IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 13, 2019, accepted May 19, 2019, date of publication June 10, 2019, date of current version June 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922011

Image Quality Assessment by Considering Multiscale and Multidirectional Visibility Differences in Shearlet Domain

WU DONG[1,2], HONGXIA BIE[1], LIKUN LU[2], AND YELI LI[2]

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunication, Beijing 100876, China
[2]Beijing Key Laboratory of Signal and Information Processing for High-end Printing Equipment, Beijing Institute of Graphic Communication, Beijing 102600, China

Corresponding author: Wu Dong (dongwu@bigc.edu.cn)

This work was supported by the Scientific Innovation Project of the Beijing Education Committee under Grant PXM2017_014223_000055.

ABSTRACT Conventional objective image assessment metrics, such as mean squared error and peak signal-to-noise ratio, which only calculates pixel-based differences between the original and the degraded images, are not in agreement with the human vision. In this paper, we present an improved objective full-reference image quality assessment method, called the multiscale and multidirectional visibility differences (MMVD) predictor. The proposed MMVD metric considers multiscale and multidirectional visibility differences in the domain of the discrete nonseparable shearlet transform, which emulates the multichannel structure of information processing of the human vision system. In the process of constructing the visual just noticeable difference threshold in the shearlet domain, the contrast sensitivity function and the visual masking effect which are important properties of the human visual perception are considered simultaneously to approximate the sensitivities of human visual responses. Both contrast masking and entropy masking are considered to tackle the visual masking issue. All subbands of the shearlet transform are evaluated, and perceptual errors of subbands are pooled together to yield the objective quality index of a distorted image. The extensive validation experiments are conducted on five public image databases, namely, TID2008, TID2013, CSIQ, IVC, and LIVE. The experimental results demonstrate the proposed method is well coherent with human perception and has better performance compared to the several state-of-the-art image quality metrics.

INDEX TERMS Discrete nonseparable shearlet transform, image quality assessment, human visual system, visual masking, visual just noticeable difference threshold.

I. INTRODUCTION

Image Quality Assessment (IQA) is a basic and challenging research field for image processing. The objective of the studies about image quality evaluation is to obtain an objective image quality method, and this method can be in agreement with subjective ratings made by humans. A triumphant objective image quality method can reduce laborious works of human, such as image quality inspection in communication, printing quality inspection, and other image system performance evaluation in manufacturing environment. Moreover, in a lot of image processing applications, such as digital image collection, intensity transformations, smoothing, sharpening, watermarking, reconstruction, display, and printing, we can employ this objective IQA method online to optimize algorithm effect and cut down computational complexity [1]. The Mean Square Error (MSE) and the Peak Signal-to-Noise Ratio (PSNR) are traditional objective IQA metrics, and they are the most widely employed image quality indices. Meanwhile, they are relatively simple and easy to be implemented. However, they only calculate differences of pixels between the reference image and the degraded image, and don't deal with the correlation between pixels and properties of human vision, so they don't correlate well with subjective evaluations and have been widely criticized [2], [3].

A class of image quality assessment metrics is based on the assumption that the Human Visual System (HVS) attemps to abide by an overarching principle when a distorted image is observed. Typical overarching principles consist of the Structural SIMilarity (SSIM) metric [4], features

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

synthesis index, and the Visual Information Fidelity (VIF) measure [5]. The SSIM metric is under the hypothesis that human vision is very sensitive to structural information of an image, and the image quality degradation is quantified by structural distortion of the image. In the SSIM method, structural distortion is differentiated from luminance and contrast distortion. The SSIM method is a landmark in the research process of IQA, and initiates a new important research direction of objective IQA by investigating the structural change of an image. However, the SSIM metric has a drawback that it cannot correctly assess blurred images. In addition, the SSIM metric don't consider visual characteristics of the HVS. In [6], the standard deviation of the pixel-wise gradient magnitude similarity map of an image is calculated as local structural distortion measurement. In [7], to identify the dominant structures of an image, the image gradient is multiplied by the anisotropy measurement and the local directionality measurement which are computed from the structure tensor. In [8], both first- and high-order image structures are extracted and are fused together by Support Vector Regression (SVR). In [9], the microstructural and macrostructural similarities of the image gradient magnitude are computed. In [10], gradient magnitudes are weighted by neighborhood gradient information and contrast sensitivity of human vision. So far, how to precisely define the structure of an image, and how to quantify structural distortion of an image, still remain open issues and are further investigated in a great many literatures on image quality measure.

Because image features are attractive to the HVS, they can be adopted in the research of image quality assessment. In the features synthesis index, multiple features are selected and then are properly combined together to derive objective evaluation scores. Zhang *et al.* [11] presented a Feature SIMilarity (FSIM) method which integrates two features, the phase congruency and the image gradient magnitude. Yang *et al.* [12] proposed the Riesz transform and Visual contrast sensitivity-based feature SIMilarity (RVSIM) metric which achieves better prediction capability than traditional models. In the RVSIM metric, both a log Gabor filter and the Riesz transform are performed to images, and in the Riesz transform domain, three features including amplitude, phase and direction are extracted and are properly combined with another feature, the image gradient magnitude, to derive the similarity index. Ding *et al.* [13] selected three image features, namely, the image gradient, the energy of the log Gabor filter, and histograms of local pattern analysis, and in [14], six basic image features and eleven auxiliary image features are employed. In addition, in [13] and [14], SVR is adopted to map these features into a predictive score. In [15], non-negative matrix factorization is applied to obtain image features about degradation, and Extreme Learning Machine (ELM) is adopted as the features pooling method. Although features synthesis methods improve objective assessment ability to some extent, it is very difficult to find appropriate features which are highly sensitive to human vision. Additionally, the features polling technique is very important and can significantly affect assessment performance. Obviously, if machine learning approaches, including support vector machine, SVR, ELM, deep learning and so on, are employed as polling techniques, computational complexity of the IQA method will be greatly increased.

In the VIF metric [5], the issue of image quality evaluation is tackled as an information retaining issue and image signals are dealt with to extract cognitive information through the HVS channels. The VIF metric employs the wavelet transform, and the IQA is accomplished by measuring information quantity loss of the distorted image relative to the undistorted image. In [16], the Discrete Wavelet Transform (DWT) is performed to decompose the weighted gradient magnitude image, and entropies of DWT subbands are computed and pooled together. Kuo *et al.* [17] used the log Gabor filter to decompose low-frequency components resulted from the Haar wavelet transform of images, and the local mutual information of log Gabor subbands is calculated and combined together. In [18], image information is classified into three types, namely, saliency information, specific information and entanglement information, and statistical features of the three types of image information are extracted to form the quality assessment model. Information measurement methods can achieve better predictive performance, yet they seldom incorporate the effects from characteristics of human perception.

Because the eyes of a person are ultimate receptors and appreciators of distorted images in the majority of applications, modeling the HVS accurately and efficiently becomes a significant research issue, and it is very important and reasonable to incorporate psychophysical characteristics of the HVS into image quality evaluation algorithms of system implementation, optimization, and testing. During the last two decades, researches have already conducted a great deal of works about HVS-based IQA. In the Visual Signal-to-Noise Ratio (VSNR) metric [19], multiple HVS properties which include contrast sensitivity, visual masking, perceived contrast and global precedence are employed. In [20], contrast sensitivity, contrast interaction and contrast masking are taken into account to obtain a Noise Quality Measure (NQM) of a degraded image which is corrupted by additive noise. In [21] and [22], a sensitive threshold of subband coefficients of different transforms in multiscale geometric analysis and the Contrast Sensitivity Function (CSF) are incorporated to develop a IQA framework. In [23], a semi-local masking of subband coefficients of the wave atoms transform is calculated. In [24], the CSF is adopted to simulate the initial vision processing in the HVS, and the structural randomness of each pixel is employed to quantify the masking effect. Uzair and Dony [25] proposed a just noticeable distortion model in the pixel domain, which combines the CSF, the foveal vision effect, the eye-movement effect, and the content-based masking effect. In [26], the internal generative mechanism of the human brain is modeled by sparse representation. In [27], both visual saliency and visual masking are employed to pool features of three different image regions:

contour region, edge-extension region and flat region. Apparently, HVS-based methods are the most intuitive and reliable IQA ones, and can achieve the best evaluation performance in theory. However, the HVS has complex perception mechanisms, so accurately modeling multiple properties of the HVS is very necessary in IQA scheme.

The wavelet transform has already been employed extensively to derive image features and emulate multiple resolutions and localization characteristics of the HVS. In [19] and [28], the wavelet transform is employed to accomplish image quality evaluation. Though the wavelet achieves remarkable success in many image processing applications, it is far from the optimal in dimensions larger than one. Wang-Q Lim proposed an image sparse representation method, namely, the Discrete Nonseparable Shearlet Transform (DNST) [29]. Multidimensional representations of the shearlet transform exhibit better mathematical and geometrical properties than the wavelet transform, which include multiscale, localization, anisotropy, and directionality [30], [31]. Inspired by these ideas, in this paper, an efficient Multiscale and Multidirectional Visibility Difference (MMVD) predictor is proposed as the Full-Reference Image Quality Assessment (FR-IQA) metric, and the image quality evaluation is conducted in DNST domain. Multiple lowlevel psychophysical characteristics of the HVS employed in this metric include the multi-channel structure [32], the CSF, the contrast masking effect, the entropy masking effect, the visual Just Noticeable Difference (JND) threshold, and error pooling. Firstly, the multi-channel structure is simulated by subband decomposition based on the five-level DNST, and a local directional bandlimited contrast is defined at each position of all subbands at different scales and different orientations in DNST domain. Then a new visual JND threshold model in DNST domain is established. Besides contrast masking, entropy masking is also taken into account to handle the visual masking problem. Entropy masking can be employed to account for the semi-local complexity of an image [33]. Finally, to yield a final quality index, the Minkowski summation is employed to poll perceptual errors of all subbands. To evaluate the efficiency of the proposed MMVD metric, we compare it with subjective evaluation and six objective assessment metrics.

This paper is organized as follows. Section II reviews the characteristics of the DNST. Section III describes the detailed implementation of our MMVD method. In Section IV, experiments conducted on five public IQA databases and thorough analyses are presented. Section VI gives general conclusions.

II. DISCRETE NONSEPARABLE SHEARLET TRANSFORM

It is known that the traditional wavelet transform only provides optimal approximation for one-dimensional piecewise continuous signals with pointwise singularities. But, the wavelet transform has a drawback, i.e., it has limited capability to tackle multivariate and directional signals, such as images and videos. In a two-dimensional image, other types of singularities, such as edge and texture, are usually primary, and the wavelet cannot represent them very efficiently for lacking of directionality. In recent years, the Multiscale Geometric Analysis (MGA) method is presented to overcome this disadvantage. The MGA method includes multiple types, such as the curvelet transfrom [34], the contourlet transform [35], and the shearlet transform [29]–[31], [36]. The shearlet transform is originally proposed in [30] and [31]. It is a polydimensional extension of the conventional wavelet transform, and inherits many advantages of the curvelet and contourlet transforms. Additionally, the shearlet transform can implement sparse representation for multidimensional signals and anisotropic information at multiple scales and multiple directions. So, it can accurately detect signal singularities of images, such as edges. The shearlet transform is being employed gradually in image processing field owing to its characteristics of multiresolution, multidirectional representation, and localized analysis [37]–[40].

In the two-dimensional case, the shearlet transform is defined as

$$SH_{\psi}f\,(j,o,l) =< f, \psi_{j,o,l} > \tag{1}$$

where f denotes a function. j, o, and l respectively represent scaling, orientation, and location parameters. The shearlet $\psi_{j,o,l}$ is given as

$$\psi_{j,o,l} = |detM_{j,o}|^{-\frac{1}{2}}\psi(M_{j,o}^{-1}\,(x-l)) \tag{2}$$

where $M_{j,o} = \begin{pmatrix} j & -\sqrt{j}o \\ 0 & \sqrt{j} \end{pmatrix} = B_oA_j, A_j = \begin{pmatrix} j & 0 \\ 0 & \sqrt{j} \end{pmatrix}$, and $B_o = \begin{pmatrix} 1 & -o \\ 0 & 1 \end{pmatrix}$. A_j and B_o respectively represent a scaling matrix and a nonexpansive shear matrix. In this paper, j and o are set to 4 and -1, respectively. ψ denotes a well localized generating function and can satisfy appropriate admissibilty conditions.

Let f represent a two-dimensional function which is C^2 except for discontinuities along C^2 curves. The N largest coefficients in the different transform are employed to reconstruct f. Here, let f_N denote the reconstructed result. The resulting asymptotic approximation error is

$$\varepsilon_N = ||f - f_N||^2 \tag{3}$$

If the Fourier transform is employed, the asymptotic approximation error is as follows:

$$\varepsilon_N \leq N^{-1/2} \tag{4}$$

The coefficients of the wavelet transform has a slow decline, and its asymptotic approximation error satisfies

$$\varepsilon_N \leq N^{-1} \tag{5}$$

The asymptotic approximation error of the wavelet transform is better than that of the Fourier transform, but it is far from the optimal theoretical approximation error, which is as follows:

$$\varepsilon_N \leq N^{-2} \tag{6}$$

For the shearlet transform, the approximation error satisfies

$$\varepsilon_N \leq N^{-2}(logN)^3 \tag{7}$$

Obviously, for two-dimensional piecewise smooth functions with discontinuities along C^2 curves, the shearlet transform provides a better approximation property than the Fourier transform and the wavelet transform.

Furthermore, the shearlet transform has a number of other advantages relative to the wavelet transform and the contourlet transform. For example, the shearlet transform don't restrict the quantity of orientations. In addition, its shearing filters have lesser support sizes than the directional filters employed in the wavelet transform and the contourlet transform, and can be implemented much more efficiently.

In [29], Lim presented a modified version of the shearlet transform, i.e., the Discrete Nonseparable Shearlet Transform (DNST), which employs a nonseparable shearlet generator. The DNST utilizes a discrete framework, which implements a faithful digitization of the continuum domain directional transform. It uses compactly supported shearlets obtained from a nonseparable generator, and offers the superior localization property of the spatial domain. In addition, its directional selectivity is improved over previous shearlet transforms, and anisotropic singularities of signals which have multiple variables can be encoded sparsely by the DNST.

Fig.1 illustrates subband decomposition of the DNST of the Zonenplatte_Cosinus image at the two scales, and a lowpass subband and multiple highpass subbands are generated. The first scale decomposition has four subbands at different directions, and the second scale decomposition yields six subbands at different directions.

In short, the DNST forms a Parseval frame that is localized fine in the space and frequency domains, and provides better sensibility of the orientation and best sparse approximation to represent images which have edges and textures. With these fine characteristics, the DNST offers more information in regard to images, and is very suitable to be applied in the IQA work.

III. PROPOSED IMAGE QUALITY ASSESSMENT METRIC IN DNST DOMAIN

The HVS is a multichannel structure, namely, different visual information components are preprocessed via different neural channels, and then are inputted into the visual cortex. Considering this multichannel behavior of the HVS, in this paper, the DNST is employed to mimic this behavior and extract an image's features. Fig.2 illustrates the framework of the MMVD method in DNST domain, and this framework includes four major parts. Firstly, in order to construct the perceptual model of the HVS, the multichannel mechanism of information processing in the HVS is emulated by applying the DNST which decomposes the reference and the degraded images into many subbands at multiple scales and multiple orientations. Secondly, coefficients of each subband of the DNST are utilized to compute the local directional bandlimited contrast in DNST domain. Thirdly, the visual JND threshold model is constructed by simultaneously exploiting the CSF and the visual masking effect in DNST

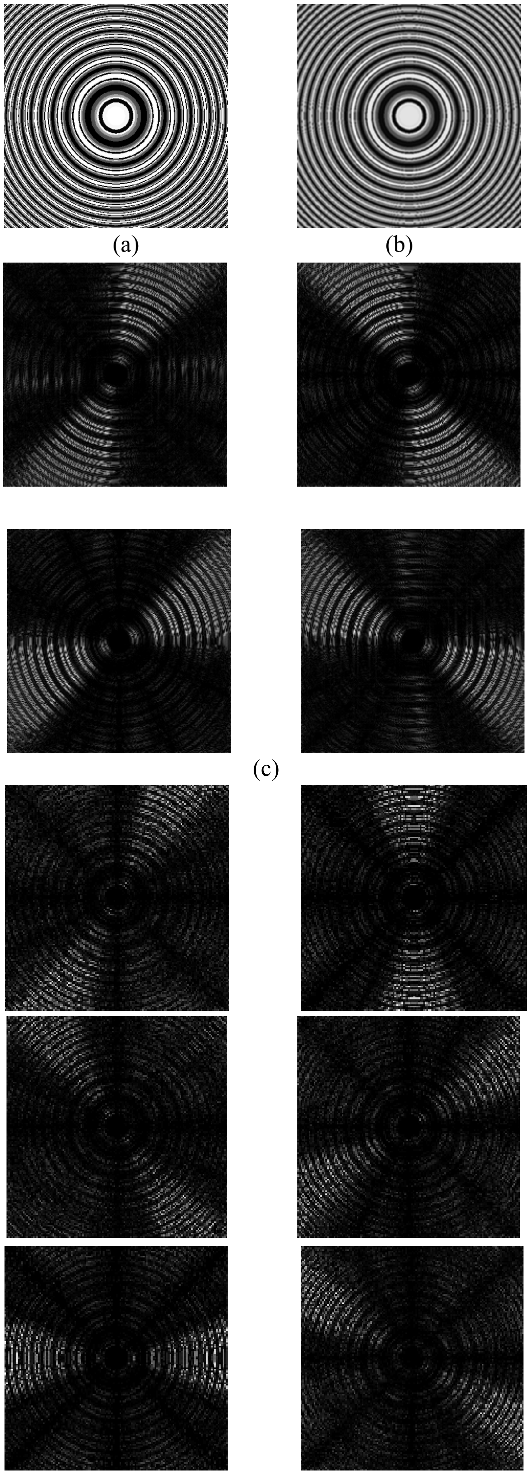


FIGURE 1. An illustration of subband decomposition in DNST domain. (a) The original Zonenplatte_Cosinus image. (b) The lowpass subband. (c) Four highpass subbands at the first scale and four directions. (d) Six highpass subbands at the second scale and six directions.

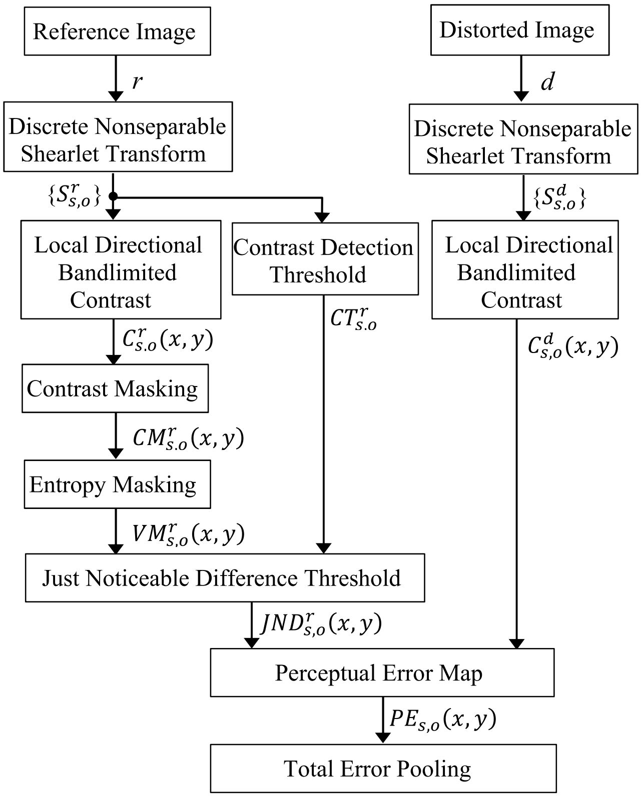


FIGURE 2. The framework of the proposed MMVD metric based on the DNST.

domain. Finally, all perceptual errors are pooled together as an objective index to denote the scalar quality value of the distorted image. Details on each component in this framework of the proposed MMVD metric will be given in the following content.

A. LOCAL DIRECTIONAL BANDLIMITED CONTRAST IN DNST DOMAIN

An image's contrast is one of the most important factors which need to be considered in image processing. We know human perception is not sensitive to the absolute luminance but the relative luminance, i.e., the local variance of the surrounding luminance. In order to account for nonstationarity and local structure of natural and complex images, the local contrast rather than the global contrast needs to be defined. Peli [41] originally presented the definition of the local bandlimited contrast at every point in an image, and the definition of this contrast takes into account both the local luminance and the spatial frequency. On the basis of this local bandlimited contrast, Winkler and Vandergheynst [42], Dauphin *et al.* [43], and Fei *et al.* [44] proposed the local directional bandlimited contrast in the wavelet domain, the Gaussian filter domain and the contourlet domain, respectively. Besides the local luminance and the spatial frequency, directional information is also considered to define the local directional bandlimited contrast. Inspired by these studies, in this paper, a new local directional bandlimited contrast at every point in all subbands at different scales and directions in the DNST domain is defined as follows:

$$C^r_{s,o}(x,y) = \frac{I^r_{s,o}(x,y)}{l^r_s(x,y)} \tag{8}$$

$$l^r_s(x,y) = l^r_0(x,y) + \sum\nolimits_{i=1}^{s-1} I^r_{i,o}(x,y) \tag{9}$$

where $C^r_{s,o}(x,y)$ denotes the local directional bandlimited contrast of the position (x, y) in the reference image's subband at the sth scale and the oth orientation, $I^r_{s,o}(x,y)$ represents the coefficient located at the position (x, y) in the reference image's subband at the sth scale and the oth orientation, and $l^r_s(x,y)$ denotes the background energy, namely, the total of coefficients in all subbands of the reference image whose scale is less than s. In (9), $l^r_0(x,y)$ denotes the coefficient at the position (x, y) in the zero-frequency subband of the reference image, and $I^r_{i,o}(x,y)$ represents the coefficient at the position (x, y) in the reference image's subband at the ith scale and the oth orientation. Analogously, the local directional bandlimited contrast of the distorted image, $C^d_{s,o}(x,y)$, is defined by the same approach.

B. CONTRAST SENSITIVITY FUNCTION IN DNST DOMAIN

The human visual contrast sensitivity is affected by properties of the visual signal, namely, its spatial frequency, orientation, and so on. The contrast sensitivity has a nonlinear bandpass characteristic in the frequency domain, and exhibits different intensities to different spatial frequencies. In general, the wellknown CSF is employed to account for variations in sensitivity about spatial frequencies [12]. The CSF reaches the maximum at the middle frequency, and decreases both the low frequency and the high frequency. Meanwhile, the HVS has the oblique effect, i.e., the eyes of a person have the most sensitivity to the visual signal in the horizontal orientation or the vertical orientation, and have the least sensitivity to the visual signal in the diagonal orientation. In this paper, a model of the CSF, $H(f,\theta)$, is applied to describe quantitatively the sensitive extent of the HVS to different spatial frequencies and directions. This CSF model, $H(f,\theta)$, is initially presented by Mannos and Sakrison [45] and is further modified by Daly [46]. Additionally, $H(f,\theta)$ has been also employed in many IQA metrics [16], [47], [48], [49]. $H(f,\theta)$ is shown as follows:

$$H(f,\theta) = \begin{cases} 2.6\,(0.0192 + \lambda f_\theta)\,exp\,[-\lambda f_\theta] & f_\theta \geq f_{peak} \\ 0.981 & \text{otherwise} \end{cases} \tag{10}$$

$$f_\theta = f/[0.15\cos(4\theta) + 0.85] \tag{11}$$

where f represents the radial frequency in cycles per degree (c/deg), θ denotes the orientation, and f_θ denotes an orientation-based modification of f to account for the oblique effect. In addition, parameters λ and f_{peak} are constants. Here, we have used $\lambda = 0.114$ and $f_{peak} = 8$ c/deg. For the horizontal orientation or the vertical orientation, θ is equal to 0 or $\pi/2$, and $\cos(4\theta)$ is equal to 1. For the diagonal orientations, θ is equal to $\pi/4$ or $3\pi/4$, $\cos(4\theta)$ is equal to -1, and at this point, $H(f, \theta)$ yields an approximately -3dB attenuation.

In this paper, an image is decomposed by the DNST into subbands with different directions and different scales. Here, in consideration of the contrast sensitivity and the oblique effect of human perception, we apply the CSF model, $H(f, \theta)$, to describe subband decomposition of the DNST. The frequency, f, in $H(f, \theta)$ is used to denote the scale, s, of a subband in DNST domain, and the direction, θ, in $H(f, \theta)$ is used to denote the direction, o, of a subband in DNST domain.

The contrast detection threshold denotes the minimum contrast value when an observer perceives a target signal. It is expressed as

$$CT^{r}_{s,o} = \frac{1}{H(f, \theta)|_{f=s,\theta=o}} \tag{12}$$

where $CT^{r}_{s,o}$ represents this threshold value of the reference image's subband at the sth scale and the oth orientation. In (12), the frequency f in $H(f, \theta)$ is equal to the scale s of the DNST, and the direction θ in $H(f, \theta)$ is equal to the direction o of the DNST.

C. VISUAL MASKING EFFECT IN DNST DOMAIN

Visual masking is one important characteristic of human vision, and reflects changes in the visibility threshold of a target signal because of the presence of the background signal. Psychovisual experiments indicate if the target and background signals have similar spatial frequency, direction, phase, location, and so on, the visual masking effect will become more obvious. The visual masking effect includes two parts, namely, the contrast masking effect and the entropy masking effect. The contrast masking effect indicates when the contrast of the background signal changes, the detection threshold of a target signal will also change accordingly. The strength of the background signal can be quantitatively represented by the detection threshold of a target signal. Daly [50] proposed a contrast masking model in the cortex transform domain, and in [44], this model is also employed in the contourlet transform domain. Here, the contrast masking effect in the DNST domain is defined by employing this Daly's contrast masking model:

$$CM^{r}_{s,o}(x, y) = \left(1 + \left(k_1\left(k_2\left|C^{r'}_{s,o}(x, y)\right|\right)^{v}\right)^{b}\right)^{\frac{1}{b}} \tag{13}$$

where $CM^{r}_{s,o}(x, y)$ denotes the visibility threshold elevation at the position (x, y) in the reference image's subband at the sth scale and the oth orientation, which results from the contrast masking effect. Parameters k_1 and k_2 are related to the pivot point of the contrast curve. The parameter b decides the adjacent degree between the curve and the asymptote in the transitional area, and its value varies from 2 to 4. The parameter v denotes the slope of the high masking contrast asymptote, and its value varies from 0.65 to 1. In this paper, k_1, k_2, v, and b are respectively set to 0.0164, 390.325, 0.75, and 4. $C^{r'}_{s,o}(x, y)$ denotes the weighted local directional bandlimited contrast of the reference image. Here, the CSF model, $H(f, \theta)$, is employed to weight the local directional bandlimited contrast. $C^{r'}_{s,o}(x, y)$ is given as follows:

$$C^{r'}_{s,o}(x, y) = C^{r}_{s,o}(x, y) \cdot H(f, \theta)|_{f=s,\theta=o} \tag{14}$$

Besides contrast masking, entropy masking also should be considered simultaneously in the visual masking effect [33]. Contrast masking only takes account of the change of the visual detection threshold because of the contrast value. Entropy masking considers the change of the visual detection threshold because of the neighboring properties. Entropy masking indicates that when the uncertainty of the masking signal changes, the masking extent will also change accordingly. Entropy masking has the important and positive impact, and is the reasonable complement to contrast masking. For instance, it is difficult to detect a distorted signal in texture regions, but not in smooth regions. So, to quantitatively describe the strength of the entropy masking effect, neighborhood properties of a target signal are needed to be considered. In [51] and [52], the Daly's contrast masking model [50] in (13) is modified to define the entropy masking effect in the wavelet transform domain, and in [33], this modified contrast masking model is employed to define the entropy masking effect in the contourlet transform domain. In [23], the same model is applied in the wave atom transform domain. Inspired by the three literatures, in this paper, we define the entropy masking effect in the DNST domain by using this modified contrast masking model, and the visibility threshold elevation in (13) is adjusted as follows:

$$VM^{r}_{s,o}(x, y) = \left(1 + \left(k_1\left(k_2\left|C^{r'}_{s,o}(x, y)\right|\right)^{v+\Delta v(x,y)}\right)^{b}\right)^{\frac{1}{b}} \tag{15}$$

where $VM^{r}_{s,o}(x, y)$ denotes the modified visibility threshold elevation at the position (x, y) in the reference image's subband at the sth scale and the oth orientation, and $\Delta v(x, y)$ represents the neighborhood complexity parameter and is estimated from components of both the reference and the degraded images. A sigmoid function is employed to map the entropy value, $E(x, y)$, into the value, $\Delta v(x, y)$, which is given by:

$$\Delta v(x, y) = \frac{t_1}{1 + e^{-t_2(E(x,y)-t_3)}} \tag{16}$$

where three parameters t_1, t_2, and t_3 are empirically computed from different types of texture in the image. Here, t_1, t_2, and t_3 are respectively set to 0.3, 2, and 1. $E(x, y)$ denotes the neighborhood activity and is computed on a n-by-n neighborhood, which is given by

$$E(x, y) = -\sum p(x, y) log(p(x, y)) \tag{17}$$

where $p(x, y)$ denotes the probability computed from the luminance histogram of the n-by-n surrounding area of the position (x, y). Here, we set $n = 8$.

D. VISUAL JUST NOTICEABLE DIFFERENCE THRESHOLD IN DNST DOMAIN

We know the HVS cannot notice every change in an image, and the visual JND threshold can be exploited to describe this characteristic of the HVS. The visual JND threshold denotes the least visibility threshold and is as a result of physiological and psychophysical phenomena of human vision [53]. If a change is lower than this threshold, it will not be perceived by most observers. The visual JND modeling deals with the issue of visual resemblance, and expresses the local property of the human perception.

In this paper, on the basis of the contrast detection threshold in (12), we define the visual JND threshold in DNST domain by incorporating the visual masking effect, which is given by

$$JND^{r}_{s,o}\left(x,y\right)=CT^{r}_{s,o}\,VM^{r}_{s,o}\left(x,y\right) \tag{18}$$

where $JND^{r}_{s,o}\left(x,y\right)$ denotes the visual JND threshold value of the coefficient at the position $(x,\ y)$ in the reference image's subband at the sth scale and the oth orientation, and $CT^{r}_{s,o}$ denotes the contrast detection threshold value mentioned before. $VM^{r}_{s,o}\left(x,y\right)$ measures the visual masking strength of the position (x,y) in the oth orientation subband at the sth scale, and denotes the increase of the HVS's detection threshold owing to the visual masking effect. Here, in this paper, visual masking measured in the visual JND threshold model consists of contrast masking as well as entropy masking.

E. ERROR POOLING IN DNST DOMAIN

According to above research, we know that if a local contrast error between the reference and the degraded images is less than the visual JND threshold, there will have no or little effect. On the contrary, enough attention should be paid to perceptual distortion caused by this error. Hence, we define the perceptual error of each coefficient of subbands in DNST domain to describe this relationship, which is given by

$$PE_{s,o}\left(x,y\right)=\frac{E_{s,o}\left(x,y\right)}{JND^{r}_{s,o}\left(x,y\right)} \tag{19}$$

where $PE_{s,o}\left(x,y\right)$ denotes the perceptual error map of the oth orientation subband at the sth scale in DNST domain, and $E_{s,o}\left(x,y\right)$ denotes the absolute value of the local directional bandlimited contrast error in DNST domain between the reference and the degraded images. $E_{s,o}\left(x,y\right)$ is represented as

$$E_{s,o}\left(x,y\right)=\left|C^{r}_{s,o}\left(x,y\right)-C^{d}_{s,o}(x,y)\right| \tag{20}$$

where $C^{r}_{s,o}\left(x,y\right)$ represents the local directional bandlimited contrast of the reference image in DNST domain, and $C^{d}_{s,o}(x,y)$ represents the local directional bandlimited contrast of the degraded image in DNST domain. Obviously, according to the definition of the perceptual error proposed in this paper, when $JND_{s,o}(x,y)$ is larger, the perceived error is less for a same amount of $E_{s,o}\left(x,y\right)$. Decreasing value of $PE_{s,o}\left(x,y\right)$ indicates decreasing detected distortions and thus increasing visual image quality. Further, a value of $PE_{s,o}\left(x,y\right)=0$ shows that the distortions in a distorted image cannot be perceived, namely, this image exhibits the optimum visual quality.

Error pooling is crucial in the implementation of image quality assessment metrics. Physiological experiments revealed that numerous cortical cells concentrate on specific areas in their receptive fields, and they pool the outputs from entire photoreceptors on the identical retinal position. Accordingly, the total visual perception of human vision is the integration of each cell's response in the primary cortex of the brain. This wellknown mechanism of human being is referred to as the summation effect. In order to mimic this mechanism, the perceptual errors of all positions of all subbands at total scales and total orientations in DNST domain should be integrated together into a unified perceptual response for a whole distorted image. In this paper, the error pooling scheme is implemented by employing the most widely used nonlinear fusion method, i.e., the Minkowski summation. The Minkowski metric is a normal characteristic of the majority of current image processing paradigms [54]. In this paper, the total error pooling in DNST domain includes two parts, i.e., the intrasubband pooling and the intersubband pooling. Here, the intrasubband error pooling is given by

$$PE_{s,o}=\left(\frac{1}{X_{S,o}Y_{S,o}}\sum\nolimits_{x=1}^{X_{S,o}}\sum\nolimits_{y=1}^{Y_{s,o}}\left[PE_{s,o}(x,y)\right]^{\beta}\right)^{1/\beta} \tag{21}$$

where $PE_{s,o}$ denotes the result of the intrasubband pooling of the oth orientation subband at the sth scale; $X_{s,o}$ and $Y_{s,o}$ denote the height and width of the oth orientation subband at the sth scale, respectively; β is based on practical experience, and is relevant to the psychometric function and probability summation. The range of β is between 2 and infinity. Psychophysical experiments showed 4 is an appropriate selection of β in the intrasubband pooling. So, in this paper, we utilize 4 as the value of β.

The intersubband error pooling includes the pooling of subbands at the same scale and whole orientations, and the polling of subbands at whole scales. They are expressed as

$$PE_{s}=(\frac{1}{O_{s}}\sum\nolimits_{o=1}^{O_{s}}[PE_{s,o}]^{\beta_{o}})^{1/\beta_{o}} \tag{22}$$

$$PE=(\frac{1}{N}\sum\nolimits_{S=1}^{N}[PE_{s}]^{\beta_{S}})^{1/\beta_{S}} \tag{23}$$

where PE_{s} denotes the pooling result value of subbands at the sth scale and whole orientations, O_{s} denotes the number of whole orientations at the sth scale, PE denotes the pooling result value of subbands at all scales and N is the number of total decomposition scales. In this paper, $\beta_{o}=2.3$ and $\beta_{s}=2.5$.

At last, in terms of the Weber-Fechner law, namely, the perceived intensity of a target signal is in direct proportion to the logarithm of the physical magnitude of this target signal, we define the scalar quality assessment value of a distorted image, QA, as follows:

$$QA=log_{10}(PE+C) \tag{24}$$

where the parameter C denotes a constant. To avoid a non-positive result of objective IQA, we set $C = 1$ in this paper.

F. SUMMARY OF IMPLEMENTATION STEPS OF PROPOSED MMVD METHOD

In summary, let $\boldsymbol{r}$ denote a reference image, let $\boldsymbol{d}$ denote the degraded image of $\boldsymbol{r}$, each stage of the proposed MMVD method is roughly summarized in Fig.2, and its detailed implementation steps are given as following:

Stage 1: Perform the DNST of $\boldsymbol{r}$ and $\boldsymbol{d}$ to derive subbands $\{S_{s,o}^{r}\}$ and $\{S_{s,o}^{d}\}$ via (1).

Stage 2: Compute the visual JND threshold.

1. Compute the local directional bandlimited contrasts $C_{s,o}^{r}(x,y)$ and $C_{s,o}^{d}(x,y)$ of each subband of $\boldsymbol{r}$ and $\boldsymbol{d}$ via (8) and (9), respectively.

2. Employ the CSF via (10) and (11) to compute the contrast detection threshold $CT_{s,o}^{r}$ of $\boldsymbol{r}$ via (12).

3. Compute the visibility threshold elevation $CM_{s,o}^{r}(x,y)$ of $\boldsymbol{r}$ owing to the contrast masking effect via (13).

4. Compute the weighted local directional bandlimited contrast $C_{s,o}^{r'}(x,y)$ of $\boldsymbol{r}$ via (14).

5. Compute the modified visibility threshold elevation $VM_{s,o}^{r}(x,y)$ of $\boldsymbol{r}$ owing to both the entropy masking effect and the contrast masking effect via (15), (16) and (17).

6. Compute the visual JND threshold $JND_{s,o}^{r}(x,y)$ of $\boldsymbol{r}$ via (18).

Stage 3: Pool all perceptual errors.

1. Compute the perceptual error map $PE_{s,o}(x,y)$ of each subband via (19) and (20).

2. Compute the intrasubband error pooling $PE_{s,o}$ via (21).

3. Compute the intersubband error polling PE via (22) and (23).

4. Compute the final assessment value QA via (24).

IV. EXPERIMENTAL RESULTS

In this section, we employ the proposed MMVD method to predict the subjective IQA. Then, we present experimental results and evaluate predictive ability of the MMVD metric. Here, in consideration of the reasonable tradeoff between computational complexity and predictive accuracy, the five-level DNST decomposition is applied in the MMVD metric, and five scales are respectively separated into 16, 16, 16, 8, and 8 directional subbands from finer to coarser scales. The Matlab source code of the DNST employed in our proposed MMVD metric is downloaded online at http://www.shearlab.org.

A. IMAGE DATABASES AND PERFORMANCE MEASURES

In order to validate the MMVD metric, we conduct extensive experiments on five public image databases, which include the Tampere Image Database (TID2008) [55], the Tampere Image Database 2013 (TID2013) [56], the Categorical Subjective Image Quality (CSIQ) database [57], the IVC database [58], and the LIVE database [59]. These databases contain vast distorted images and many distortion types, and their properties are listed in Table 1. Moreover, they provide subjective ratings values of distorted images, namely, Differential Mean Opinion Scores (DMOS) or Mean Opinion Scores (MOS), which are exploited as the ground truths to assess predictive ability of objective image quality evaluation metrics.

TABLE 1. Properties of five image databases.

	Reference Images	Degraded Images	Distortion Categories	Typical Image Size	Image Type	Subjects
TID2008	25	1700	17	384×512	Color	838
TID2013	25	3000	24	384×512	Color	971
CSIQ	30	886	6	512×512	Color	35
IVC	10	185	5	512×512	Color	15
LIVE	29	779	5	634×438	Color	29

In our experiments, all color images in these image databases are converted into grayscale images, and this pixel-wise conversion is calculated by

$$I = 0.2989R + 0.587G + 0.114B \tag{25}$$

where I denotes the resultant grayscale image; R, G, and B denotes, respectively, the red, green, and blue component of a color image. In this paper, the Matlab function "rgb2gray" is utilized to implement this pixel-wise conversion.

To quantify predictive ability of image quality evaluation methods, the Video Quality Experts Group (VQEG) [60] recommended three performance criteria, i.e., the prediction accuracy, the prediction monotonicity, and the prediction consistency. A metric's prediction correlation with subjective ratings values, as well as average errors between objective assessment values and subjective ratings values, are employed to quantify the prediction accuracy. Here, the Pearson Correlation Coefficient (CC) and the Root-Mean-Squared Error (RMSE) are employed to quantify correlation and average errors, respectively. Before calculating CC and RMSE, in order to derive a linear relationship between objective values and subjective values, a nonlinear regression is employed to each objective value. Here, we employ a monotonic logistic function to implement this regression, which is given by

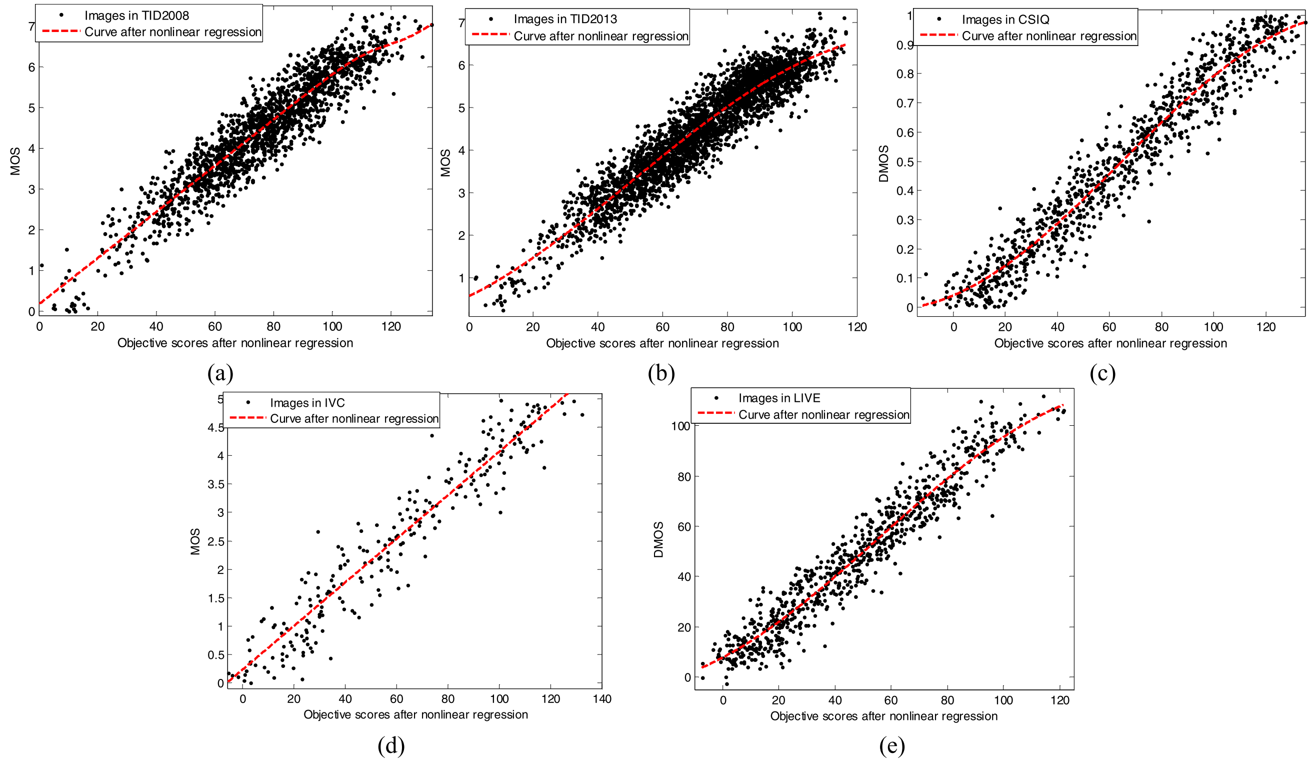
$$f(z) = c_1 + \frac{c_2 - c_3}{\exp\left(\frac{z-c_4}{c_5}\right) + 1} \tag{26}$$

where z denotes the original objective assessment value, and $f(z)$ represents the objective assessment value after the regression. In (26), we employ five parameters c_1, c_2, c_3, c_4, and c_5 to derive the least average square errors between converted objective assessment values and corresponding subjective ratings values. In this paper, the Matlab function "nlinfit" is utilized to accomplish this regression.

Let O_i denote the ith predicted quality value after nonlinear regression, and S_i denote the ith corresponding subjective quality value. Meanwhile, let N denote the total number of images. The formula of CC is given by:

$$CC = \frac{\sum_{i=1}^{N}\left(O_i - \overline{O}\right)\left(S_i - \overline{S}\right)}{\sqrt{\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2}\sqrt{\sum_{i=1}^{N}\left(S_i - \overline{S}\right)^2}} \tag{27}$$

where $\bar{O}$ and $\bar{S}$ denote the average values of O_i and S_i, respectively.



(a) (b) (c)

(d) (e)

FIGURE 3. Scatter plots and fitted curves of the proposed MMVD metric output values versus the subjective ratings values from five image databases (after the nonlinear regression). (a)TID2008. (b)TID2013. (c) CSIQ. (d)IVC. (e) LIVE.

RMSE is computed from

$$RMSE = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}(O_i - S_i)^2} \tag{28}$$

The prediction monotonicity refers to measuring the IQA metric's ability according to predicting the rank-ordering of subjective ratings values. Here, we utilize the Spearman Rank-Order Correlation Coefficient (SROCC) and the Kendall Rank-Order Correlation Coefficient (KROCC) to quantify the monotonicity. SROCC is calculated as:

$$SROCC = 1 - \frac{6}{N(N^2-1)}\sum\nolimits_{i=1}^{N}(X_i - Y_i)^2 \tag{29}$$

where X_i and Y_i denote the ranks of the ith image in the predicted quality values and the subjective quality values, respectively.

KROCC is defined as:

$$KROCC = \frac{2(N^{'} - N^{''})}{N(N-1)} \tag{30}$$

where $N^{'}$ represent the quantity of concordant pairs in the image database, and $N^{''}$ represent the quantity of discordant pairs in the image database.

In order to quantify the prediction consistency between objective assessment values and subjective ratings values, the Outlier Ratio (OR) is employed in this paper. OR is given by

$$OR = \frac{N_o}{N} \tag{31}$$

where N is the quantity of predicted values, and N_o is the quantity of predictions out of the range of two standard deviation of the subjective rating values.

If an objective IQA metric can simultaneously offer higher values of CC, SROCC and KROCC, and lower values of RMSE and OR, it is deemed that this metric achieves better predictive performance.

B. OVERALL PERFORMANCE COMPARISON

Fig.3 depicts scatter plots and fitted curves of the MMVD metric output values versus subjective ratings values of perceived distortion on five image databases including TID2008, TID2013, CSIQ, IVC, and LIVE. Each dot in these plots represents a degraded image of databases. In all graphs, the vertical axis represents distorted images' subjective ratings values, namely, MOS or DMOS, and the horizontal axis denotes MMVD metric output values transformed by the nonlinear regression. These graphs show that the objective evaluation values predicted by the MMVD metric are highly consistent with the subjective evaluation values.

In our experiments, the predictive ability of the proposed MMVD metric is compared with eleven representative FR-IQA metrics, including two extensively used traditional metrics, i.e., PSNR and MultiScale Structural SIMilarity (MSSSIM) [61], four state-of-the-art FR-IQA metrics, i.e., VIF [5], VSNR [19], FSIM [11], NQM [20], and the latest RVSIM [12], EFS [27], GDRW [2], SLY [3], SCQI [62], which are published in 2018, 2018, 2017, 2015, 2016,

TABLE 2. Overall performance comparison of twelve objective evaluation metrics.

		PSNR	MSSSIM	VIF	VSNR	FSIM	NQM	RVSIM	EFS	GDRW	SLY	SCQI	MMVD
TID2008	CC	0.5446	0.8420	0.8078	0.6797	0.8733	0.6162	0.7924	0.8793	0.8817	**0.8918**	**0.8884**	**0.9271**
	SROCC	0.5340	0.8531	0.7495	0.6871	0.8805	0.6242	0.7381	**0.9382**	0.8967	0.8857	**0.9103**	**0.9347**
	KROCC	0.3696	0.6556	0.5862	0.5290	0.6946	0.4524	0.5634	0.7107	0.7131	**0.7183**	**0.7284**	**0.7574**
	RMSE	1.1237	0.7251	0.7900	0.9937	0.6539	1.0594	0.8121	0.6335	0.6317	**0.6176**	**0.6116**	**0.4279**
TID2013	CC	0.6729	0.8266	0.7708	0.7134	0.8577	0.6357	0.7847	**0.9107**	0.8906	0.8237	**0.9067**	**0.9171**
	SROCC	0.6397	0.7809	0.6747	0.6808	0.8015	0.6195	0.6741	**0.8937**	0.8817	0.7973	**0.9047**	**0.9106**
	KROCC	0.4696	0.6047	0.5147	0.5087	0.6289	0.4247	0.5134	**0.7186**	0.6984	0.6104	**0.7334**	**0.7426**
	RMSE	0.9176	0.6976	0.7898	0.8704	0.6255	1.0681	0.7692	**0.5227**	0.5636	0.7125	**0.5224**	**0.5157**
CSIQ	CC	0.7999	0.8807	0.9255	0.8016	0.9101	0.7420	0.9215	**0.9294**	**0.9557**	0.9184	0.9257	**0.9365**
	SROCC	0.8055	0.9031	0.9030	0.8058	0.9230	0.7411	0.8984	**0.9362**	**0.9584**	0.9137	**0.9447**	0.9351
	KROCC	0.6042	0.7395	0.7535	0.6239	0.7567	0.7558	0.7221	**0.7773**	**0.8174**	0.7454	**0.7861**	0.7674
	RMSE	0.1577	0.1197	0.0983	0.1579	0.1080	0.1758	0.0987	**0.0981**	**0.079**	0.1047	0.0994	**0.0845**
	OR	0.3423	0.2459	0.2261	0.3210	0.2346	0.3705	0.2184	0.2364	**0.1847**	0.1965	**0.1767**	**0.1754**
IVC	CC	0.7385	0.8987	0.9110	0.8416	0.9377	0.8494	0.8956	0.9254	**0.9431**	**0.9484**	0.9175	**0.9423**
	SROCC	0.7154	0.8915	0.8963	0.8333	0.9261	0.8344	0.8894	0.9257	**0.9473**	0.9178	**0.9341**	**0.9354**
	KROCC	0.5219	0.7012	0.7165	0.5947	0.7564	0.6038	0.6947	0.7364	**0.7764**	**0.7674**	0.7462	**0.7847**
	RMSE	0.8524	0.5520	0.5239	0.7259	**0.4236**	0.6422	0.5264	**0.4238**	0.5047	0.5284	0.5194	**0.4132**
LIVE	CC	0.8547	0.9403	0.9583	0.9153	**0.9610**	0.9019	0.9561	0.9487	**0.9593**	**0.9592**	0.9367	0.9567
	SROCC	0.8540	0.9446	**0.9610**	0.9161	**0.9663**	0.8997	0.9587	0.9548	**0.9597**	0.9567	0.9454	0.9557
	KROCC	0.6865	0.7996	0.8275	0.7581	**0.8337**	0.7048	0.8194	0.8118	**0.8293**	0.8235	0.7973	**0.8501**
	RMSE	12.4620	8.7735	**7.4048**	9.2801	7.7812	10.4680	7.9284	8.4804	7.6251	**4.6107**	8.8967	**7.4964**
	OR	0.6841	0.5995	0.5421	0.6043	**0.3894**	0.6670	0.5394	0.5567	0.5014	0.5149	**0.4987**	**0.3910**
Weighted mean	CC	0.6805	0.8537	0.8280	0.7447	0.8837	0.6837	0.8291	**0.9101**	0.9070	0.8743	**0.9084**	**0.9278**
	SROCC	0.6625	0.8390	0.7660	0.7322	0.8619	0.6775	0.7617	**0.9192**	0.9073	0.8587	**0.9173**	**0.9263**
	KROCC	0.4891	0.6622	0.6090	0.5616	0.6915	0.5155	0.5966	0.7360	**0.7363**	0.6869	**0.7471**	**0.7638**
	RMSE	2.2147	1.5651	1.4600	1.7835	1.3919	2.0286	1.5177	1.4201	**1.3387**	**1.0551**	1.4666	**1.2468**

respectively. Matlab source codes of these methods are downloaded from websites provided in corresponding literatures, or are derived from their authors. Here, CC, SROCC, KROCC, RMSE, and OR are calculated as experimental results listed in Table 2. For each row of Table 2, three optimal objective evaluation entries are bolded. As TID2008, TID2013, and IVC do not release standard deviations of subjective ratings values, OR cannot be calculated on these databases.

From Table 2, we can see the MMVD metric achieves good performance on total image databases. Especially on TID2008, TID2013, CSIQ, and IVC, the MMVD metric is the most consistent objective image assessment index according to CC, SROCC, KROCC, RMSE, and OR. The MMVD metric demonstrates quite good performance that the CC values are 0.9271, 0.9171, 0.9567, 0.9365, and 0.9423 for TID2008, TID2013, LIVE, CSIQ, and IVC, respectively. In addition, in order to derive an overall performance comparison, we calculate the weighted mean values of performance measures including CC, SROCC, KROCC, and RMSE, which are also given in Table 2. The weight of each database is derived by the calculation, i.e., the quantity of images in each database is divided by the total quantity of images in all databases. Table 3 gives these weights.

Our MMVD metric also obviously outperforms the other metrics according to these weighted mean experimental results in Table 2 except for the RMSE value. Furthermore, the RMSE value of our proposed MMVD metric is very competitive and near with the least value, i.e. the RMSE value of the SLY method.

TABLE 3. Weights of five image database.

Database	TID2008	TID2013	CSIQ	IVC	LIVE
Weight	0.2579	0.4552	0.1344	0.0357	0.1168

C. STATISTICAL SIGNIFICANCE

To verify the statistical significance of the MMVD method compared to the other methods, an F-test is conducted in this paper. The F-test is based on predictive differences between subjective ratings values and nonlinear regression results of objective evaluations values [63], [64]. Predictive differences are defined as follows:

$$D_i = O_i - S_i \quad i = 1, 2, \ldots, N \tag{32}$$

where O_i denotes the ith objective evaluations value after the regression, S_i represents the ith subjective ratings values, D_i represents the difference between O_i and S_i, and N represents the quantity of distorted images.

Predictive differences are supposed to satisfy a Gaussian distribution. If the variance of predictive differences of an objective evaluation metric is smaller than that of predictive differences of the other metric, this metric is deemed to have more accurate prediction performance than the other metric. The variances of predictive differences of each method on five databases are listed in Table 4, and the smallest variance on each database is shown in boldface. Here, F-ratio is employed to represent the proportion between the variances of predictive differences of two objective evaluation methods. In the F-ratio, the bigger variance and the smaller

TABLE 4. Variances of predictive differences of objective evaluation metrics on five image databases.

	PSNR	MSSSIM	VIF	VSNR	FSIM	NQM	RVSIM	EFS	GDRW	SLY	SCQI	MMVD
TID2008	1.124	0.497	0.658	0.956	0.452	1.047	0.678	0.436	0.427	0.431	0.439	**0.408**
TID2013	1.097	0.547	0.647	0.968	0.527	1.174	0.947	0.521	0.548	0.577	0.516	**0.497**
CSIQ	0.0257	0.0157	0.0104	0.0408	0.0148	0.0387	0.0415	0.0173	0.026	0.0143	0.0155	**0.0101**
IVC	0.784	0.297	0.217	0.493	0.287	0.387	0.223	0.281	0.225	0.238	0.236	**0.206**
LIVE	175.93	81.57	56.82	107.21	55.47	121.56	57.46	78.49	54.93	**53.73**	82.34	57.56

TABLE 5. Results of comparison in regard to statistical significance.

	PSNR	MSSSIM	VIF	VSNR	FSIM	NQM	RVSIM	EFS	GDRW	SLY	SCQI
TID2008	1	-	1	1	1	1	1	1	1	1	1
TID2013	1	1	1	1	1	1	1	1	1	1	-
CSIQ	1	1	1	1	-	1	1	1	-	1	0
IVC	1	1	-	1	-	1	1	1	1	_	1
LIVE	1	-	0	1	0	1	-	1	0	0	1

TABLE 6. Mean execution time of twelve FR-IQA Metrics.

FR-IQA metric	PSNR	MSSSIM	VIF	VSNR	FSIM	NQM	RVSIM	EFS	GDRW	SLY	SCQI	MMVD
Time (second)	0.0028	0.1255	2.6122	0.0516	0.5947	0.4453	0.4287	0.3714	0.5658	6.7933	0.2403	0.4361

variance are placed, respectively, as the numerator and the denominator. If an F-ratio is large than the F-test threshold, then the performance comparison of two objective metrics is deemed to be significant in statistical performance, and the metric in the denominator of this F-ratio is deemed to have the statistically superior performance to the metric in the numerator. The F-test threshold relates to both the quantity of predictive differences and the assigned confidence coefficient. Here, the confidence coefficient employed in our experiments is 95%. Table 5 gives the results of comparison which are concerning statistical significance. The entry values "1", "0", and "-" respectively denote that, with 95% confidence, the statistical performance of the MMVD metric is superior, inferior, and indistinguishable to that of the other metrics on the image database in the first column of Table 5.

Results of Table 6 shows the proposed MMVD method has the minimum variance in comparison with the other metrics on most databases except the LIVE database. On LIVE, the variance of the MMVD metric is competitive with the least value, namely, the variance of the SLY method.

Results of Table 5 demonstrate the MMVD metric is better than the most of the other metrics in terms of statistical significance. On TID2008, the proposed metric has better statistical performance except for MSSSIM. On TID2013, the proposed metric exceeds all other metrics statistically except for SCQI. On CSIQ, the MMVD metric outperforms other metrics statistically, except for FSIM, GDRW and SCQI. On IVC, the proposed metric surpasses other metric statistically, except for VIF, FSIM and SLY. On LIVE, the statistical ability of the MMVD method is superior to PSNR, VSNR, NQM, EFS and SCQI, and is inferior to VIF, FSIM, GDRW and SLY. In Table 5, the number of total statistical comparisons between two metrics is 5*11=55, and the number of comparisons in which the proposed MMVD method surpasses the other methods statistically is 41. Hence, the proposed metric shows significant improvement in 74.55% of the cases. In all, the proposed MMVD metric obtains very promising statistical performance when compared with the most of other metrics.

D. COMPUTATIONAL COMPLEXITY ANALYSIS

Computational complexity is one important concern of an objective image quality assessment metric. Here, evaluation experiments are conducted on a computer which has a 3.2-GHz Intel Core4 CPU and a 4-Gbyte RAM. MATLAB R2017a is utilized as the software development platform to implement all codes of these objective evaluation metrics. For each metric, the mean execution time of a distorted image is calculated on the TID2008 database, and the resolution of distorted images is 384×512 pixels. The unit of the execution time is second. Table 6 gives the mean execution time of each metric. Results of Table 6 show the proposed MMVD metric spends less time than VIF, FSIM, NQM, GDRW, and SLY. Specifically, in all metrics, SLY needs the longest time to finish the evaluation. Furthermore, PSNR, MSSSIM, VSNR, RVSIM, EFS, and SCQI spend less time than the MMVD metric, but their performance is worse than that of the MMVD metric, which can be viewed from Tables 2 and 4. In a word, it is apparent that in comparison to the other methods, computational complexity of the MMVD method is moderate and competitive. Our next step is to optimize the programming and further reduce its execution time.

E. DISCUSSION

The proposed MMVD metric is a multiscale image quality evaluation scheme and tries to address the issue of image quality assessment more effectively. Extensive experiments on five image databases are conducted to investigate prediction accuracy and consistency of the MMVD metric. The MMVD metric demonstrates better prediction accuracy than the other metrics. This remarkable performance is attributed to two points. The first point is that the shearlet transform is employed by the MMVD metric, and the shearlet transform has many better characteristics than the traditional wavelet transform in processing two-dimensional image signals, which are already mentioned before. The IQA is accomplished on the basis of the multiscale and multidirectional subbands of the DNST. The second point is that many properties of the HVS are precisely modeled in the MMVD metric. These characteristics include the multichannel mechanism, the local directional bandlimited contrast, the contrast detection threshold, the contrast masking effect, the entropy masking effect, the visual just noticeable difference threshold, and the error pooling.

However, the MMVD metric has two drawbacks. Firstly, it can be only applied to a greyscale image and cannot be applied to a color image. To accomplish the color IQA, the MMVD metric needs to be extended, and some perceptual chrominance characteristics of the HVS should be taken into account. For example, chrominance perception is also aggregate responses of a large number of single space-frequency local channels, which is similar to luminance perception of a greyscale image. The masking effect of the chrominance also affects the performance of color image quality assessment. Secondly, the MMVD metric has the relatively long run time. Therefore, its computational complexity should be further reduced and its implementation codes should be optimized. Additionally, our MMVD metric can also be further extended to handle some other problems of visual quality measure, such as video quality assessment and stereoscopic image quality assessment, which are our future research work.

V. CONCLUSIONS

This paper presents an effective FR-IQA metric on the basis of multiscale and multidirectional visibility differences in DNST domain, and multiple properties of human vision are considered to mimic human responses to incoming image signal. In the proposed MMVD metric, the DNST is employed to emulate the multichannel property of human perception. The image is decomposed by the DNST into multiple different subbands. The CSF and the visual masking effect are dealt with simultaneously and are incorporated into the visual JND threshold in DNST domain. For the visual masking effect, both contrast masking and entropy masking are applied in the proposed metric. The perceptual error polling is implemented by the Minkowski summation and a scalar value is yielded to denote a distorted image's quality. Extensive validation experiments are conducted on five subjective-rated image databases which are specially established to be used in studies of image quality assessment. Experimental results indicate the proposed metric achieves better prediction accuracy and consistency with subjective ratings of human beings in comparison with the existing state-of-the-art IQA metrics, and demonstrate generally better predictive performance.

REFERENCES

[1] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, Oct. 2011.

[2] 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.

[3] Y. Yuan, Q. Guo, and X. Lu, "Image quality assessment: A sparse learning way," *Neurocomputing*, vol. 159, no. 1, pp. 227–241, 2015.

[4] 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.

[5] 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.

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

[7] L. Ding, H. Huang, and Y. Zang, "Image quality assessment using directional anisotropy structure measurement," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1799–1809, Apr. 2017.

[8] T. Dai, K. Gu, L. Niu, Y.-B. Zhang, W. Lu, and S.-T. Xia, "Referenceless quality metric of multiply-distorted images based on structural degradation," *Neurocomputing*, vol. 290, pp. 185–195, May 2018.

[9] 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.

[10] J. Ma, P. An, L. Shen, and K. Li, "Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation," *IEEE Access*, vol. 6, pp. 2768–2780, 2017.

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

[12] G. Yang, D. Li, F. Lu, Y. Liao, and W. Yang, "RVSIM: A feature similarity method for full-reference image quality assessment," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, Dec. 2018, Art. no. 6.

[13] 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.

[14] T. Liu, K. Liu, J. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 107–121, Jan. 2017.

[15] S. Wang, C. Deng, W. Lin, G. 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.

[16] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5293–5303, Nov. 2016.

[17] T.-Y. Kuo, P.-C. Su, and C.-M. Tsai, "Improved visual information fidelity based on sensitivity characteristics of digital images," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 76–84, Oct. 2016.

[18] H. Gao, Q. Miao, J. Yang, and Z. Ma, "Image quality assessment using image description in information theory," *IEEE Access*, vol. 6, pp. 47181–47188, 2018.

[19] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[20] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.

[21] A. Gao, W. Lu, X. Li, and D. Tao, "Wavelet-based contourlet in quality evaluation of digital images," *Neurocomputing*, vol. 72, nos. 1–3, pp. 378–385, Dec. 2008.

[22] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[23] Z. Haddad, A. Beghdadi, A. Serir, and A. Mokraoui, "Image quality assessment based on wave atoms transform," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 305–308.

[24] S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C.-C. J. Kuo, "Compressed image quality metric based on perceptually weighted distortion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5594–5608, Dec. 2015.

[25] M. Uzair and R. D. Dony, "Estimating just-noticeable distortion for images/videos in pixel domain," *IET Image Process.*, vol. 11, no. 8, pp. 559–567, Aug. 2017.

[26] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.

[27] Z. Shi, J. Zhang, Q. Cao, K. Pang, and T. Luo, "Full-reference image qua-lity assessment based on image segmentation with edge feature," *Signal Process.*, vol. 145, pp. 99–105, Apr. 2018.

[28] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A Haar wavelet-based perceptual similarity index for image quality assessment," *Signal Process., Image Commun.*, vol. 61, pp. 33–43, Feb. 2018.

[29] W.-Q. Lim, "Nonseparable shearlet transform," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2056–2065, May 2013.

[30] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *Proc. SPIE*, vol. 5914, Sep. 2005, Art. no. 59140U.

[31] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Anal.*, vol. 39, no. 1, pp. 298–318, 2007.

[32] D. Liu, F. Li, and H. Song, "Regularity of spectral residual for reduced reference image quality assessment," *IET Image Process.*, vol. 11, no. 12, pp. 1135–1141, Dec. 2017.

[33] M. Liu and X. Yang, "Image quality assessment using contourlet transform," *Opt. Eng.*, vol. 48, no. 10, 2009, Art. no. 107201.

[34] J. Yang, H. Wang, W. Lu, B. Li, A. Badii, and Q. Meng, "A no-reference optical flow-based quality evaluator for stereoscopic videos in curvelet domain," *Inf. Sci.*, vol. 414, pp. 133–146, Nov. 2017.

[35] B. H. Ismail, B. Soufiene, and A. Bessaid, "Quality assessment of medical image compressed by contourlet quincunx and SPIHT coding," *J. Mech. Med. Biol.*, vol. 17, no. 6, Sep. 2017, Art. no. 1750097.

[36] G. Kutyniok, W.-Q. Lim, and R. Reisenhofer, "ShearLab 3D: Faithful digital shearlet transforms based on compactly supported shearlets," *ACM. Trans. Math. Softw.*, vol. 42, no. 1, Jan. 2016, Art. no. 5.

[37] C. Wei, B. Zhou, and W. Guo, "Multi-focus image fusion based on non-subsampled compactly supported shearlet transform," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8327–8358, Apr. 2018.

[38] L. Li, Y. Si, and Z. Jia, "Microscopy mineral image enhancement based on improved adaptive threshold in nonsubsampled shearlet transform domain," *AIP Adv.*, vol. 8, no. 3, Mar. 2018, Art. no. 035002.

[39] X. Ma, S. Liu, S. Hu, P. Geng, M. Liu, and J. Zhao, "SAR image edge detection via sparse representation," *Soft Comput.*, vol. 22, no. 8, pp. 2507–2515, Apr. 2018.

[40] H. R. Shahdoosti and O. Khayat, "Image denoising using sparse representation classification and non-subsampled shearlet transform," *Signal Image Video Process.*, vol. 10, no. 6, pp. 1081–1087, Sep. 2016.

[41] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 7, no. 10, pp. 2032–2040, Oct. 1990.

[42] S. Winkler and P. Vandergheynst, "Computing isotropic local contrast from oriented pyramid decompositions," in *Proc. Int. Conf. Image Process. (ICIP)*, Kobe, Japan, vol. 4, Oct. 1999, pp. 420–424.

[43] G. Dauphin, A. Beghdadi, and P. V. de Lesegno, "A local directional band-limited contrast," in *Proc. 7th Int. Symp. Signal Process. Appl. (ISSPA)*, Paris, France, vol. 2, Jul. 2003, pp. 197–200.

[44] X. Fei, L. Xiao, Y. Sun, and Z. Wei, "Perceptual image quality assessment based on structural similarity and visual masking," *Signal Process. Image Commun.*, vol. 27, no. 7, pp. 772–783, Aug. 2012.

[45] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 525–536, Jul. 1974.

[46] S. Daly, "Subroutine for the generation of a two dimensional human visual contrast sensitivity function," Eastman Kodak, Rochester, NY, USA, Tech. Rep. 233203Y, 1987.

[47] 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, 2010, Art. no. 011006.

[48] S. A. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DNT coefficients of locally weighted gradients," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 4117–4120.

[49] J. Ma, P. An, L. Shen, and K. Li, "Joint binocular energy-contrast perception for quality assessment of stereoscopic images," *Signal Process. Image Commun.*, vol. 65, pp. 33–45, Jul. 2018.

[50] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1993, pp. 179–206.

[51] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "On the performance of human visual system based image quality assessment metric using wavelet domain," *Proc. SPIE*, vol. 6806, Feb. 2008, Art. no. 680610.

[52] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Which semi-local visual masking model forwavelet based image quality metric?" in *Proc. Int. Conf. Image Process. (ICIP)*, San Diego, CA, USA, Oct. 2008, pp. 1180–1183.

[53] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *J. Vis. Commun. Image Represent.*, vol. 19, no. 1, pp. 30–41, Jan. 2008.

[54] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 717–730, May 1999.

[55] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. (2008). *Tampere Image Database*. [Online]. Available: http://www.ponomarenko.info/tid2008.htm

[56] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. J. Kuo. (2013). *Tampere Image Database 2013*. [Online]. Available: http://www.ponomarenko.info/tid2013.htm

[57] D. M. Chandler. (2010). *CSIQ Database*. [Online]. Available: http://vision.okstate.edu/csiq

[58] P. L. Callet and F. Autrusseau. (2005). *Subjective Quality Assessment IRCCyn/IVC Database*, 2005. [Online]. Available: http://ivc.univnantes.fr/en/databases/Subjective_Database

[59] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2005). *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality

[60] (2003). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II, Video Quality Expert Group (VQEP)*. [Online]. Available: http://www.vqeg.org

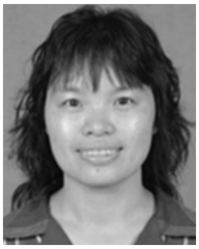
[61] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, vol. 2, Nov. 2003, pp. 1398–1402.

[62] 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.

[63] W. Sun, Q. Liao, J.-H. Xue, and F. Zhou, "SPSIM: A superpixel-based similarity index for full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4232–4244, Sep. 2018.

[64] M. Oszust, "Image quality assessment with lasso regression and pairwise score differences," *Multimedia Tools Appl.*, vol. 76, no. 11, pp. 13255–13270, Jun. 2017.

WU DONG received the M.S. degree in signal and information processing from Anhui University, China, in 2004. He is currently pursuing the Ph.D. degree with the Digital Media Laboratory, Multimedia Technology Center, Beijing University of Posts and Telecommunications, Beijing, China. He is also an Associate Professor with the School of Information Engineering, Beijing Institute of Graphic Communication. His research interests include image processing, machine learning, and data mining.



HONGXIA BIE received the Ph.D. degree from Jilin University, China, in 2000. She is currently a Professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. Her current research interests include multimedia information processing and wireless data transmission.

LIKUN LU received the Ph.D. degree in communications and information systems from the Beijing University of Technology, China, in 2008. He is currently a Professor with the Beijing Institute of Graphic Communication. His research interests include pattern recognition, machine learning, and image processing.

YELI LI received the Ph.D. degree in computer science from Northeastern University, China, in 2000. She is currently a Professor with the Beijing Institute of Graphic Communication. Her current research interests include signal control, data mining, and image processing.