
[46] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2379–2394, 1987.
[47] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 471–478.
[48] S. Söderqvist, H. Matsson, M. Peyrard-Janvid, J. Kere, and T. Klingberg, "Polymorphisms in the dopamine receptor 2 gene region influence improvements during working memory training in children and adolescents," *J. Cognit. Neurosci.*, vol. 26, no. 1, pp. 54–62, 2014.
[49] P. J. Olesen, H. Westerberg, and T. Klingberg, "Increased prefrontal and parietal activity after training of working memory," *Nature Neurosci.*, vol. 7, no. 1, pp. 75–79, 2004.
[50] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[51] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, 2010.
[52] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.



Wujie Zhou received the Ph.D. degree from Ningbo University, Ningbo, China, in 2014. He is currently an Associate Professor with the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, China. He is also a Post-Doctoral Fellow with the Institute of Information and Communication Engineering, Zhejiang University, China. His research interests include multimedia signal processing and communication. He is a reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON BROADCASTING, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS, the *Information Sciences*, the *Neurocomputing*, and SPIC.

Lu Yu received the B.Eng. degree in radio engineering and the Ph.D. degree in communication and electronic systems from Zhejiang University, Hangzhou, China, in 1991 and 1996, respectively. She is currently a Professor with the Institute of Information and Communication Engineering, Zhejiang University. Her current research interests include video coding, multimedia communication, and relative application-specific integrated circuit design.



Yang Zhou received the B.S. degree in automation from Zhejiang Normal University, Jinhua, China, in 2006, the M.S. degree in communication from Beihang University, Beijing, China, in 2009, and the Ph.D. degree from the Shanghai University of Science and Technology, Shanghai, China, in 2013. He is currently a Lecturer with the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, China. His research interests include signal processing and optical measurement technology.

Weiwei Qiu received the B.E. and Ph.D. degrees in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2007 and 2013, respectively. She is currently a Lecturer with the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, China. Her current research focuses on optical fiber sensors and optical signal processing.

Ming-Wei Wu received the B.E. degree (Hons.) in electrical engineering, the M.E. and Ph.D. degrees in electrical engineering from the National University of Singapore, in 2000, 2003, and 2011, respectively. From 2002 to 2004, she was a Research Engineer with the Institute for Infocomm Research, Singapore, where she focused on ethernet passive optical networks standardization and implementation. In 2004, she joined the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, China, as an Associate Professor. Her research interests include wireless communication, detection, and estimation theory and performance analysis. She received the Best Paper Award at the IEEE ICC2011, Kyoto, Japan. She serves as a Technical Program Committee Member for the international communications conferences including the IEEE ICC2011, the IEEE ICC2012, the IEEE VTC2013-Spring, and the IEEE ICCC2012.

Ting Luo received the B.S. degree in computer science from Ningbo University, Ningbo, China, in 2003, the M.S. degree in business information technology from Middlesex University, London, U.K., in 2004, and the Ph.D. degree from Ningbo University in 2016. He is currently an Associate Professor with the College of Science and Technology, Ningbo University. His research interests include multimedia security, image processing, data hiding, and pattern recognition.