Full-Reference Image Quality Assessment by Combining Features in Spatial and Frequency Domains

Zhisen Tang, Yuanlin Zheng, Ke Gu, *Member, IEEE*, Kaiyang Liao, Wei Wang, and Miaomiao Yu

***Abstract*—Objective image quality assessment employs mathematical and computational theory to objectively assess the quality of output images based on the human visual system (HVS). In this paper, a novel approach based on multifeature extraction in the spatial and frequency domains is proposed. We combine the gradient magnitude and phase congruency maps to generate a local structure (LS) map, which can perceive local structural distortions. The LS matches well with HVS and highlights differences with details. For complex visual information, such as texture and contrast sensitivity, we deploy the log-Gabor filter, and spatial frequency, respectively, to effectively capture their variations. Moreover, we employ the random forest (RF) to overcome the limitations of existing pooling methods. Compared with support vector regression, RF can obtain better prediction results. Extensive experimental results on the five benchmark databases indicate that the proposed method precedes all the state-of-the-art image quality assessment metrics in terms of prediction accuracy. In addition, the proposed method is in compliance with the subjective evaluations.**

***Index Terms*—Image quality assessment (IQA), log-Gabor filter, full-reference, contrast sensitivity function (CSF), random forest (RF).**

I. Introduction

With the high-speed development of image processing and communication systems, image quality assessment (IQA) has played a significant role in plentiful visual signals applications [1], including image acquisition, compression, transmission, enhancement, and watermarking. Image quality is inevitably affected by a variety of distortions that degrade the perceived visual quality in the above processes. Therefore, accurately measuring visual quality is very important. Since humans are the ultimate receiver of visual signals, it is an effective technique for judging image quality. However, subjective assessment is time consuming, expensive and laborious, and thus it is urgent to contrive objective image quality assessment approaches that measure image quality automatically. These approaches must maintain consistency with the human visual system (HVS). These objective image quality assessment metrics can be widely applied to image monitoring systems and parameters optimization in image processing algorithms [2].

On the basis of the degree of availability of reference image information, objective IQA can be divided into three categories: full reference (FR) [3]–[5], reduced reference (RR) [6], [7], and no reference (NR) [8]–[12] methods. This paper only discusses the topic of full reference image quality assessment. The FR-IQA models include a wide range of real applications. The main two applications are provided as follows.

First, a well-designed FR-IQA model can be used for optimizing the coding technologies [52]. The second application of FR-IQA models, as noted in several recently proposed studies [53], [54], can be deployed as weak supervision for learning robust NR-IQA models.

The development processes for FR-IQA, peak signal-to-noise ratio (PSNR) and mean-squared error (MSE) are the earliest and most widely used algorithms; they measure the intensity change between pixels. Although these two methods have low complexity, they are not well aligned with subjective perceptions of visual quality, i.e., the Human Visual System (HVS). Thus, the vast majority of algorithms based on the HVS have been put forward recently. Structural similarity (SSIM) [13] is a milestone in many algorithms based on HVS; it predicts visual quality by capturing changes in luminance, contrast, and structural information. Meanwhile, its extended versions, namely, multiscale SSIM (MSSIM) [14], measures structural information by embedding multiscale space in an image. Information content weighted SSIM (IW-SSIM) [15] improves image quality prediction by introducing weighted local information. Information theoretic approaches address the IQA problem from the perspective of information

Manuscript received January 13, 2018; revised April 8, 2018 and June 28, 2018; accepted September 4, 2018. Date of publication October 2, 2018; date of current version March 4, 2019. This work was supported in part by the National Natural Science Foundation of China Project under Grant 61671376, and in part by the Natural Science Foundation of Shaanxi Province under Grant 2016JM6022. *(Corresponding author: Yuanlin Zheng.)*

Z. Tang, W. Wang, and M. Yu are with the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China (e-mail: zhisentang@163.com).

Y. Zheng is with the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China, and also with the Key Laboratory of Printing and Packaging Engineering of Shaanxi Province, Xi'an University of Technology, Xi'an 710048, China (e-mail: zhengyuanlin@xaut.edu.cn).

K. Gu is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: guke.doctor@gmail.coms).

K. Liao is with the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China, and also with the Printing and Packaging Engineering Technology Research Centre of Shaanxi Province, Xi'an University of Technology, Xi'an 710048, China (e-mail: 17878837@qq.com).

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/TBC.2018.2871376

communication and the amount of information shared within the reference image and the distorted image. In [16] and its extended version, visual information fidelity (VIF) [17] quantifies the information of the reference image and then compares how much information similar to this is available in a distorted image. Most apparent distortion (MAD) [18] considers the image distortions to be classified as near-threshold and supra-threshold; the MAD uses two quality indexes to obtain the overall image quality score.

Many studies [13], [19]–[21] show that the HVS is strongly sensitive to the structural information in an image. Gradient similarity measure (GSM) [20] captures image quality by combining gradient and luminance information. Supposing that the HVS is susceptible to the low-level features in the image, feature-similarity (FSIM) [21] uses gradient magnitude (GM) and phase congruency (PC) features as complementary aspects. Similarly, in visual saliency-induced index (VSI) [22], visual saliency (VS) and GM are utilized to measure the distortion degree. Unlike the average pooling in SSIM, the GSM, FSIM, and VSI use a weighting function to improve IQA accuracy to some extent, but it may increase algorithmic complexity. Hence, gradient magnitude similarity deviation (GMSD) [23] only calculates the GM similarity of a reference image and the corresponding distorted image and then utilizes the standard deviation of GM similarity as the quality score. The GMSD [23] is much faster than the algorithms mentioned above, but it cannot measure color image quality. In addition, it is not enough for GMSD, FSIM and VSI to describe image structural information by using GM, PC and VS, respectively. At the same time, one common shortcoming is that the prediction accuracies of these methods are too low. To overcome the above problems, the proposed algorithm captures the image structure by combining spatial and frequency domain features; i.e., gradient magnitude, phase congruency and texture features are computed as structural information. The FSIM introduces two feature maps to describe structure in images. Compared with the FSIM, the proposed method, which is motivated by a previous work [11], combines the GM and PC maps to obtain a localized edge map. Although [11] is a NR-IQA model that can only be used to evaluate blurred images, the localized edge map proposed by [11] is also applied to the FR-IQA model. Considering texture as important structural information, the Log-Gabor filter, which agrees with the HVS, characterizes textural information. Meanwhile, this paper employs three features to simulate contrast sensitivity changes of HVS that have different sensitivities to distortion depending on spatial frequency [24]. Details can be found in Sections II and III.

This paper makes two contributions. 1) We consider the properties of HVS in spatial and frequency domains, and depict image quality by multiple complementary features in spatial and frequency domains. We overcome the common limitation of most existing FR-IQA models that only analyze image quality from spatial or frequency domain. We use these features include GM and color information in the spatial domain, texture energy maps extracted from log-Gabor features and a descriptor of contrast sensitivity information in the frequency domain; 2) The learning-based scheme, i.e., random forest (RF) is adopted for the distortion effects pooling. Human vision knowledge toward IQA is incorporated into the process of distortion pooling with the help of RF. RF-based pooling technique has the ability to overcome conventional pooling limitation (summation or multiplication operation may cause the relationship between the distortions and the quality score to be linear in the conventional pooling strategies). Compared with other state-of-the-art models, our experiments prove the effectiveness and robustness of our approach across various databases; and it correlates well with the subjective evaluations.

The rest of this paper is organized as follows. Section II briefly reviews two-stage frameworks in existing FR-IQA approaches and describes the related schemes of this paper. The details of the proposed model are presented in Section III. Experiments and analysis are presented in Section IV. Finally, conclusions are presented in Section V.

II. Related Work

In general, these existing FR-IQA models are typically made up of two-stage frameworks. The first stage is to measure the local distortions and calculate similarity, and the second stage (pooling strategy) is to convert the local similarity maps into the final quality score. This section briefly summarizes structural feature extraction and pooling strategies of existing FR-IQA approaches and presents the related schemes of this paper.

As depicted above, the gradient has the ability to reflect structure and contrast change. In the first stage, the gradient is usually used to measure the structural information in the spatial domain. At the same time, it is not enough to express the structural information strictly in terms of the gradient. Therefore, the FSIM [21] introduces phase congruency (PC) in the frequency domain to act as a complementary feature. As shown in Fig. 1, Fig. 1(b)-(c) are the GM and PC maps of reference image (a), respectively. Unfortunately, FSIM cannot completely represent the image structure information with only one structure map. Inspired by a previous work [11], we combine the normalized GM and PC maps to obtain a localized structure map (LS) that can perceive local structural distortions.

In other words, Fig. 1(b)-(c) are replaced by Fig. 1(d), which combines the advantages of Fig. 1(b)-(c). It can be seen from the red square area in Fig. 1(b) that the edge of the tower cannot be accurately captured because both sides of the edge of the tower have similar luminance (white cloud and white tower). At this point, PC is able to extract the edge of the tower well, as observed from Fig. 1(c). On the other hand, the blue and yellow square areas in Fig. 1(b), the structures of the house wall and fences can be captured by GM in the spatial domain. In this regard, it is difficult for PC to extract structures because the blue and yellow square areas of Fig. 1(c) have similar frequency behaviors. The local structure (LS) map addresses the above issues. It uses GM and PC features as two complementary aspects of HVS to maximally combine GM and PC so that it can represent the structures comprehensively.

Fig. 1. Illustration of the reference image and its GM, PC and LS maps. (a) Reference image, (b) GM map, (c) PC map, (d) LS map.

The second stage (pooling strategy) is to convert these single quality scores into the final quality score. The pooling strategies of existing approaches are generally based on average pooling [13], [14] and weighted pooling [15], [25], [26]. From the perspective of average pooling, all regions of the image are treated as having the same importance. In fact, these different regions contribute differently to perceptual visual quality. In [21] and [22], PC and VS are used as weight functions to embody the significance of local regions in the image, respectively. Percentile pooling [27] was also proposed to improve algorithmic accuracy to some extent. Furthermore, standard deviation pooling was successfully applied to [23] and [28]; GMSD [23] is one of the most typical examples with low complexity.

Regrettably, the above pooling strategies fuse different distortion factors by summations or multiplications, which depend on appropriate weighting coefficients, and there is no universal approach available for this. Meanwhile, a summation or multiplication operation may cause the relationship between the distortions and the image quality score to be linear. In other words, such pooling strategies seem to be ad-hoc with limited theoretical grounds. Fortunately, pooling techniques based on machine learning have the ability to address these problems. A convolutional neural network (CNN)-based model [50] is able to automatically predict the object quality score by computing the difference in high-frequency relevance information between the reference image and its distorted version. Local linear model (LLM) [50] obtains higher prediction accuracy than other methods on some databases. Because non-negative matrix factorization (NMF) [33] can effectively reflect high-level visual perception, it is used to measure image degradation. In [33], the extreme learning machine (ELM) is employed for fusing the distortion effects and subjective scores to predict object image quality. Compared with other learning-based models (e.g., neural networks and support vector regression), ELM shows faster learning speed in [33]. Moreover, support vector regression (SVR)-based models [9], [10], [29], [31], [51] are widely used in the field of image quality assessment. In [29], SVR is implemented for synthesis of multiple distortion effects and mapping similarity vectors into a quality score. Similarly, SVR is used to regress distortion effects that are represented by singular vectors in [31]. To overcome the limitations of a single approach, multi-method fusion (MMF) [51] combines multiple object image quality models using SVR. Although the overall performance of this model is not bad, there is no single model that performs well on all databases. For CNN and SVR, some additional parameters (e.g., convolution layers/kernels, pooling/fully connected layers in CNN [34], kernel parameters γ and C in SVR [31].) need to be tuned, and this induces heavy computation.

In order to solve the problems above, we use random forests (RF) [40] to deduce a mathematical function to simulate the relationship between distortion factors and image quality. Because there are only two parameters in RF, the parameters are set by default. In addition, the RF regression algorithm is composed of a multitude of regression trees, which is simple and easily parallelized. RF has shown prominent capability of precise mapping functions in IQA [32]. It has been proved that the RF regression significantly surpasses SVR in [32]. This is why we choose RF as a regression tool instead of SVR in this paper (in fact, we find that the

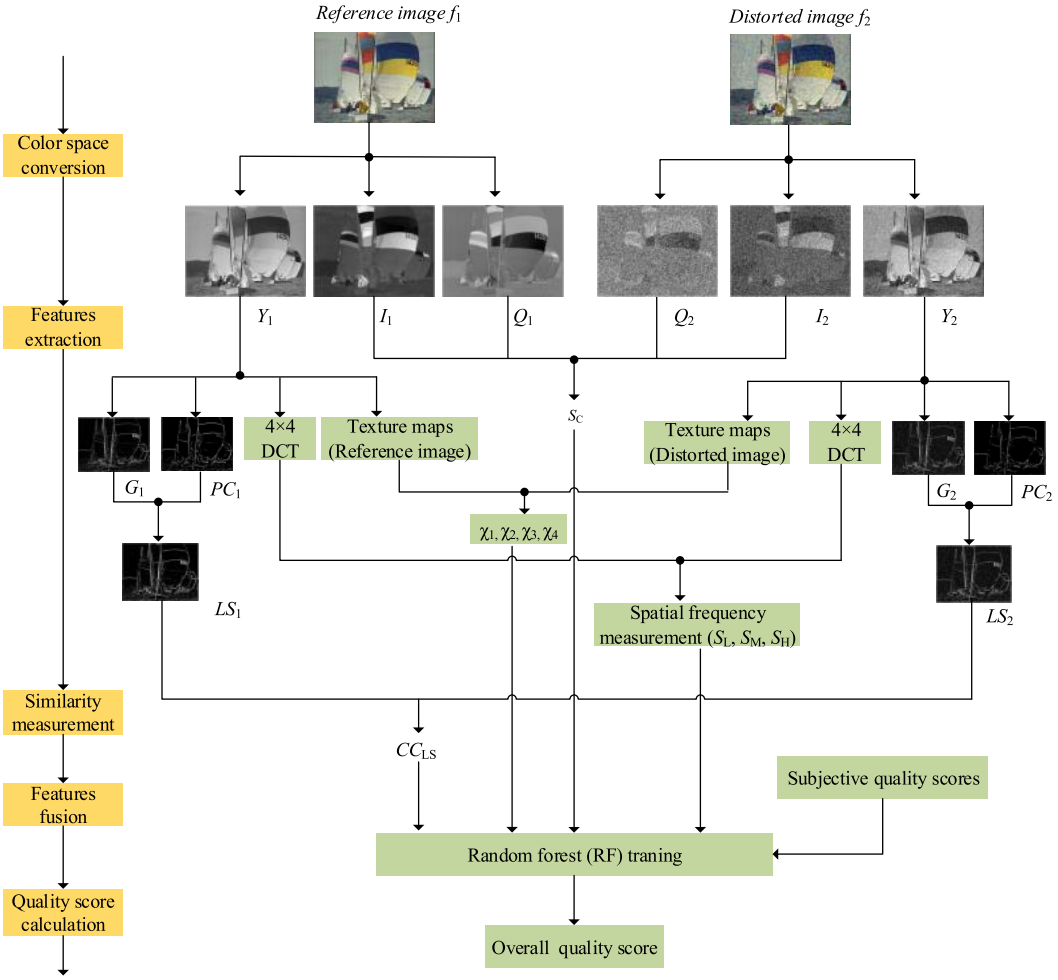

Fig. 2. The flow chart of the proposed algorithm.

performance of RF notably outperforms SVR). Furthermore, the performance comparison of RF and SVR is also shown in Section IV.

It is well known that HVS is highly adapted for extracting structural information from nature scenes and has the multi-scale, multi-direction, masking effect and band-pass properties of contrast sensitivity. These properties of HVS are reflected in the form of spatial and frequency domains. Most of the existing FR-IQA models [13]–[17], [20], [23] have a common limitation that only analyze image quality from the spatial or frequency domain. Because the properties of HVS are not fully considered in the spatial-frequency domain, the results predicted by these methods are far less than ideal. We combine existing approaches [11], [24], [29] to overcome the limitations of a single approach in this paper. Namely, GM in the spatial domain, the PC, texture, and spatial frequency features in the frequency domain are employed to describe the properties of HVS.

III. Proposed Method

A novel FR-IQA scheme by combining spatial and frequency domain features is proposed in this paper, referred to as the perceptual feature similarity index (PFSI). Before the extraction GM, PC, texture and spatial frequency features, the input RGB color images are converted into the YIQ space [35]. It is noteworthy that the GM, PC, texture and spatial frequency features are extracted from the luminance channel. Specifically, texture and spatial frequency features are extracted using log-Gabor and discrete cosine transform (DCT) coefficients in a 4×4 DCT block, respectively. Then, the image color similarity is measured in the I and Q channels. Finally, RF is employed as a regression tool fusing similarity vectors and mean opinion scores (MOS) to predict the final quality score. The flow chart of the proposed algorithm is exhibited in Fig. 2. Note that the f_1 and f_2 mentioned in this paper represent the reference image and the distorted image, respectively.

A. Phase Congruency (PC) Extraction From Frequency Domain

Psychophysical and physiological studies have affirmed that the various feature points can be detected by phase congruency (PC). The main principle of the scheme is the following: The visual features correspond to those points where the Fourier components are maximal in phase [36], [37]. Moreover, PC is dimensionless and is thus insensitive to the contrast and brightness information. Therefore, PC is

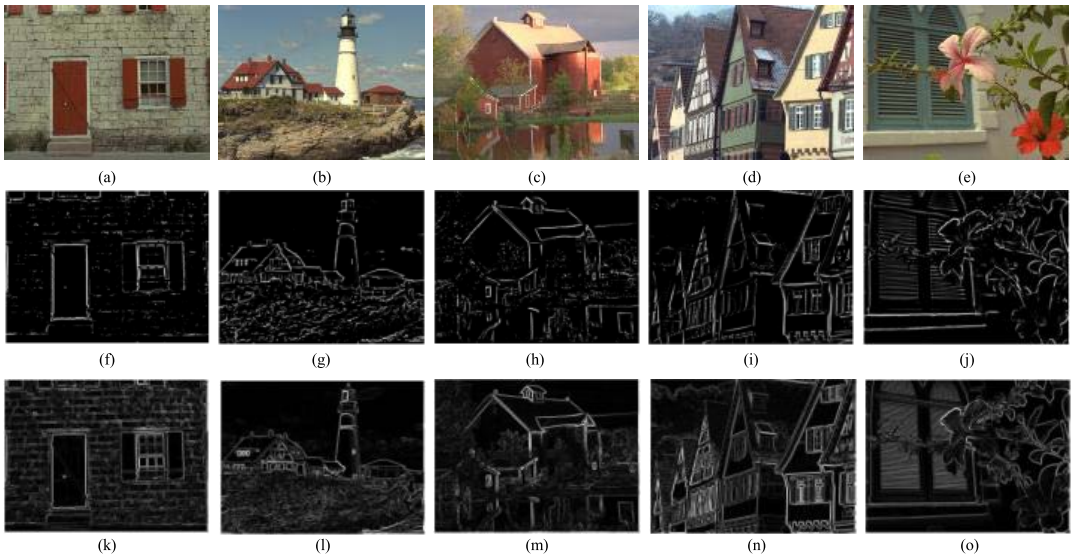

Fig. 3. Illustration of the original image and its phase congruency and gradient magnitude maps. (a)-(e) original images with different image content, (f)-(j) PC maps extracted from (a)-(e), respectively. The GM maps of (a)-(e) are shown in (k)-(o), respectively.

employed as a feature detector in [11], [21], and [38]. This paper utilizes the method proposed by [37] to calculate PC. The calculation process is as follows.

To obtain PC information in 2-D, Gaussian [37] is employed as the spreading function to extend log-Gabor filters from 1-D to 2-D. And then, the 2-D log-Gabor function is given the following transfer function:

$$G_2\left(\omega, \theta_j\right) = \exp\left(-\frac{\left(\log\left(\frac{\omega}{\omega_0}\right)\right)^2}{2\sigma_r^2}\right) \cdot \exp\left(-\frac{\left(\theta - \theta_j\right)^2}{2\sigma_\theta^2}\right) \quad (1)$$

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

$M^{\mathrm{o}}_{n,\theta_j}$ and $M^{\mathrm{e}}_{n,\theta_j}$ show the odd- and even-symmetric filters of the 2-D log-Gabor filters on scale n and orientation θ_j. By regulating these parameters in Eq. (1) and convolving the input image with G_2, a set of the responses are generated at each point x as $[e_{n,\theta_j}(x), o_{n,\theta_j}(x)]$, where $o_{n,\theta_j}(x)$ and $e_{n,\theta_j}(x)$ denote responses of the even- and odd-symmetric filters. Then, the local amplitude on scale n and orientation θ_j is computed as

$$A_{n,\theta_j}(x) = \sqrt{e_{n,\theta_j}(x)^2 + o_{n,\theta_j}(x)^2} \quad (2)$$

the local energy with θ_j is defined as

$$E_{\theta_j}(x) = \sqrt{F_{\theta_j}(x)^2 + H_{\theta_j}(x)^2} \quad (3)$$

where $F_{\theta_j}(x) = \sum ne_n, \theta_j(x)$ and $H_{\theta_j}(x) = \sum ne_n, \theta_j(x)$. Finally, the 2-D PC at x is calculated as:

$$PC(x) = \frac{\sum_j E_{\theta_j}(x)}{\lambda + \sum_n \sum_j A_{n,\theta_j}(x)} \quad (4)$$

where λ is a positive constant to maintain the fraction stability. Take the Fig. 3(f)-(j) as an example of PC information; Fig. 3(f)-(j) are PC maps extracted from original images in Fig. 3(a)-(e), respectively. As can be seen, the PC map is able to reflect the main contour information of the image.

B. Gradient Magnitude Extraction From the Spatial Domain

Edge is able to convey the important visual information. The gradient is widely used in the field of image quality assessment [13], [19], [20], [23] since it has the ability to capture structure and contrast changes. And the gradient magnitude map can be represented by horizontal and vertical components, which are calculated by convolving the input image with an edge detector. The famous Prewitt filter [39] is utilized in this paper because it possesses the lowest computational complexity. The horizontal and vertical components of GM map are defined as:

$$G_x(x) = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} * f(x)$$
$$G_y(x) = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * f(x) \quad (5)$$

where $G_x(x)$ and $G_y(x)$ denote the horizontal and vertical GM maps, respectively, and "$*$" is the convolving filter. The $f(x)$ denotes an input image. And then, the GM of an input image is computed as $G(x) = \sqrt{G_x^2(x) + G_y^2(x)}$. As exhibited in Fig. 3 (k)-(o), the GM maps are extracted from Fig. 3(a)-(e). The structural information in the original image can be presented by the GM map.

Actually, HVS tends to pay more attention to the strong edges in the image. As depicted in Section II, PC and GM maps have their own merits in different image regions. To obtain strong edges (i.e., localized structure map) in different image regions, we combine the normalized GM values and PC information to emphasize the edges. So, the localized structure

map (LS) is defined as:

$$LS(i,j) = \max\{GM(i,j)/GM_{\max}, PC(i,j)\} \tag{6}$$

where (i, j) denotes each point in the LS, GM and PC maps.

That is to say, the LS map is computed by Eq. (6), which highlights differences with details in an image. Since the PC values are distributed from 0 to 1, GM values are normalized to 0-1 by dividing the maximum value in the GM map. The LS map can be seen from Fig. 1(d). In each position, the larger value in the GM and PC maps is taken as the structural feature points. Therefore, the LS map is able to capture structural information in the image thoroughly. Furthermore, the local structure dissimilarity between f_1 and f_2 is defined as correlation coefficient (CC_{LS}):

$$CC_{LS}(r,d) = \frac{\sum_{i=1}^{N}(r_i - \bar{r}) \cdot \left(d - \bar{d}\right)}{\sqrt{\sum_{i=1}^{N}(r_i - \bar{r})^2 \cdot \sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2}} \tag{7}$$

where r_i and d_i are pixels in LS maps of f_1 and f_2, respectively. N denotes the number of pixels in an image. It is noteworthy that a higher CC value denotes higher similarity or better image quality and vice versa.

C. Texture Feature Extraction in Frequency Domain and Similarity Computation

Because of the simplicity of operation, PC and GM have been widely employed as edge extraction tools to detect the local structure in an image. Unfortunately, the complex HVS cannot be simulated only with PC and GM. Satisfactory structures cannot be obtained from PC and GM maps alone. A great deal of information that needs to be represented by complex features still remains underused. It should be observed that features like texture descriptor express information correlating with HVS. Since the texture also exists in structural information, this paper performs log-Gabor filter for extracting texture features in the frequency domain.

The expression of the log-Gabor filter has been given by Eq. (1); it is not difficult to see that two parameters inform the log-Gabor filter, i.e., scale and orientation. This implies that the log-Gabor filter has the attributes of orientation and multiscale, which is consistent with HVS. Therefore, we select the log-Gabor filter to extract texture from an input signal. The log-Gabor filter bank consists of four directions, which are 0°, 45°, 90°, 135°, and four scales in this paper. Specifically, the window for the $(s + 1)$-th scale is twice as large as the s-th scale. After filtering, an $a \times b$ image can be described by an $a \times b$ complex matrix. Then, the energy map is given by

$$E_{a,b}(x,y) = \sqrt{\left(D^r_{a,b}(x,y)\right)^2 + \left(D^i_{a,b}(x,y)\right)^2} \tag{8}$$

where $D^r_{a,b}(x, Y)$ and $D_{a,b}(x, Y)$ are real and imaginary parts of the complex matrix, respectively. The texture information is represented by the energy maps that are generated from the log-Gabor filter bank with four scales and four orientations. Fig. 4 shows an original image and its energy maps. According to Fig. 4, the noises are reduced, while the image loses more detailed information with the larger scale.

Benefiting from the energy map, the $a \times b$ complex matrix can be transformed into a real matrix of the same size. The texture similarity is represented by the chi-square distance of the two real matrices obtained from the reference image and corresponding distorted one. The filters are divided into s (s = 1, 2, 3, 4) groups according to the number of scales. In addition, there are four directions for each scale. Hence, the texture similarity at the s-th scale is calculated as the chi-square distance between two $a \times 4b$ real matrices,

$$\chi^2_s(E_R, E_D) = \frac{1}{4ab}\sum_{i=1}^{4ab}\frac{(E_R(i) - E_D(i))^2}{E_R(i) + E_D(i)} \tag{9}$$

where E_R and E_D denote energy maps at s-th scale of f_1 and f_2, respectively. The indicator implies that the image with more severe distortion has less similarity to its reference image. After filtering, four similarity indicators can be obtained from each image in this paper.

D. Spatial Frequency Feature Extraction and Dissimilarity Computation

In fact, except structural information, contrast sensitivity function (CSF) plays a vitally important role in HVS which has different sensitivities to distortions depending on spatial frequency [24]. Therefore, we decide to employ the subsequent three characteristics in the luminance channel to simulate the CSF of HVS. These three similarity indicators are expressed by comparing the contrast energy values in low frequency (LF), middle frequency (MF) and high frequency (HF) districts in 4×4 discrete cosine transform (DCT) blocks. The contrast energy map in the LF region extracted from an image is defined as [24]:

$$\phi_L = \sum_{(u,v)\in R_L} p(u,v) \tag{10}$$

where $p(u, v)$ denotes the normalized magnitude of a DCT coefficient at (u, v). R_L is the LF region. Similarly, contrast energy maps in MF and HF regions are calculated as:

$$\phi_M = \sum_{(u,v)\in R_M} p(u,v) \tag{11}$$

$$\phi_H = \sum_{(u,v)\in R_H} p(u,v) \tag{12}$$

Similar to SSIM, the similarity indicator has the form of $(2a \cdot b + c)/(a^2 + b^2 + c)$. Therefore, the similarity indicator in the LF region between the reference image and its distorted one is defined as:

$$S_L = \frac{1}{N}\sum_{(u,v)\in R_L}\frac{2\phi_{RL}\cdot\phi_{DL} + C_1}{\phi^2_{RL} + \phi^2_{DL} + C_1} \tag{13}$$

where ϕ_{RL} and ϕ_{DL} represent the contrast energy maps of f_1 and f_2 in the LF region, respectively. N denotes the number of pixels in an image. C_1 is a positive constant to maintain the fraction stability. Similarly, similarity indicators in MF and HF regions between f_1 and f_2 are defined as:

$$S_M = \frac{1}{N}\sum_{(u,v)\in R_M}\frac{2\phi_{RM}\cdot\phi_{DM} + C_2}{\phi^2_{RM} + \phi^2_{DM} + C_2} \tag{14}$$

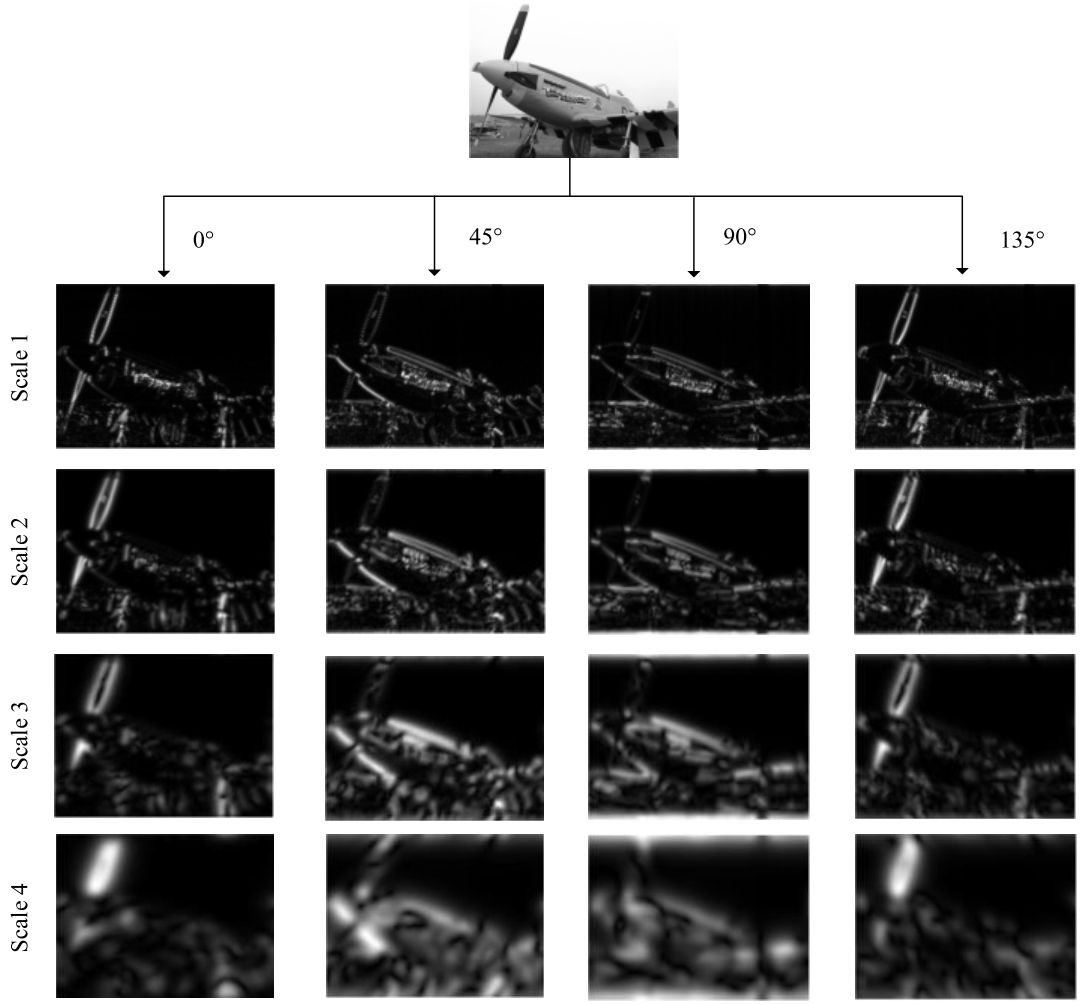

Fig. 4. Texture information of an original image is obtained from a log-Gabor filter bank with four scales and four orientations.

$$S_H = \frac{1}{N} \sum_{(u,v)\in R_H} \frac{2\phi_{RH} \cdot \phi_{DH} + C_3}{\phi_{RH}^2 + \phi_{DH}^2 + C_3} \tag{15}$$

where ϕ_{RM} and ϕ_{DM} denote the contrast energy maps of f_1 and f_2 in the MF region, respectively. ϕ_{RH} and ϕ_{DM} represent the contrast energy maps of f_1 and f_2 in the HF region, respectively. N denotes the number of pixels in the image. C_2 and C_3 are two positive invariables to control numerical stability.

E. Chromatic Feature Extraction and Similarity Measure

It is common knowledge that color information can also affect the visual quality in color images. Most of the existing IQA algorithms just measure the image quality in the luminance channel, which overlooks the chrominance information. Color information is not considered in the above features. Consequently, we introduce a similarity indicator to reflect the color distortions. For this purpose, an input RGB color image is converted into the YIQ color space [35]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{16}$$

where Y denotes the luminance component, and Y and Q represent chromatic components. In this paper, Y is used to compute LS, texture and spatial frequency features. I and Q are employed to capture color distortions between the f_1 and f_2 as follows:

$$S_I = \frac{1}{N} \sum_x \frac{2I_R(x) \cdot I_D(x) + C_4}{I_R^2(x) + I_D^2(x) + C_4} \tag{17}$$

$$S_Q = \frac{1}{N} \sum_x \frac{2Q_R(x) \cdot Q_D(x) + C_4}{Q_R^2(x) + Q_D^2(x) + C_4} \tag{18}$$

$$S_C = \frac{1}{N} \sum_x \left(\frac{2I_R(x) \cdot I_D(x) + C_4}{I_R^2(x) + I_D^2(x) + C_4} \cdot \frac{2Q_R(x) \cdot Q_D(x) + C_4}{Q_R^2(x) + Q_D^2(x) + C_4} \right) \tag{19}$$

where x denotes the pixel of the image, I_R, Q_R, I_D and Q_D represent the chromatic components of the f_1 and f_2, respectively. N denotes the number of pixels in an image. C_4 is a positive constant, which is of similar function to C_1, C_2, and C_3.

F. Regression Tool

Machine learning has been extensively applied in the domain of image quality assessment, such as random forest (RF) [32], [40] and support vector regression (SVR) [29]–[31]. Regression tools are contained in these

TABLE I
INFORMATION ABOUT FIVE PUBLIC DATABASES

Database	Reference Images NO.	Distorted Images NO.	Distortion Types NO.	Distortion Levels No.	Subjects No.
TID2013.	25	3000	24	5	971
TID2008	25	1700	17	4	838
CSIQ	30	866	6	4-5	35
LIVE	29	779	5	5	161
CCID2014	15	655	5	<=5	22

learning-based techniques. We adopt RF to learn a mapping from the feature vectors to quality score because the RF regression distinctly outperforms SVR. The performance comparisons of random forest and support vector regression are in Section IV.

After features extraction, a 9-D feature vector can be generated by the proposed method. In this paper, the feature vector is represented by

$$f = \left[CC_{LS}, \chi_1^2, \chi_2^2, \chi_3^2, \chi_4^2, S_L, S_M, S_H, S_C\right] \tag{20}$$

where the nine indicators have already been introduced in Sections III-A–III-E, which contain local structure similarity (CC_{LS}), the texture dissimilarity at the four scale ($\chi_1^2, \chi_2^2, \chi_3^2, \chi_4^2$), the similarity between the contrast energy in low frequency region (S_L), the similarity between the contrast energy maps in middle frequency region (S_M), the similarity between the contrast energy maps in high frequency region (S_H) and the chromaticity similarity (S_C). A set of 9-D feature vectors and MOS are employed to set up a regression model in the training stage. In the test stage, the feature vectors extracted from the testing dataset are directly mapped to the quality scores through the learned model.

IV. Experimental Results and Analysis

A. Databases and Methods for Comparison

In this paper, the performance of the PFSI is tested on five databases, including TID2013 [41], TID2008 [42], CSIQ [43], LIVE [44] and CCID2014 [45]. Specifically, the contrast distorted images of the CCID2014 database are also utilized in the experiments. The information of these databases is summarized in Table I.

To quantitatively describe the performance of the proposed method, this paper employs four popular performance metrics, i.e., Pearson linear correlation coefficient (PLCC), Spearman rank-order coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC) and root mean squared error (RMSE). PLCC is used to represent the prediction accuracy of a model, i.e., the ability to predict subjective evaluations with low error. RMSE is used to judge the prediction consistency of a model. SROCC and KROCC are used to represent the prediction monotonicity of a model, i.e., the degree to which subjective evaluations can be predicted. Furthermore, we adopt the regression analysis method recommended by the video quality experts group [46], which leads the objective scores on the same scale as the subjective scores and provides a better fit for all data. The logistic regression function is defined as [46]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}}\right) + \beta_4 x + \beta_5 \tag{21}$$

where x denotes the raw predicted score, and $f(x)$ is mapped out. β_i $(i = 1, 2, 3, 4, 5)$ are the free parameters to be fitted by minimizing the sum of squared differences $f(x)$ and MOS. The reference values of initial parameters are selected based on the recommendation in [46]. The first metric is the PLCC between the predicted scores and subjective MOS after regression. Similar to PLCC, the fourth metric is the RMSE between the predicted scores and subjective MOS after regression.

Because the viewing distance is closely related to the visual quality [47], like most of the previous FR-IQA models [13], [21]–[24], the process of down-sampling helps to reduce computational complexity. Hence, images are down-sampled by a factor of 2 using a local 2×2 average filter in this paper.

This paper compares 12 state-of-the-art FR-IQA models, i.e., SSIM [13], IW-SSIM [15], FSIM$_C$ [21], GSM [20], GMSD [23], IFS [48], VSI [22], DSCSI [49], PSIM [27], DOG-SSIM$_C$ [32], MDSI [28] and SCQI [24] with the proposed method in terms of prediction accuracy.

B. Cross Validation

Generally, some learning-based models [9], [29] randomly divided images into two subsets in each database; i.e., 80% are used for training and the remaining 20% are used for testing. However, the testing data of these methods may contain the same image content in the training data. Since the impact of the image content in quality assessment is significant, it may affect the accuracy of the prediction. Moreover, it is also not advisable to use the identical dataset for training and testing. This kind of model would lead to a perfect score but fail to predict anything useful on unseen data.

The k-fold cross validation technique [31]–[33] is widely applied to the validation of machine learning-based IQA models. This technique can effectively avoid the occurrence of over-fitting and under-fitting. Therefore, we also adopt k-fold cross validation to test the robustness of the PFSI. The whole dataset is divided into k disjoint subsets, and the number of samples in each subset is equal or roughly equal. Then, each subset is treated as testing data separately, and the remaining (k-1) subsets are treated as training data. The final result depends on the average performance of the k test experiments. To obtain a more stable result, the process of training-test is repeated 1000 times, and the median is retained.

As mentioned above, the impact of image content in quality assessment is significant. To this end, the training and testing subsets are separated by the image content in each database. Specifically, k is usually set as 5∼10 [31]. The TID2013 database is divided into eight subsets according to the different image contents. Similarly, the TID2008 database is divided into eight subsets, CSIQ is divided into ten subsets, and LIVE is divided into ten subsets.

TABLE II
PERFORMANCE COMPARISON OF 13 IQA MODELS ON FIVE DATABASES

		SSIM	IW-SSIM	$FSIM_c$	GSM	GMSD	IFS	VSI	DSCSI	PSIM	MDSI	SCQI	DOG-$SSIM_c$	PFSI	PFSI-R
TID 2013	PLCC	0.7895	0.8319	0.8769	0.8464	0.8590	0.8791	0.9000	0.8782	0.9080	0.9085	0.9071	0.9279	0.9426	**0.9449**
	SROCC	0.7417	0.7779	0.8510	0.7946	0.8044	0.8697	0.8965	0.8744	0.8925	0.8899	0.9052	0.9145	0.9315	**0.9356**
	KROCC	0.5588	0.5977	0.6665	0.6255	0.6339	0.6785	0.7183	0.6862	0.7161	0.7123	0.7327	0.7542	0.7777	**0.7815**
	RMSE	0.7608	0.6880	0.5959	0.6603	0.6346	0.5905	0.5404	0.5930	0.5193	0.5181	0.5219	0.4606	0.4095	**0.4048**
TID 2008	PLCC	0.7732	0.8579	0.8762	0.8422	0.8788	0.8810	0.8762	0.8445	0.9078	0.9160	0.8899	0.9288	**0.9429**	0.9411
	SROCC	0.7749	0.8559	0.8840	0.8504	0.8907	0.8903	0.8979	0.8634	0.9120	0.9208	0.9051	0.9235	**0.9382**	0.9366
	KROCC	0.5768	0.6636	0.6991	0.6596	0.7092	0.7009	0.7123	0.6651	0.7395	0.7515	0.7294	0.7721	**0.7921**	0.7863
	RMSE	0.8511	0.6895	0.6468	0.7235	0.6404	0.6349	0.6466	0.7187	0.5628	0.5383	0.6120	0.5036	**0.4422**	0.4536
CSIQ	PLCC	0.8613	0.9144	0.9192	0.8964	0.9541	0.9576	0.9279	0.9313	0.9642	0.9531	0.9268	0.9445	**0.9783**	0.9758
	SROCC	0.8756	0.9213	0.9310	0.9108	0.9570	0.9581	0.9423	0.9417	0.9621	0.9569	0.9434	0.9488	**0.9732**	0.9700
	KROCC	0.6907	0.7529	0.7690	0.7374	0.8129	0.8158	0.7857	0.7787	0.8273	0.8130	0.7870	0.8128	**0.8625**	0.8507
	RMSE	0.1334	0.1063	0.1034	0.1164	0.0786	0.0757	0.0979	0.0956	0.0696	0.0795	0.0986	0.0857	**0.0523**	0.0574
LIVE	PLCC	0.9449	0.9522	0.9613	0.9512	0.9603	0.9586	0.9482	0.9434	0.9584	0.9659	0.9344	0.9747	**0.9780**	0.9771
	SROCC	0.9479	0.9567	0.9645	0.9561	0.9603	0.9599	0.9524	0.9487	0.9622	0.9667	0.9406	0.9710	**0.9749**	0.9743
	KROCC	0.7963	0.8175	0.8363	0.8150	0.8268	0.8254	0.8058	0.7982	0.8294	0.8395	0.7835	0.8607	**0.8709**	0.8657
	RMSE	8.9455	8.3472	7.5296	0.8427	7.6214	7.7764	8.6817	9.0635	7.7996	7.0790	9.7355	6.1241	**5.6082**	5.7907
CCID 2014	PLCC	0.8308	0.8342	0.8204	0.8073	0.8521	0.7936	0.8209	0.7586	0.8386	0.8576	0.8200	0.8612	**0.8771**	0.8721
	SROCC	0.8174	0.7811	0.7657	0.7768	0.8077	0.7227	0.7734	0.7400	0.8003	0.8128	0.7811	0.8270	0.8418	**0.8427**
	KROCC	0.6106	0.5898	0.5707	0.5711	0.6100	0.5337	0.5735	0.5468	0.6038	0.6181	0.5812	0.6398	**0.6596**	0.6556
	RMSE	0.3640	0.3606	0.3739	0.3859	0.3422	0.3978	0.3734	0.4260	0.3562	0.3363	0.3734	0.3318	**0.3140**	0.3182
W. A	PLCC	0.8156	0.8617	0.8861	0.8596	0.8862	0.8901	0.8956	0.8726	0.9140	0.9175	0.9002	0.9291	**0.9449**	0.9446
	SROCC	0.7964	0.8348	0.8736	0.8388	0.8619	0.8819	0.8972	0.8757	0.9050	0.9070	0.9022	0.9190	0.9347	**0.9357**
	KROCC	0.6108	0.6566	0.6970	0.6636	0.6936	0.7037	0.7214	0.6919	0.7376	0.7396	0.7301	0.7669	**0.7910**	0.7888

TABLE III
PERFORMANCE COMPARISON OF 5 LEARNING-BASED IQA MODELS ON FOUR DATABASES

		LLM	[29]	CD-MMF	DOG-SSIMc	PFSI
TID 2013	PLCC	0.9068	0.8939	0.9249	0.9279	**0.9426**
	SROCC	0.9037	0.8802	0.9143	0.9145	**0.9315**
	KROCC	0.7209	0.6953	0.7611	0.7542	**0.7777**
	RMSE	0.5277	0.5527	0.4714	0.4606	**0.4095**
TID 2008	PLCC	0.8971	0.8881	**0.9476**	0.9288	0.9429
	SROCC	0.9077	0.8880	**0.9422**	0.9235	0.9382
	KROCC	0.7368	0.7025	**0.8864**	0.7721	0.7921
	RMSE	0.5982	0.6121	**0.4289**	0.5036	0.4422
CSIQ	PLCC	0.9000	0.9680	0.9675	0.9445	**0.9783**
	SROCC	0.9050	0.9627	0.9668	0.9488	**0.9732**
	KROCC	0.7238	0.8358	0.8266	0.8128	**0.8625**
	RMSE	0.1232	0.0650	0.0664	0.0857	**0.0523**
LIVE	PLCC	0.9578	**0.9829**	0.9801	0.9747	0.9780
	SROCC	0.9608	0.9755	**0.9798**	0.9710	0.9749
	KROCC	0.8230	**0.8778**	0.8574	0.8607	0.8709
	RMSE	7.7678	**4.1701**	5.4239	6.1241	5.6082

Information about the subsets classification can be found in [33].

C. Overall Performance Evaluation

Table II shows the prediction performance of the PFSI and the other 12 state-of-the-art IQA models on five public image databases in terms of PLCC, SROCC, KROCC and RMSE. Note that the best result is highlighted in boldface for each database. To be fair, the training-test process of the DOG-SSIMc [32] is the same as the proposed method.

Because image classification is critical to regression, we adopt two schemes to prove the effectiveness of the proposed method. One way is to classify images by content and perform *k*-fold cross validation as described in Section IV-B, and the other is to randomly divide images into two subsets in each database (80% for training, and the remaining 20% are used for testing). The performance of the former is represented by PFSI and that of the latter by PFSI-R. As can be seen from Table II, the predictive capability of the PFSI/PFSI-R outperforms all the other methods by a large margin on the five databases, i.e., TID2013, TID2008, CSIQ, LIVE and CCID2014. Meanwhile, PFSI/PFSI-R almost attains the best results every time among all methods and consistent performance across the five databases. This means that PFSI/PFSI-R is an effective model to reflect the information processing of HVS for image quality perception. In comparison, there is no other algorithm that can obtain the best results on each database. Moreover, the PFSI/PFSI-R can effectively predict the images with contrast distortion in the CCID2014 database. If the database becomes larger, the images in the testing data that contain the same content of the training data will increase through random classification. For this reason, PFSI-R surpasses PFSI in TID2013 database. The

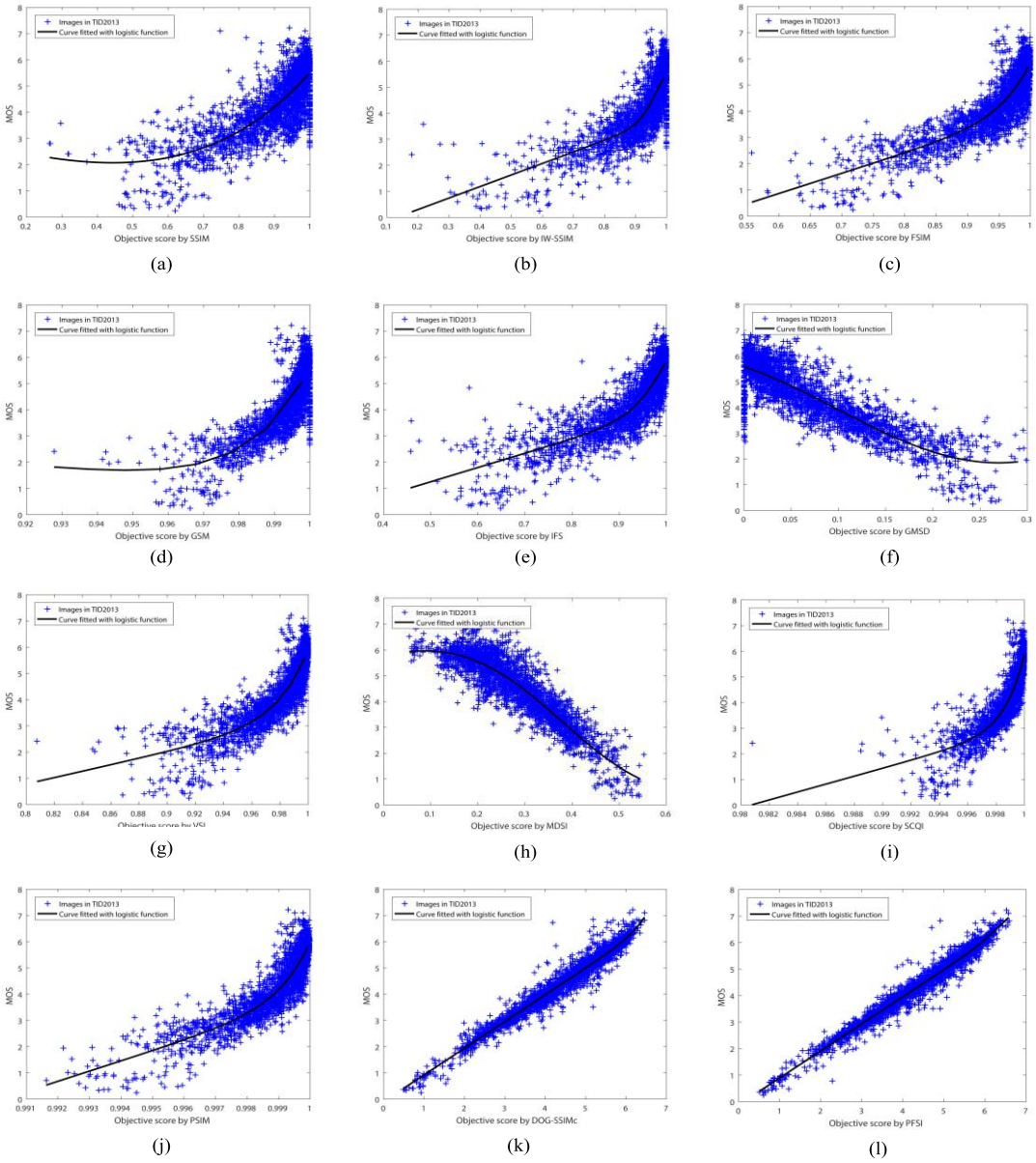

Fig. 5. Scatter plots of subjective scores versus predicted scores obtained by objective IQA models on the TID2013 database. (a) SSIM, (b) IW-SSIM, (c) FSIM, (d) GSM, (e) IFS, (f) GMSD, (g) VSI, (h) MDSI (i) SCQI, (j) PSIM, (k) DOG-SSIMc and (l) PFSI.

overall performance is calculated by taking weighted averages (W. A.) that depend on the numbers of distorted images. For all the IQA models, their weighted-average PLCC, SROCC and KROCC performances are given in Table II. Note that PFSI/PFSI-R, which is largely superior to other models, consistently achieves the best results on the average performance comparison.

To further demonstrate the effectiveness of the proposed method,the performance of PFSI indices will be compared with four representative learning-based methods. The best result is highlighted in boldface for each database. As tabulated in Table III, PSFI outperforms all learning-based methods on the TID2013, and CSIQ databases. On the TID2008 and LIVE databases, the performance of the proposed method is slightly worse than the best method. Note that no IQA model performs very well on four databases except for proposed method PFSI. For other learning-based models, they may work well on some databases, but fail to provide good results on other databases. For example, although LLM and [29] can get good results on LIVE database, they perform poorly on TID2013 and TID2008 databases.

In addition, Fig. 5 shows scatter plots of subjective MOS values versus predicted quality scores for PFSI and twelve representative IQA methods (SSIM, IW-SSIM, FSIM, GSM, IFS, GMSD, VSI, MDSI, SCQI, PSIM and DOG-SSIMc) along with the best fitting logistic functions on the TID2013 database. The horizontal and vertical axis corresponds to the object quality score and subjective MOS; each point on the plot denotes one image in the database. The curves (meaning perfect prediction) are obtained by a nonlinear fitting according to Eq. (21). As can be seen from Fig. 5, these sample points of PFSI tend to be clustered closer to the fitting curve

TABLE IV
PERFORMANCE COMPARISON AMONG DIFFERENT REGRESSION TOOLS

		SVR	RF
TID 2013	PLCC	0.9180	**0.9426**
	SROCC	0.9083	**0.9315**
	KROCC	0.7426	**0.7777**
	RMSE	0.4907	**0.4095**
TID 2008	PLCC	0.9159	**0.9429**
	SROCC	0.9118	**0.9382**
	KROCC	0.7497	**0.7921**
	RMSE	0.5333	**0.4422**
CSIQ	PLCC	0.9645	**0.9783**
	SROCC	0.9655	**0.9732**
	KROCC	0.8422	**0.8625**
	RMSE	0.0672	**0.0523**
LIVE	PLCC	0.9696	**0.9780**
	SROCC	0.9685	**0.9749**
	KROCC	0.8511	**0.8790**
	RMSE	6.5926	**5.6082**
CCID 2014	PLCC	0.8477	**0.8771**
	SROCC	0.8229	**0.8418**
	KROCC	0.6335	**0.6596**
	RMSE	0.3482	**0.3140**

compared to other scatter plots, and this implies a higher consistency with the subjective scores. Specifically, PFSI is able to acquire better performance for low-quality images. When RF and SVR are used as regression tools, the predicted results are listed in Table IV. Obviously, the predictive results using RF as a regression tool on each database are better than those using SVR.

In order to analyze the statistical significance of proposed method PFSI in relative to the competing approaches, we employ the left-tailed F-test to decide whether a metric is statistically superior to another index [23]. The F-test is based on the residuals between the objective quality scores by an IQA model after nonlinear regression and the subjective scores. The results of significance tests are shown in Fig. 6. A value of $H = 1$ for the left-tailed F-test at a significance level of 0.05 represents that the first method (indicated by the row in Fig. 6) is superior in IQA performance to the second method (indicated by the column in Fig. 6) with a confidence greater than 95%. A value $H = 0$ represents that first method is not significantly better than the second one, i.e., the two methods have no significant difference in performance. Fig. 6(a-d) shows the statistical significant results on TID2013, TID2008, CSIQ and LIVE databases, respectively. It can be observed that the performances of statistical significance tests are consistent with the results displayed in Table II and Table III. The proposed method PFSI is significantly better than all the other methods on TID2013 and CSIQ databases. On TID2008 and LIVE database, PFSI, CD-MMF and DOG-SSIMc all perform very well and they have no significant difference. Note that no IQA metric performs significantly better than proposed model PFSI on four databases. In other words, the overall statistical performance of the proposed method PFSI is the best.

TABLE V
PLCC VALUES OF CROSS DATABASE VALIDATION WITH DIFFERENT METHODS

raining database	LIVE	TID2008	CSIQ	CSIQ
Testing database	CSIQ	CSIQ	LIVE	TID2008
[29]	0.8620	0.8863	0.8852	0.7761
DOG-SSIMc	0.8924	0.9154	0.8934	0.8554
Proposed	**0.9376**	**0.9471**	**0.9243**	**0.8902**

TABLE VI
AVERAGE ERUNNING TIME FOR DIFFERENT MODELS ON TID2013

Methods	Running times (s)
SSIM	0.0140
IW-SSIM	0.2412
FSIM	0.0983
GSM	0.0129
IFS	0.0382
GMSD	0.0083
VSI	0.0837
MDSI	0.0196
SCQI	0.0528
PSIM	0.0333
DOG-SSIMc	0.0427
PFSI	0.1348

D. Cross Database Validation

To comprehensively analyze the robustness and generality of the PFSI, cross database validation is conducted. We adopt a database to train the regression model and the testing process on another database. Ideal cross validation implies that the databases of training and testing are absolutely non-overlapping. Because the same images exist in TID2013, TID2008, and LIVE, this paper employs CSIQ as the training or testing database and the TID2008 and LIVE databases act another role in turn. Table V records the experimental results in terms of PLCC, the best result produced by a train-test is highlighted in boldface. In fact, similar results can also be obtained by using SROCC, KROCC and RMSE. As can be seen in Table V, for all training models on any database, the PFSI far surpasses the other state-of-the-art methods. For instance, when the LIVE and TID2008 databases are used to train the regression model, the test results obtained by the CSIQ database are 0.9376 and 0.9471, respectively. On the other hand, the CSIQ database is employed as a training target; the test results are 0.9243 and 0.8902 on the LIVE and TID2008 databases, respectively.

E. Computational Complexity

Table VI lists the average running time of each image on the TID2013 database, which contains 3000 images of size

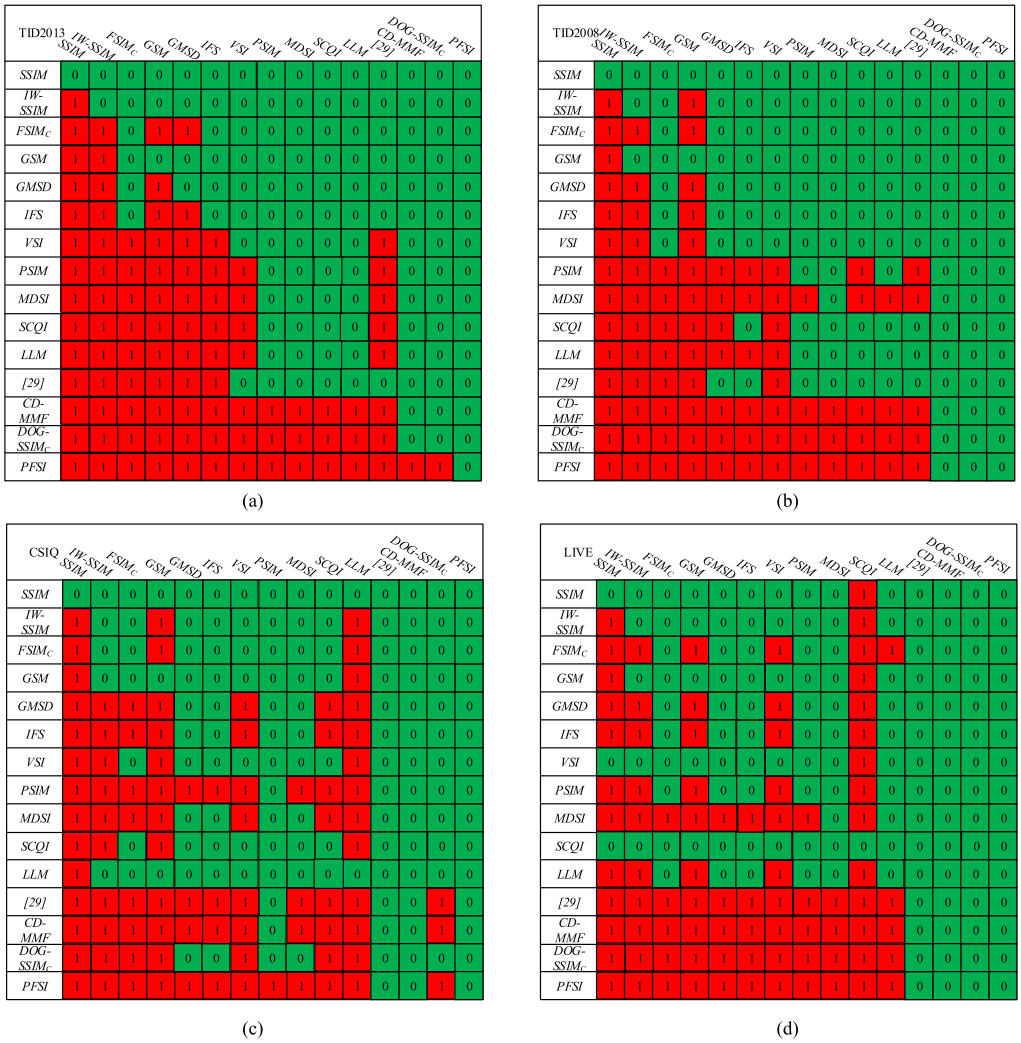

Fig. 6. The results of statistical significance tests of the competing IQA methods on the databases of (a) TID2013, (b) TID2008, (c) CSIQ and (d) LIVE. A value of '1' (highlighted in red) means that the method in the row is significantly better than the method in the column, while a value of '0' (highlighted in green) means the first method is not significantly better than the second method.

512×384. All approaches are executed on a 3.60 GHz Intel Core i7-4790 PC with 8GB RAM. Meanwhile, all codes were received from the original authors and were implemented in MATLAB 2016a. It is clear that GMSD is the fastest, followed by SSIM. However, their performances are quite poor. The running speed of the proposed method is similar to FSIM. Although the average running time of PFSI is relatively high, it maintains the highest prediction accuracy than the other IQA models.

F. Parameters Setting

There are some parameters that need to be set in the proposed method PFSI. C_1, C_2, C_3 and C_4 are for stability when the denominator is closer to zero. In our experiments, we find that the prediction accuracy of PFSI is relatively insensitive to the choice of C_i $(i = 1, 2, 3, 4)$ for the TID2013 database. In fact, we can reach the same conclusion on other databases. Fig. 7 plots the SROCC curves against C_i $(i = 1, 2, 3, 4)$ by applying PFSI to the TID2013 database. As can be seen from Fig. 7(a), SROCC changes slightly when $C_1 = 0.6, 6, 60$ and 600. Fig. 7(b)-(d) also shows similar results. For that, we set $C_1 = 0.6$, $C_2 = 2000$, $C_3 = 1.7$ and $C_4 = 200$. Moreover, for the RF regression

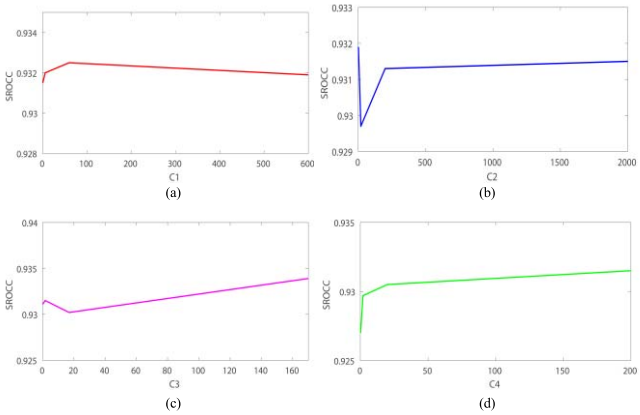

Fig. 7. The performance of PFSI in terms of SROCC vs. constant $C_i (i = 1, 2, 3, 4)$ on TID2013 database.

TABLE VII
THE INFORMATION ABOUT PARAMETERS SETTING ON EACH DATABASE

Parameters	TID2013	TID2008	CSIQ	LIVE	CCID2014
(ntree, mtry)	(500,3)	(500,3)	(500,3)	(500,3)	(500,3)
(C, γ)	(24.6066, 6.6165)	(2.5432, 3.2589)	(100, 5.7719)	(100, 5.7719)	(100, 5.7719)

tool, we adopt a default setting for the number of trees (*ntree*) and variables per level (*mtry*); i.e., these two parameters are taken as $(ntree, mtry) = (500, 3)$. The radial basis function (RBF) is chosen for support vector regression; the penalty parameter C and kernel parameter are set to be $(C, \gamma) = (31.1304, 29.0028)$ on TID2013, $(C, \gamma) = (24.6066, 6.6165)$ on TID2008, $(C, \gamma) = (2.5432, 3.2589)$ on CSIQ, $(C, \gamma) = (100, 5.7719)$ on LIVE and $(C, \gamma) = (18.2299, 0.94845)$ on CCID2014. The information regarding the parameter setting on each database is listed on Table VII.

V. CONCLUSION

In this paper, an effective and reliable image quality assessment is proposed, namely, PFSI, which combines spatial and frequency domain features. The local structure map, which highlighted differences with details, is a combination of the GM and PC maps. Afterward, we take the Log-Gabor filter and spatial frequency as complementary aspects of image quality to reflect texture and CSF in the frequency domain. Finally, the random forest (RF) employs a regression tool to predict overall quality scores by effective mapping of distortion effects. Extensive experimental results on five publicly available databases indicate that the proposed method can highly agree with subjective perceptions and outperforms all the state-of-the-art IQA metrics since we combine existing approaches to overcome the limitations of a single approach. The performance on the cross database validation demonstrates that the proposed approach maintains effectiveness and robustness. The source code of the proposed metric will be publicly available.

REFERENCES

[1] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.

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

[3] W. Lu, T. Xu, Y. Ren, and L. He, "On combining visual perception and color structure based image quality assessment," *Neurocomputing*, vol. 212, pp. 128–134, Nov. 2016.

[4] A. Saha and Q. M. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures," *Signal Process.*, vol. 128, pp. 186–197, Nov. 2016.

[5] Z. Shi, K. Chen, K. Pang, J. Zhang, and Q. Cao, "A perceptual image quality index based on global and double-random window similarity," *Digit. Signal Process.*, vol. 60, pp. 277–286, Jan. 2017.

[6] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.

[7] M. Liu, K. Gu, G. Zhai, P. L. Callet, and W. Zhang, "Perceptual reduced-reference visual quality assessment for contrast alteration," *IEEE Trans. Broadcast.*, vol. 63, no. 1, pp. 71–81, Mar. 2017.

[8] L. Li *et al.*, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

[9] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.

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

[11] Y. Liu *et al.*, "Quality assessment for real out-of-focus blurred images," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 70–80, Jul. 2017.

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

[13] W. Zhou, 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.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals Syst. Comput.*, vol. 2. Pacific Grove, CA, USA, 2003, pp. 1398–1402.

[15] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[16] H. R. Sheikh, A. C. Bovik, and G. D. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

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

[18] 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, pp. 1–21, Jan. 2010.

[19] K. Gu *et al.*, "Evaluating quality of screen content images via structural variation analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 10, pp. 2689–2701, Oct. 2018.

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

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

[22] 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.

[23] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[24] S.-H. Bae and M. Kim, "A novel image quality assessment with globally and locally consilient visual quality perception," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2392–2406, May 2016.

[25] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[26] K. Gu *et al.*, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.

[27] 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.

[28] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.

[29] Y. Ding, Y. Zhao, and X. Zhao, "Image quality assessment based on multi-feature extraction and synthesis with support vector regression," *Signal Process. Image Commun.*, vol. 54, pp. 81–92, May 2017.

[30] S. Du, Y. Yan, and Y. Ma, "Blind image quality assessment with the histogram sequences of high-order local derivative patterns," *Digit. Signal Process.*, vol. 55, pp. 1–12, Aug. 2016.

[31] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 347–364, Apr. 2012.

[32] S.-C. Pei and L.-H. Chen, "Image quality assessment using human visual DOG model fused with random forest," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3282–3292, Nov. 2015.

[33] S. Wang, C. Deng, W. Lin, G.-B. Huang, and B. Zhao, "NMF-based image quality assessment using extreme learning machine," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 232–243, Jan. 2017.
[34] S. Bosse, D. Maniry, K. R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[35] C. Yang and S. H. Kwok, "Efficient gamut clipping for color image processing using LHS and YIQ," *Opt. Eng.*, vol. 42, no. 3, pp. 701–711, Mar. 2003.
[36] L. Henriksson, A. Hyvärinen, and S. Vanni, "Representation of cross-frequency spatial phase relationships in human visual cortex," *J. Neurosci. Official J. Soc. Neurosci.*, vol. 29, no. 45, pp. 14342–14351, 2009.
[37] P. D. Kovesi, "Image features from phase congruency," *J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.
[38] A. Saha and Q. M. J. Wu, "Perceptual image quality assessment using phase deviation sensitive energy features," *Signal Process.*, vol. 93, no. 11, pp. 3182–3191, Nov. 2013.
[39] R. K. R. Jain and B. G. Schunck, *Machine Vision*. New York, NY, USA: McGraw-Hill, 1995.
[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[41] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process. (EUVIP)*, Paris, France, 2013, pp. 106–111.
[42] N. Ponomarenko *et al.*, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv Mod. Radioelectron*, vol. 10, no. 4, pp. 30–45, 2009.
[43] E. C. Larson and D. M. Chandler. (2010). *Categorical Image Quality (CSIQ) Database*. [Online]. Available: http://vision.okstate.edu/csiq
[44] Z. W. H. R. Sheikh, L. Cormack, and A. C. Bovik. *Live Image Quality Assessment Database Release 2*. Accessed: Feb. 2014. [Online]. Available: http://live.ece.utexas.edu/research/quality
[45] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.
[46] Video Quality Experts Group. (Mar. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Vedio Quality Assessment*. [Online]. Available: http://www.vqeg.org/
[47] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.
[48] H.-W. Chang, Q.-W. Zhang, Q.-G. Wu, and Y. Gan, "Perceptual image quality assessment by independent feature detector," *Neurocomputing*, vol. 151, pp. 1142–1152, Mar. 2015.
[49] D. Lee and K. N. Plataniotis, "Towards a full-reference quality assessment for color images using directional statistics," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3950–3965, Nov. 2015.
[50] H. Wang *et al.*, "Image quality assessment based on local linear information and distortion-specific compensation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 915–926, Feb. 2017.
[51] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.
[52] 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./Feb. 2018.
[53] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.
[54] K. Gu *et al.*, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

Zhisen Tang received the B.E. degree in printing engineering from the Hunan University of Technology, Zhuzhou, China, in 2016. He is currently pursuing the M.E. degree with the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an, China. His research interests include image quality assessment and machine learning.

Yuanlin Zheng received the B.S. degree in printing engineering from the Zhuzhou Institute of Technology, Zhuzhou, China, in 1999, the M.S. degree in pulp and papermaking engineering from the Xi'an University of Technology, Xi'an, China, in 2002, and the Ph.D. degree in pulp and papermaking engineering from the Tianjin University of Science and Technology, Tianjin, China, in 2007. He is currently a Full Associate Professor of printing engineering with the Xi'an University of Technology. His research interests include color management, evaluation of quality of color image and color science, and pattern recognition.

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. His research interests include image analysis, environmental perception, quality assessment, and deep learning. He was a recipient of the Best Paper Award from 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 VCIP 2016 and ICIP 2017. He is currently an Associate Editor for IEEE ACCESS and is a reviewer for 20 top journals.

Kaiyang Liao received the B.S. degree in computer science from Xidian University, Xi'an, China, in 2004, the M.S. degree in computer science from the University of Science and Technology Liaoning, Anshan, China, and the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, in 2013. He is currently a Full Lecturer with the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an. His research interests include data mining and pattern recognition.

Wei Wang received the B.E. degree from the Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an, China, where he is currently pursuing the M.E. degree. His main research interests include image quality assessment and machine learning.

Miaomiao Yu received the B.E. degree in printing engineering from Qufu Normal University, Rizhao, China, in 2017 and from the Xi'an University of Technology, Xi'an, China, where she is currently pursuing the M.E. degree in printing engineering. Her research interest includes image quality evaluation.