IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 31, 2020, accepted September 28, 2020, date of publication October 1, 2020, date of current version October 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028282

Full Reference Image Quality Assessment by Considering Intra-Block Structure and Inter-Block Texture

JUNFENG YANG[1], WEI ZHANG[2], XIAOLONG LI[1], TING ZHOU[2], AND BO OU[2]

[1]Key Laboratory of Hunan Province for New Retail Virtual Reality Technology, Hunan University of Technology and Business, Changsha 410205, China
[2]College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

Corresponding author: Wei Zhang (zhangweidoccc@gmail.com)

This work was supported in part by the NSFC under Grant 61874120, in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ50288, in part by the Natural Science Foundation of Hunan Province under Grant 2018JJ2196, in part by the Research Foundation of Education Bureau of Hunan Province under Grant 19B309, and in part by the Hunan Province Science and Technology Program under Grant 2017GK2274.

ABSTRACT The purpose of the image quality assessment is to evaluate image quality consistently with human's subjective evaluation. In image quality assessment, many structural features have been used to measure the quality degradation of an image. However, most of the existing image quality assessment methods are block-based, which ignores the features among neighboring blocks. In this paper, we argue that the human visual system perceives distortions not only depends on local structural (intra-block structure) distortions, but also relates to the structural distortions of their neighborhoods (inter-block texture). Based on this insight, we propose a novel image quality assessment method, called the diffusion speed structure similarity (DSSIM), by considering both intra-block structure and inter-block texture. Specifically, to characterize the inter-block texture, we devise a novel visual feature based on the image diffusion speed. To measure the changes of the intra-block structure, we adopt the image gradient magnitude. Furthermore, to differentiate the importance of a local region, we devise a weighting function based on the image diffusion speed. Extensive experimental results on six benchmark databases demonstrate that our proposed method yields a better performance than the state-of-the-art methods.

INDEX TERMS Image quality assessment, nonlinear diffusion, human visual system.

I. INTRODUCTION

The image quality assessment has become a widely used method to improve the performance of image processing and computer vision applications (e.g., image enhancement, compression, restoration, and reproduction [1]–[3]). To get an accurate image quality assessment, an alternative way is to conduct a user survey, and get the Mean Opinion Score (MOS) from a variety of demographics. However, this method is expensive and time-consuming, especially when a huge amount of images require to be assessed. Therefore, it is imperative to devise an objective Image Quality Assessment (IQA) metric, which can simulate people's subjective perception reliably and efficiently.

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

The objective IQA aims to predict the visual quality by simulating subjective ratings with mathematical models. According to the availability of a reference image, the objective IQA methods can be categorized into three categories: (1) full reference (FR-IQA) [1], [4], [5], (2) no-reference (NR-IQA) [6]–[8], and (3) reduced-reference (RR-IQA) [9]–[11]. In this paper, we focus on the FR-IQA methods.

The mean square error (MSE) and peak signal-to-noise ratio (PSNR) metrics are widely used in FR-IQA due to its computational efficiency and clear physical meaning. However, they cannot correlate well with the Human Visual System (HVS). Subsequently, many HVS-based FR-IQA methods are proposed during the recent ten years. Many methods employ the properties of HVS, e.g., visual masking effect [12], contrast [13], and just noticeable differences [14], directly. Chandler and Hemami [15] proposed a method called visual signal to noise ratio (VSNR), by using

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

near-threshold and supra-threshold properties of HVS. However, due to the complexity of the HVS, and the uncertainty of our current knowledge about it, it is difficult to construct an efficient and accurate FR-IQA model.

Since the structural information is recognized as an important feature to simulate HVS, some structure-based methods are proposed. In these methods, the structural similarity (SSIM) [1] index is considered to be one of the most representative FR-IQA. In SSIM, the comparison of image luminance, contrast, and structural features are adopted to perceive the distortions of image quality. In a follow-up study, many novel methods extending SSIM are proposed [16]–[19], such as multi-scale SSIM, information weighted SSIM, gradient-based SSIM, and complex wavelet SSIM. Later on, inspired by the SSIM, some new structure features are introduced to enhance the performance of FR-IQA methods [20]–[25]. Based on the hypothesis that variations of gradient magnitudes in image signals are important in perceived visual quality, Liu *et al.* [20] proposed a gradient similarity (GSM) index. Based on the assumption that image low-level features are important for HVS to understand an image, Zhang *et al.* [22] proposed a feature similarity (FSIM) index. In FSIM, the comparison of the phase congruency and gradient magnitude is adopted to perceive the local distortions of image quality. Zhang *et al.* [23] proposed a visual saliency-induced index (VSI). In VSI, the visual saliency information is incorporated into the IQA model to improve its performance. Recently, some FR-IQA methods using other techniques are proposed. Yang *et al.* [24] proposed an image decomposition based structural similarity (IDSSIM) index, and evaluated the edge and texture components of an image. Cakir and Cetin [26] proposed a FR-IQA using two-dimensional complex mel-cepstrum. The method tries to integrate Fourier transform phase information into the two-dimensional mel-cepstrum, which is shown to be an efficient feature extraction scheme. Nafchi *et al.* [27] proposed a mean deviation similarity index using gradient similarity and chrominance similarity, to measure the local distortions. Sun *et al.* [28] proposed to incorporate image superpixel analysis with IQA. In this method, superpixel luminance and superpixel chrominance similarity are adopted to improve the performance of IQA algorithms. Pei and Chen [29] proposed the multiband IQA method. In this method, the IQA measurement is obtained in several difference of Gaussian bands, and then the importance of the distortion in different frequency bands are learned by the random forest. Bosse *et al.* [30] proposed a FR-IQA using the deep neural networks. Firstly, the features of the reference image and the distorted image are extracted by the Siamese network, then the extracted feature are fused, finally, the quality score is obtained by a regression of the fused feature. Reisenhofer *et al.* [31] proposed a Haar wavelet-based perceptual similarity index (HaarPSI). In HaarPSI, the coefficients obtained from a Haar wavelet decomposition is adopted to measure the local similarities and image region weight. Gu *et al.* [32] proposed a FR-IQA based on Perceptual SIMilarity (PSIM) measure. In this method, the microstructural and macrostructural similarity are adopted to improve the performance of IQA algorithms.

The existing structure features only take advantage of the spatial correlations between image pixels in a local block, while ignoring the inter-block spatial correlations. By considering that the image region of human visual perception is much larger than the block sizes, in [33], the inter-block similarity is taking into account in FR-IQA tasks. In this method, the inter-block is represented by the spatial relationship between a central block and its neighbors within a fixed radius. We argue that the Human Vision System (HVS) perceives distortions, not only depends on local (intra-block) structural distortions, but also relates to the distortions of their neighborhoods (inter-block). For the same structure, when located in a very smooth or rough region, HVS may have different assessments accordingly. For example, a local block with high structure value in smooth-region would have a higher impact on HVS than that in texture region. This indicates that the inter-block texture has an important impact on HVS's image distortion perception.

In this paper, we propose a FR-IQA method called the diffusion speed structure similarity (DSSIM). Our method handles the IQA by considering both intra-block structure and inter-block texture. In this paper, the inter-block is defined as ''the background texture complexity'', while the intra-block is defined as ''the local structural information''. In our method, we adopt the image gradient magnitude to measure the changes of the intra-block structure. We devise the normalized image diffusion speed to characterize the inter-block texture. Specifically, we employ the Total Variation (TV) based nonlinear diffusion method to compute the image diffusion speed. The speed that the pixels change their values depends on their neighbouring pixels' gradient, pixels falling within a texture region move faster than those belonging to a smooth region. Thus, we can use the diffusion speed to measure the spatial correlations of neighbouring blocks. Additionally, the normalized image diffusion speed, the gradient magnitude map, and the chrominance information are used together to evaluate the local image quality. In the pooling stage, the normalized image diffusion speed is further employed as a weighting coefficient to compute a final similarity score. To improve the computational efficiency of the nonlinear diffusion, we employ the Additive Operator Splitting (AOS) scheme. The experimental results on six benchmark databases demonstrate that our proposed method yields a better performance than state-of-the-art methods.

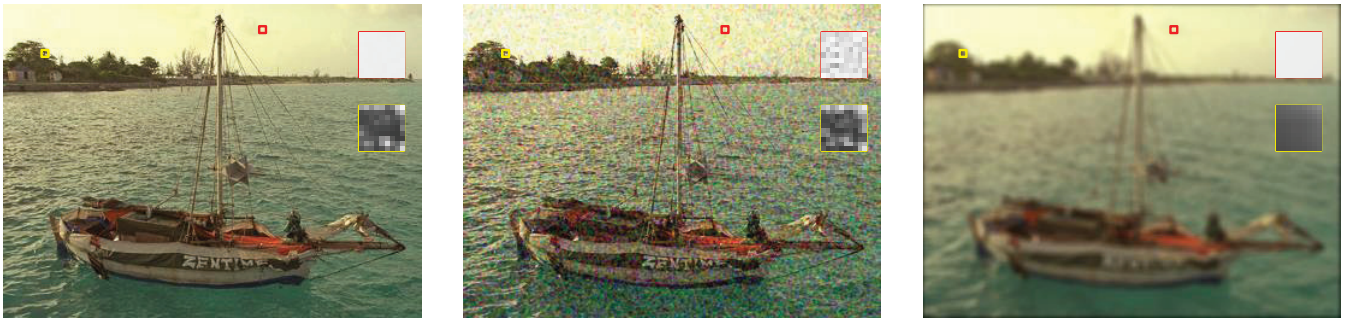
The contributions of this paper are summarized as follows:

1) We propose to use the intra-block structure and inter-block texture to measure the image quality.

2) We devise a novel visual feature based on the image diffusion speed to characterize the inter-block texture.

3) We devise a weighting function based on the normalized image diffusion speed to differentiate the importance of local regions.

The rest of this paper is organized as follows. Section 2 analyzes our new observation for visual quality perception. The



(a) (b) (c)

FIGURE 1. (a) is the original image, (b) and (c) are the corresponding distorted images. (b) is distorted by the high frequency noise (HFN), (c) is distorted by the gaussian blur (GB). The red and yellow boxes bound the homogeneous and the complex texture images regions, respectively. The gray image is displayed in the enlarged boxes (RGB to YIQ).

proposed FR-IQA method based on the image diffusion speed is elaborated in section 3. Experimental results on six databases are illustrated in Section 4, and the conclusion is given in Section 5.

II. MOTIVATION

The existing FR-IQA methods propose to use the structural information to represent perceived visual quality characteristics of HVS, such as variance [1], gradient operators [22], [23], four directional high-pass filter (4D-HF) [20].

However, in many scenarios, the structural based FR-IQA is inconsistent with the HVS perceptions [34]. As indicated by the HVS's contrast masking (CM) effect [35], the HVS has different sensitivities to distortions based on different texture characteristics in the background image region.

To illustrate this, we give an example in Fig. 1. Images in Fig. 1 are obtained from the TID2013. To help describe the changes of image structure information on different distortion types, the gray image are displayed in the enlarged boxes. Fig. 1-(a) is the original image, Fig. 1-(b) is the distorted image caused by High Frequency Noise (HFN), and Fig. 1-(c) is the distorted image caused by Gaussian Blur (GB). The red and yellow boxes bound the homogeneous, and the complex texture images regions, respectively. To estimate the local structural distortions, we use the absolute differences between two gradient values, calculated as $|f(x) - f(y)|$, where $f(\cdot)$ is a gradient operator. The pixels in the yellow and the red box have similar gradient difference value in Fig. 1-(b) and Fig. 1-(c). Obviously, for these two kinds of distortions, the gradient features yield similar distortions, which cannot make clear distinctions.

However, HVS has different strategies for these two kinds of distortions. For HFN distortions, as shown in Fig. 1-(a), and Fig. 1-(b), HVS is more sensitive to the quality degradations in the red box than that in the yellow box. In contrast, for GB distortions, as shown in Fig. 1-(a), and Fig. 1-(c), HVS is more sensitive to the quality degradations in the yellow box than that in the red box. From Fig. 1-(b), we observe that the structural degradations in homogeneous regions are more serious than that in complex texture regions. From Fig. 1-(c), conversely, the structural degradations in complex texture regions are more serious than that in homogeneous regions. This phenomenon illustrates that, for different distortion types, the HVS has different perceive capability for texture regions with different complexity. In this paper, we define the inter-block feature, i.e., the background texture complexity, to capture this characteristics of HVS.

As pointed out in [36], the inter-block texture should be considered when modeling HVS's CM effect. In this paper, we devise a novel feature based on diffusion speed to characterize the inter-block texture. Furthermore, we adopt the image gradient magnitude to capture the intra-block structure. Based on these features, we propose the DSSIM. In the following section, we explain the proposed method in detail.

III. PROPOSED DSSIM

The design of the proposed method is shown in Fig. 2. In our method, the normalized image diffusion speed, which represents the inter-block texture, is adopted as the primary feature when computing our proposed DSSIM. Meanwhile, the gradient magnitude, and chrominance information are employed to estimate the local structure (intra-block) and color degeneration, respectively.

A. INTER-BLOCK TEXTURE FEATURE

The inter-block is the area with the same texture characteristics. In our method, we propose to adopt the normalized diffusion speed to represent the texture characteristics. Inspired by the study [37] proposed by Brox and Weickert, which exploited the total variational (TV) flow based nonlinear diffusion to measure local scale, we adopt the TV-flow based nonlinear diffusion to calculate the image diffusion speed. For a given image I, the TV-flow based nonlinear diffusion calculates a filtered image w^t as a solution of the diffusion equation

$$w^{t+1} = w^t + div\left[\left(g\left(\left|\nabla w^t\right|\right)\nabla w^t\right] \quad (1)$$

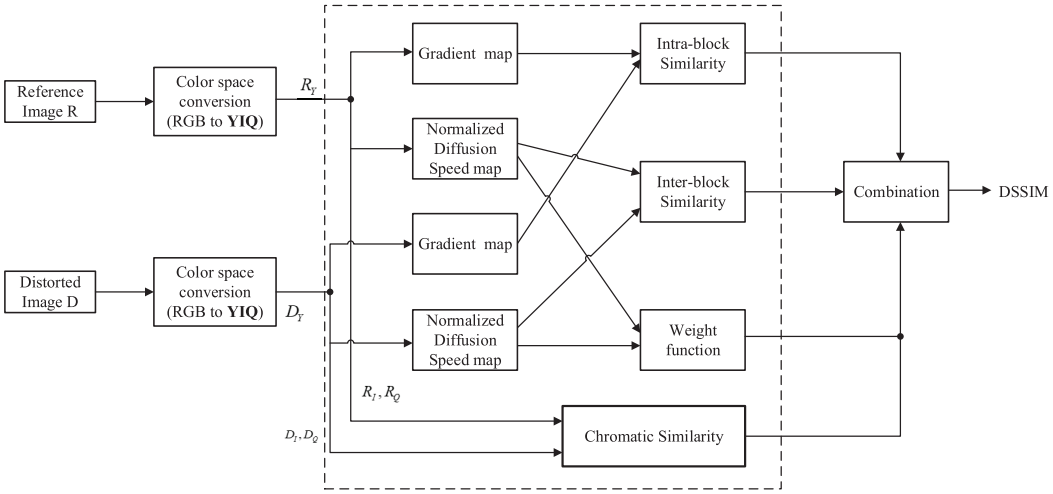


FIGURE 2. The framework of the proposed approach.

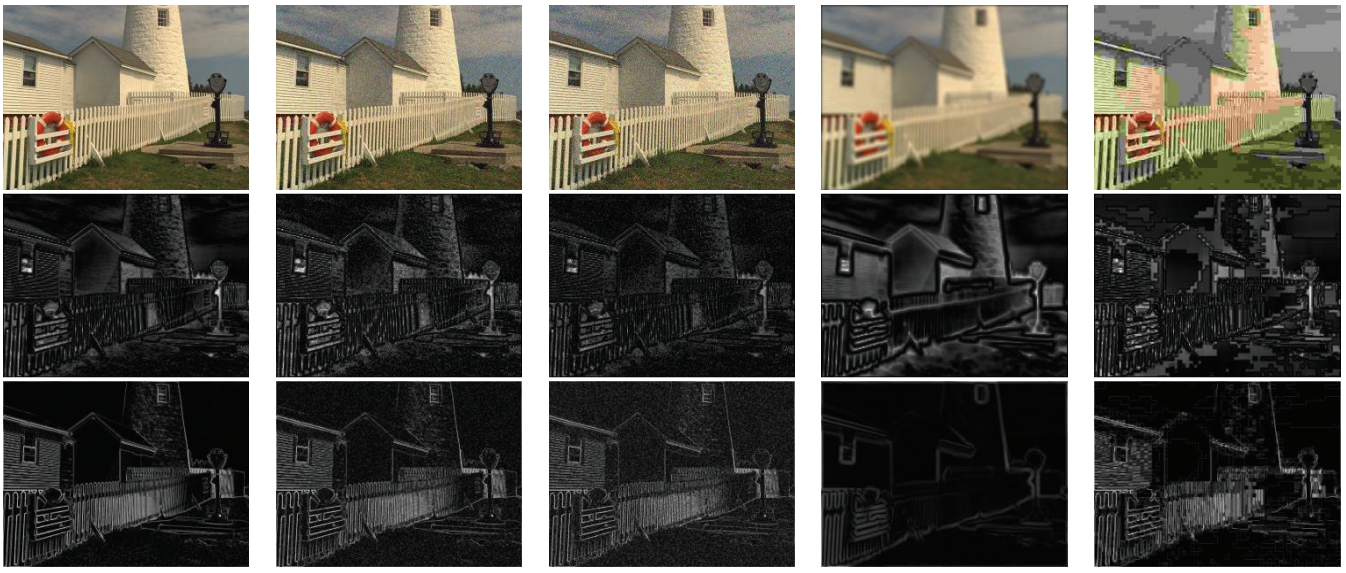


FIGURE 3. Example images from TID2013 (the first row), the normalized diffusion speed map and the gradient map (the second row). (a) Original image; (b) Additive Gaussian noise; (c) High frequency noise; (d) Gaussian blur; (e) JPEG compression.

with the original image as initial state

$$w^0 = I \tag{2}$$

where t denotes the total number of iterations, div denotes the divergence operator, ∇ denotes the gradient operator. For the diffusivity function $g(\cdot)$, we choose the TV flow [38], which is given as

$$g(x) = \frac{1}{\epsilon + x} \tag{3}$$

where ϵ is a small positive constant.

The TV flow has been proven to comply with the following rules [37]. (1) The pixels diffusion speed depends on the size of the region they belong to. (2) The two boundary pixels adapt their value with half of that speed. These rules lead to a very useful consequence: by simply computing the difference in pixel values of the original image and diffused image generated by the TV flow, we can obtain the inter-block texture effectively. The image diffusion speed is defined as

$$DS = I - w^t \tag{4}$$

where t is the number of iterations. Due to the reflecting boundary conditions in TV-flow based nonlinear diffusion, some edge pixels will have higher diffusion values. Therefore, we use the gradient value of the pixel to normalize the diffusion speed, which is given as

$$NDS(x) = \frac{DS(x))}{GM(x) + \theta} \tag{5}$$

where GM is the gradient operator, θ is a small positive constant. The second row in Fig. 3 shows an example of normalized diffusion speed map. In Fig. 3, the original image and four distorted images corrupted by "Additive Gaussian noise", "High frequency noise", "Gaussian blur", and

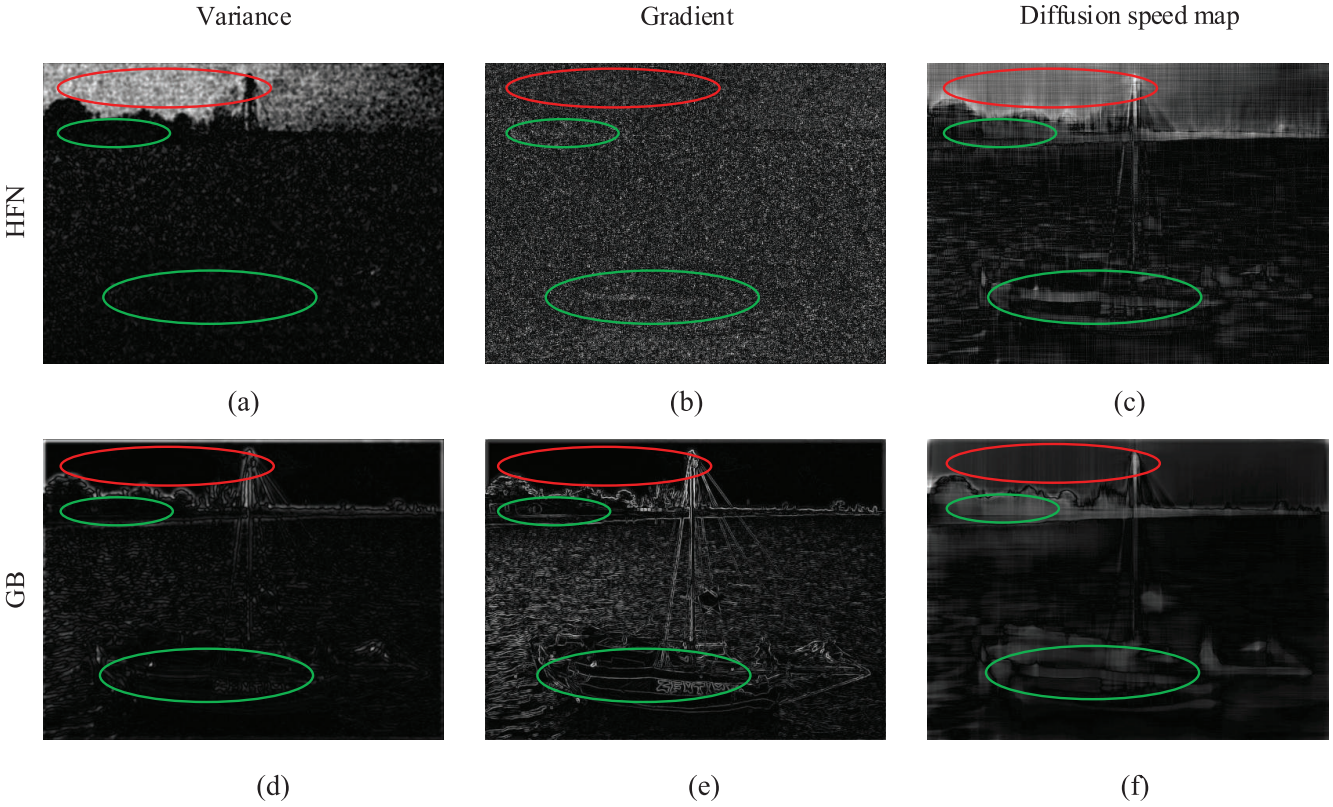


FIGURE 4. The distortion maps of the HFN-distorted image (Fig. 1(b)), and the GB-distorted image (Fig. 1(c)), respectively, for the original image (Fig. 1(a)). A higher intensity value (bright) indicates a relatively larger distortion (difference) in each image.

"JPEG compression" are selected from TID2013. We can observe that the normalized diffusion speed map can represent the region distortion effectively. To support our claim, we compare the normalized diffusion speed with other two popular structural features: variance [1] and gradient operator [22], [23]. Fig. 4 shows the distortion maps of images in Fig. 1(a)-(c). The distortion map between two feature values are calculated by $|f(x) - f(y)|$, where $f(\cdot)$ is a feature function. The first and second rows in Fig. 4 illustrate the distortion maps for Fig. 1-(b) and Fig. 1-(c), respectively. A higher intensity value indicates a relatively larger distortion (difference) in each image. For HFN distortion, the variance exhibits slightly positive correlations with the perceived distortions as shown in Fig. 4-(a), the gradient operator produces similar distortions both in homogeneous and texture regions. For GB distortion, as shown in Fig. 4-(d) and (e), the variance and gradient operator can only reflect the perceived distortions in structural regions.

From Fig. 4-(c) and (f), we can see that diffusion speed map can effectively reflect the perceived distortions for both HFN and GB distortions. For HFN distortions, the diffusion speed map yields small distortion values in complex texture regions (e.g., green ellipse) and lager distortion values in flat regions (e.g., red ellipse), as shown in Fig. 4-(c). In Fig. 4-(f), it produces lager distortion values in complex texture regions (e.g., green ellipse) and small distortion values in homogeneous regions (e.g., red ellipse). Therefore, the normalized diffusion speed map can be used to stimulate the human's perception. Based on the observation, we adopt the normalized diffusion speed as a primary feature in our DSSIM.

In some case, the low efficiency of nonlinear diffusion limits its practical application. To overcome its limitation, several approaches, which based on the simplest finite difference discretization by means of explicit or Euler-forward scheme, are proposed. However, these methods require very small time steps in order to be stable. Hence, the whole filtering procedure is rather time-consuming [39]. Here, we adopt an efficient approach, called the additive operator splitting (AOS) scheme, which is defined as

$$u^{t+1} = \frac{1}{2}\left(\left(I - 2\tau A_x\left(u^t\right)\right)^{-1} + \left(i - 2\tau A_y\left(u^t\right)\right)u^{t-1}\right)u^t \tag{6}$$

where τ is the time step, the A_x and A_y denote the diffusion matrices computed in the horizontal and vertical directions, respectively. More details about the method can be found in [39].

B. INTRA-BLOCK STRUCTURE FEATURE

The image gradient has been successfully applied in many FR-IQA methods to capture image structures. In this paper,

the image gradient is used to extract the image local structure(intra-block) information. For a given image f, the gradient is defined as

$$GM(f) = \sqrt{G_v^2(f) + G_h^2(f)} \tag{7}$$

where $G_v = O_v * f$ and $G_h = O_h * f$ with the symbol $*$ being the convolution operation. O_h and O_v are horizontal and vertical gradient operators. There are several operators that can accomplish this task, such as Sobel operator, Prewitt operator and Scharr operator. In our method, these operators perform almost the same. In this paper, we choose Prewitt operator, which is defined as

$$p_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad p_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Fig. 3 shows an example of gradient magnitude map. The original image and four distorted images corrupted by "Additive Gaussian noise", "High frequency noise", "Gaussian blur", and "JPEG compression" are selected from TID2013. It is evident that the gradient magnitude can extract the image intra-block structure effectively.

C. DIFFUSION SPEED STRUCTURE SIMILARITY (DSSIM)

In DSSIM, we first measure the local similarity maps between the reference image F^{ref} and its distorted one F^{dis}, then an overall quality score is obtained though the pooling stage.

To argument the FR-IQA metric with the capability of dealing with color distortions, chrominance information should be given special considerations. Therefore, the original RGB color images need to be converted into another color space, where the luminance can be separated from the chrominance. To this end, there are some standard color spaces, such as YIQ, YUV, YCbCr and LMN. In this paper, we choose the YIQ color space.

Before computing the diffusion speed and gradient magnitude maps, we first transform RGB color images into an opponent color space [40], where the luminance can be separated from the chrominance. The RGB color inputs are converted into YIQ color space according to

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{8}$$

Then, the diffusion speed and gradient magnitude maps are computed form Y channel, and the I and Q channels of image are used as the chromatic feature to measure color differences. The diffusion speed, the gradient magnitude, and the chromatic similarities between $F^{ref}(x)$ and $F^{dis}(x)$ are defined as follow.

1) for the intra-block structure comparison, we use the gradient magnitude similarity

$$GS(x) = \frac{2 \cdot GM_{ref}(x) \cdot GM_{dis}(x) + C_1}{GM_{ref}(x)^2 + GM_{dis}(x)^2 + C_1} \tag{9}$$

2) for the inter-block texture comparison, we use the normalized diffusion speed similarity

$$NDSS(x) = \frac{2 \cdot NDS_{ref}(x) \cdot NDS_{dis}(x) + C_2}{NDS_{ref}(x)^2 + NDS_{dis}(x)^2 + C_2} \tag{10}$$

3) for the color information comparison, we compute it with

$$CS_I(x) = \frac{2I_{ref}(x) \cdot I_{dis}(x) + C_3}{I_{ref}(x)^2 + I_{dis}(x)^2 + C_3} \tag{11}$$

$$CS_Q(x) = \frac{2Q_{ref}(x) \cdot Q_{dis}(x) + C_4}{Q_{ref}(x)^2 + Q_{dis}(x)^2 + C_4} \tag{12}$$

where C_1, C_2, C_3, and C_4 are positive fixed number to avoid the instability. Finally, the local comparison is defined as

$$M(x) = [NDSS(x)]^{\alpha} \cdot [GS(x)]^{\beta} \cdot \left[CS_I(x) \cdot CS_Q(x)\right]^{\gamma} \tag{13}$$

where α, β, and γ are three parameters used to adjust the relative importance of diffusion speed, gradient, and chrominance features.

Numerous studies have shown that, the pooling is very important for achieving an IQA metric, which is highly correlated with the perception. The most commonly used pooling strategy is average pooling, which simply averages the local quality map as the final FR-IQA score. However, such pooling strategies are not consistent with the intuition that different locations on an image would have different contributions to HVS' perception. A recent study reveals that, less complex texture regions play more important roles in visual quality perception [36]. From the Eq. 5, we know that the normalized image diffusion speed can express the difference between the local area and the background area of the pixel. The higher the normalized diffusion speed value, the less complex texture the located region. Therefore, we utilize the normalized diffusion speed as the weighting coefficient in the pooling stage. The weighting function is defined as $W = max\left(NDS_{ref}, NDS_{dis}\right)$, and the DSSIM index is defined as

$$DSSIM = \frac{\sum_{x \in \eta} M(x) \cdot W(x)}{\sum_{x \in \eta} W(x)} \tag{14}$$

where η means the whole image spatial domain.

IV. SIMULATION RESULT AND DISCUSSION

In this section, we testify and compare the performance of the DSSIM with 13 state-of-the-art FR-IQA metrics to verify its effectiveness on five benchmark IQA databases. These 13 methods include VIF [41], PSNR [42], VSNR [15], SSIM [1], GSM [20], RFSIM [21], FSIMc [22], SFF [43], DSCSI [44], IISIM [33], HaarPSI [31], VSI [23] and PSIM [32].

A. DATABASES AND EVALUATION CRITERIA

The five benchmark databases used in the following experiments include TID2013 database [45], TID2008

database [46], CSIQ database [47], LIVE database [48], IVC database [49] and LIVEMD database [50]. The detailed characteristics of these databases are shown in Table 1.

TABLE 1. Benchmark datasets.

Dataset	Reference Images	Distorted Images	Distorted Types	Subjects
TID2013	25	3000	24	971
TID2008	25	1700	17	838
CSIQ	30	866	6	35
LIVE	29	779	5	161
IVC	10	185	4	15
LIVEMD	15	450	5	37

To make a thorough comparison on the FR-IQA performance, we use four benchmark criteria in our experiments, i.e., Spearman rank order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), Kendall rank order correlation coefficient (KROCC), and the root mean squared error (RMSE).

Before computing the four criteria, we first compute the correlation between the subjective mean opinion scores and the the objective scores with a non-linear transform. To measure the prediction monotonicity, we compute the SROCC and KROCC. To get the similarity degree between subjective mean opinion scores (MOS) and objective scores, we compute PLCC. To measure the prediction consistency, we compute RMSE. In the above four indices, an excellent IQA model should attain high values in SROCC, KROCC, and PLCC, while low values in RMSE.

B. DETERMINATION OF PARAMETERS

The nonlinear regression function is computed as [48]

$$f(x)=\beta_1\cdot\left(\frac{1}{2}-\frac{1}{1+exp(\beta_2\cdot(x-\beta_3))}\right)+\beta_4\cdot x+\beta_5 \tag{15}$$

where $\beta_i, i = 1, 2, \ldots, 5$, are parameters to be fitted. More detailed illustration for the four criteria is elaborated in [17].

In this paper, we used the method proposed by Wang [1] to reduce the complexity. First, the scale S for images viewed from a typical distance is calculated as $S = max(1, round(Num/256))$, where Num is the the number of pixels in image height or width. Then, the image is downscaled by a factor of S. In the LIVE, CSIQ, TID2008, and TID2013 databases, S is equal to 2. We find that this downsizing process helps reduce computational complexity while slightly degrading the performances of the DSSIM. To obtain the best performance of our proposed method, we need to tune the value of parameters. In this paper, we divide the TID2013 database into training set and test set. The training set contains 960 distorted images from 8 reference images, and the test set contains the remaining 1920 distorted images. First, the training data set is used to get the initialization parameters. Second, the initialization parameters are fine tuned on the test set, and we choose the parameters that get the highest SROCC value. To investigate the impact of the diffusion speed map parameters on the final performance, we set the size of the time step τ to different values. In this experiment, we set $t = 5$. Fig. 5 plots the SROCC curves against τ, by applying DSSIM to the LIVE, CSIQ, TID2008, and TID 2013 databases. Note that for all these databases, DSSIM shows similar preference to the value of τ. In our implementation, we set $\tau = 400$. The parameters of DSSIM are listed in Table 2.

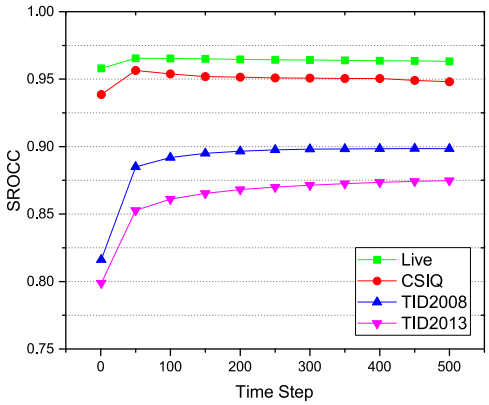


FIGURE 5. The performance of DSSIM in terms of SROCC vs. constant τ on the four databases.

TABLE 2. Parameters setting for DSSIM.

Parameter	t	τ	α	β	γ	C_1	C_2	C_3	C_4
Value	5	400	1	0.29	0.02	170	200	200	200

In DSSIM, the role of the diffusion speed map is twofold: first, the diffusion speed map is used as a critical feature to characterize the distortions of image local quality. Second, the diffusion speed map is used to differentiate the importance of a local region. In our method, the diffusion speed map is used as a weighting coefficient to compute the final score. To demonstrate the effectiveness of our weighting function, we conduct a series experiments on four databases, and employ the SROCC as our evaluation criterion. As demonstrated in Fig. 6, the weighted DSSIM, which uses the

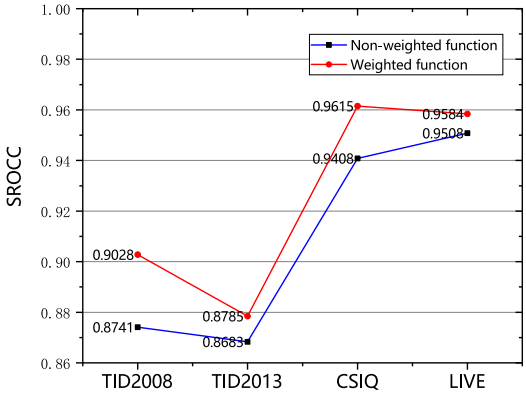


FIGURE 6. The SROCC of DSSIM with weight function and DSSIM without weight function evaluated on LIVE, TID2008, TID2013, and CSIQ.

diffusion speed map as the weighting coefficient, is better than the DSSIM that does not use it.

In our method, the original RGB color images are converted into another color space, where the luminance is separated from the chrominance. There are some standard color spaces, such as YIQ, YUV, YCbCr, and LMN. To determine which color space should be chosen, we tested DSSIM with different color space in TID2013, the experimental results are recorded in Table 3. As shown in table 3, the DSSIM achieves better performance when using YIQ and LMN. Since the YIQ and LMN have similar performance, in the following experiments, we choose the YIQ color space.

TABLE 3. Performance comparison for DSSIM on TID2013 with four color spaces.

	YCbCr	YUV	LMN	YIQ
RMSE	0.6144	0.6134	**0.5587**	**0.5502**
PLCC	0.8686	0.869	**0.8927**	**0.8941**
SROCC	0.8171	0.82	**0.8817**	**0.8785**
KROCC	0.6483	0.6504	**0.6997**	**0.6978**

There are three commonly used gradient operators, the Sobel operator, the Prewitt operator, and the Scharr operator. To determine which operator should be chosen, we tested DSSIM with gradient operators in TID2013, the experimental results are recorded in Table 4. It should be noted that, all the DSSIM parameters used in this experiment are fixed. As shown in Table 4, the DSSIM achieves slightly better performance when using the Prewitt operator. Therefore, in this paper, we choose the Prewitt operator.

TABLE 4. Performance comparison for DSSIM on TID2013 using three gradient operators.

	Sobel	Scharr	Prewitt
SROCC	0.8776	0.8750	**0.8785**
KROCC	0.6964	0.6926	**0.6978**
PLCC	0.8929	0.8902	**0.8941**
RMSE	0.5580	0.5646	**0.5502**

C. PERFORMANCE EVALUATION

We conduct the experiment on five benchmark FR-IQA databases, and compare DSSIM with the state-of-the-art, the comparison is conducted based on four evaluation criteria, i.e., the SROCC, KROCC, PLCC, and RMSE. The experimental results are recorded in Table 5. In Table 5, we highlight the top three FR-IQA methods with boldface font that achieve the top results. As shown in Table 5, it can be observed that none of the methods can work well on all databases. The proposed DSSIM achieves top 3 on TID2013, TID2008 and IVC databases. HaarPSI works well on TID2008, CSIQ and LIVE. PSIM achieves better prediction accuracy on TID2013, TID2008, CSIQ and IVC. For the largest database TID2013, the proposed DSSIM is the third best method on all criteria. Compared with the VSI and PSIM, our DSSIM show similar performances. For the CSIQ and LIVE database, the proposed DSSIM also achieves a comparable performance compared with the state-of-the-art.

Since the performance of the existing methods varies according to different databases, for easy comparison, we compute a weighted-average evaluation score for each method, where the weight coefficient assigned to each database is based on the size of the database. The evaluation criteria are SROCC, KROCC, and PLCC. Table 5 records the results. The FR-IQA method that achieves the top 3 best performance is also highlighted with boldface font. As illustrated in Table 5, our proposed DSSIM achieves the 3 best performance in SROCC and PLCC values.

In Fig. 7, we demonstrate the scatter plots that records the subjective scores and its corresponding objective scores, the objective scores are computed on the TID2013. As we can see from the result, the objective score computed by the proposed DSSIM fits consistently with the objective score, which means that DSSIM is an excellent feature to characterize the HVS behavior.

In order to recommend a more reliable IQA model, we evaluated the 14 FR-IQA methods on TID2008, TID2013, and LIVE databases with the perceptually weighted rank correlation (PWRC) indicator [51]. The PWRC evaluates the rank accuracy though a confidence-aware rank correlation from the area under the curve (AUC). For a given threshold range $[T_{min}, T_{max}]$, the measurement is defined as

$$AUC_{ca} = \int_{T_{min}}^{T_{max}} S(x, y, T)dT \tag{16}$$

where $S(x, y, T)$ is the overall sorting accuracy indicator, AUC_{ca} is a scalar that reflects the overall sorting performance under a given confidence interval. More details about the PWRC can be found in [51].

The experimental results are listed in Table 6. In Table 6, we highlight the top 3 FR-IQA methods with boldface font that achieve the top results. As demonstrated in Table 6, PSIM and our proposed DSSIM rank the top 3 on TID2013 and TID2008, HaarPSI ranks the top 3 on TID2008 and LIVE. As we can see, compared with most of the existing methods, our proposed DSSIM achieves competitive and promising performance on all the five benchmark databases.

D. PERFORMANCE MEASURES ON MULTIPLY DISTORTIONS

To verify the performance of the proposed DSSIM, we conducted the performance comparison on LIVEMD, which is a multiply distorted dataset. The results are listed in Table 7. We highlight the top 3 FR-IQA methods with boldface font that achieve the top results. As shown in Table 7, the methods (such as the VIF, SSF, and FSIMc) having the best performance on LIVEMD, also have competitive performance on LIVE. On the LIVE and LIVEMD dataset, the proposed DSSIM behaves worse slightly than these methods. However, as shown in Table 5, the proposed DSSIM method

TABLE 5. Comparison of 14 IQA indices on five benchmark datasets.

		PSNR	VSNR	VIF	SSIM	GSM	RFSIM	FSIMc	SSF	DSCSI	IISIM	HaarPSI	VSI	PSIM	DSSIM
TID2013	SROCC	0.6862	0.6812	0.6769	0.7417	0.7946	0.7744	0.8510	0.8513	0.8653	0.8109	0.8731	**0.8965**	**0.8925**	**0.8785**
	KROCC	0.5043	0.5084	0.5147	0.5588	0.6255	0.5951	0.6665	0.6588	0.6771	0.6401	0.6922	**0.7183**	**0.7161**	**0.6978**
	PLCC	0.6902	0.7402	0.7720	0.7895	0.8464	0.8333	0.8769	0.8706	0.8708	0.8643	0.8934	**0.8999**	**0.9080**	**0.8941**
	RMSE	0.8976	0.8392	0.7880	0.7608	0.6603	0.6852	0.5959	0.6099	0.6099	0.6099	0.5567	**0.5404**	**0.5193**	**0.5502**
TID2008	SROCC	0.5245	0.7046	0.7491	0.7749	0.8504	0.8680	0.8840	0.8767	0.8573	0.8981	**0.9066**	0.8979	**0.9120**	**0.9028**
	KROCC	0.3696	0.5340	0.5860	0.5768	0.6596	0.6780	0.6991	0.6882	0.6596	0.7167	**0.7372**	0.7123	**0.7395**	**0.7205**
	PLCC	0.5309	0.6820	0.8084	0.7732	0.8422	0.8645	0.8762	0.8817	0.8401	0.8831	**0.9122**	0.8762	**0.9078**	**0.8844**
	RMSE	1.1372	0.9815	0.7899	0.8511	0.7235	0.6746	0.6468	0.6333	0.7278	0.6295	**0.5660**	0.6466	**0.5628**	**0.6261**
CSIQ	SROCC	0.8057	0.8106	0.9195	0.8756	0.9108	0.9295	0.9310	**0.9627**	0.9303	0.9593	**0.9603**	0.9423	**0.9621**	0.9516
	KROCC	0.6080	0.6247	0.7537	0.6907	0.7374	0.7645	0.7690	**0.8281**	0.7572	0.8175	**0.8234**	0.7857	**0.8273**	0.8013
	PLCC	0.8001	0.8002	0.9277	0.8613	0.8964	0.9179	0.9192	**0.9643**	0.9217	0.9536	**0.9579**	0.9279	**0.9642**	0.9401
	RMSE	0.1575	0.1575	0.0980	0.1334	0.1164	0.1042	0.1034	**0.0695**	0.1018	0.0790	**0.0753**	0.0979	**0.0696**	0.0829
LIVE	SROCC	0.8755	0.9274	0.9636	0.9479	0.9561	0.9401	**0.9645**	**0.9649**	0.9480	0.9592	**0.9682**	0.9524	0.9622	0.9584
	KROCC	0.6864	0.7616	0.8282	0.7963	0.8150	0.7816	**0.8363**	**0.8365**	0.7962	0.8251	**0.8447**	0.8058	0.8294	0.8204
	PLCC	0.8721	0.9231	0.9604	0.9449	0.9512	0.9354	**0.9613**	**0.9632**	0.9430	0.9568	**0.9695**	0.9482	0.9584	0.9545
	RMSE	13.3680	10.5060	7.6137	8.9455	8.4327	9.6642	**7.5269**	**7.3460**	9.0890	7.9380	**6.6935**	8.6816	7.7990	8.1397
IVC	SROCC	0.6885	0.7983	0.8966	0.9018	0.8937	0.8270	**0.9293**	0.9256	0.9207	0.9072	0.9169	0.8954	**0.9322**	**0.9267**
	KROCC	0.5220	0.6036	0.7165	0.7223	0.7077	0.6491	**0.7636**	0.7557	0.7471	0.7271	0.7462	0.7156	**0.7700**	**0.7615**
	PLCC	0.7199	0.8032	0.9028	0.9119	0.9038	0.8397	**0.9392**	0.9323	0.9308	0.9150	0.9280	0.9084	**0.9410**	**0.9370**
	RMSE	0.8456	0.7258	0.5239	0.4999	0.5211	0.6616	**0.4183**	0.4403	0.4452	0.4914	0.4537	0.5093	**0.4114**	**0.4250**
WeightedAvg.	SROCC	0.6826	0.7444	0.7692	0.8007	0.8508	0.8485	0.8892	0.8917	0.8887	0.8779	0.9097	**0.9160**	**0.9206**	**0.9101**
	KROCC	0.5433	0.5916	0.6261	0.6359	0.6960	0.6903	0.7303	0.7318	0.7231	0.7292	**0.7601**	**0.7606**	**0.7725**	0.7552
	PLCC	0.7007	0.7648	0.8311	0.8209	0.8711	0.8738	0.8978	0.9024	0.8855	0.8983	**0.9205**	0.9098	**0.9271**	**0.9109**

TABLE 6. Comparison of 14 IQA indices with PWRC.

	PSNR	VSNR	VIF	SSIM	GSM	RFSIM	FSIMc	SSF	DSCSI	IISIM	HaarPSI	VSI	PSIM	DSSIM
TID2013	3.9624	3.6949	4.2022	4.5800	5.0631	5.0186	5.3872	5.3178	5.4102	5.1808	5.5507	**5.6912**	**5.7013**	**5.6897**
TID2008	2.8453	3.8609	4.4149	4.3688	4.9467	5.0661	5.1949	5.1614	4.9171	5.2910	**5.4519**	5.2518	**5.4565**	**5.3126**
LIVE	5.0034	5.4324	5.7642	5.6306	5.7041	5.5418	**5.8007**	**5.7780**	5.6213	5.7292	**5.8286**	5.6667	5.7592	5.7289

TABLE 7. Performance comparison of 14 IQA indices on LIVEMD.

	PSNR	VSNR	VIF	SSIM	GSM	RFSIM	FSIMc	SFF	DSCSI	IISIM	HaarPSI	VSI	PSIM	DSSIM
SROCC	0.6771	0.7727	**0.8801**	0.6534	0.8454	0.8272	0.8635	**0.8700**	0.8266	0.8492	**0.8803**	0.8121	0.8506	0.8501
KROCC	0.5003	0.5775	**0.6946**	0.4703	0.6550	0.6311	0.6728	**0.6824**	0.6358	0.6521	**0.6942**	0.6189	0.6598	0.6568
PLCC	0.7418	0.8117	**0.9009**	0.7384	0.8808	0.8605	**0.8933**	0.8893	0.8665	0.8639	**0.9004**	0.8577	0.8847	0.8834
RMSE	12.6820	11.0450	**8.2078**	12.7550	8.9562	9.6343	**8.4991**	8.6483	9.4575	8.8440	**8.2288**	9.7221	8.8160	8.8613

outperforms these methods on TID2008 and TID 2013. The main reason is that, the LIVE and LIVEMD dataset have less distorted types, and smaller number of distorted images, while TID2008 and TID2013 have more distorted types, and larger number of distorted images. This result indicates that the proposed DSSIM would have a higher generalization capability. Additionally, compared with the methods (such as VSI and IISIM) that have excellent performance on TID 2008 and TID 2013, the proposed DSSIM not only outperforms them on TID 2008 and TID 2013, but also has better performance on LIVE and LIVEMD.

E. PERFORMANCE COMPARISON ON INDIVIDUAL DISTORTION TYPES

An excellent IQA model should not only have high prediction accuracy on a whole databases, but also have high performance on a specific distortion type. To compare the generalization ability of DSSIM with the-state-of-the-art on different distortion types, we conduct a series of experiments on the database TID2013 and TID2008. For the evaluation criterion, we adopt the SROCC due to its stable performance for fitting the objective scores. The experimental results are recorded in Table 8. We highlight the top 3 results with boldface font. As demonstrated in Table 8, our proposed DSSIM outperforms the state-of-the-art. Specifically, DSSIM ranks the top three on TID2013 for 14 distortion types. On TID2008, DSSIM also ranks the top three for 8 distortion types. With these results, we can conclude that, compared with the VSI, PSIM and HaarPSI, our proposed DSSIM has an excellent capability of generalization for different distortion types.

F. STATISTICAL SIGNIFICANCE

To make a through comparison with the state-of-the-art, we conduct the statistical significance tests with F-test. The F-test is computed for every two methods, the detailed

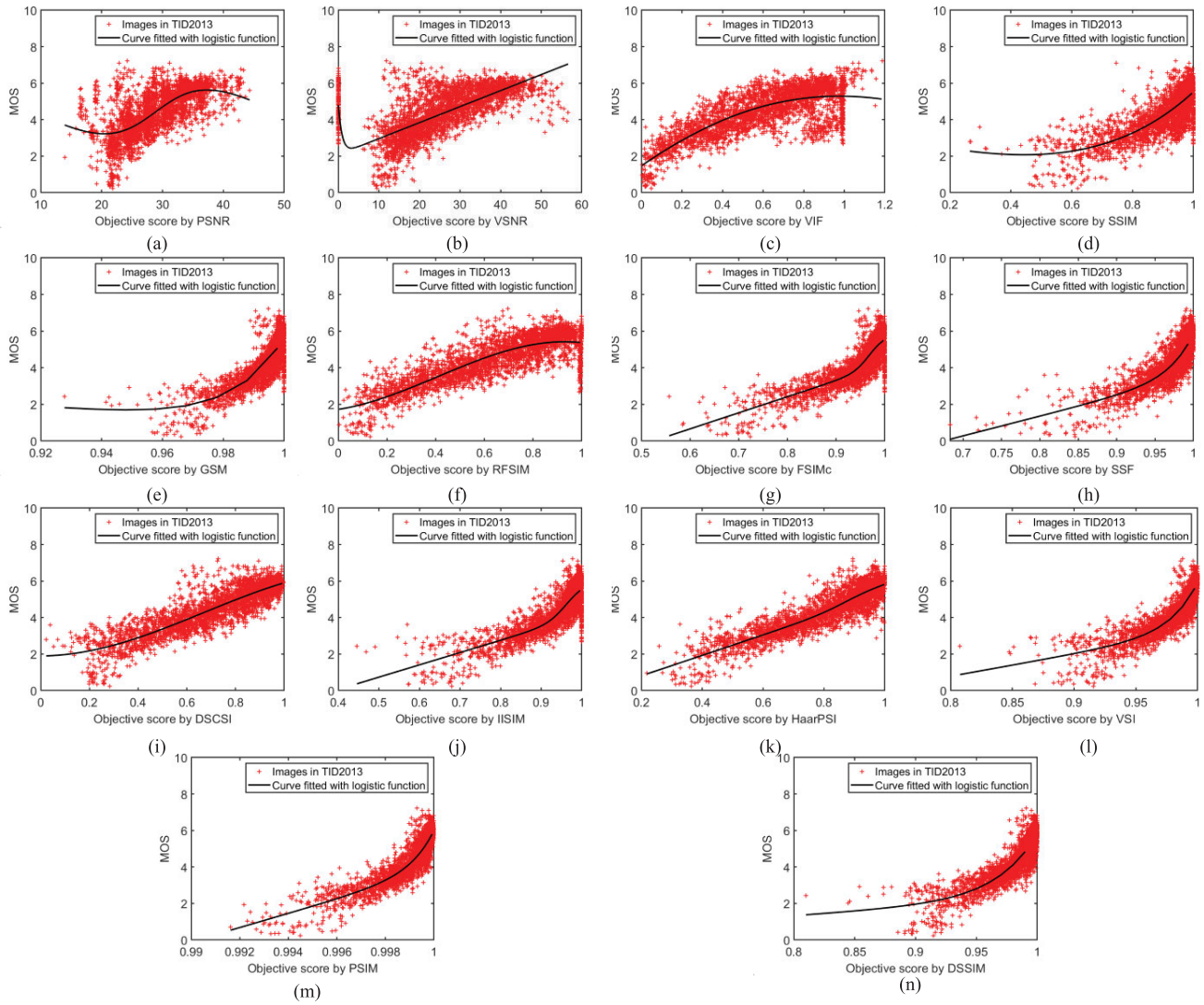


FIGURE 7. Scatter plots that record the subjective MOS and its corresponding objective scores based on the TID2013 database. (a)VIF, (b) PSNR, (c) VSNR, (d) GSM, (e) IWSSIM, (f) RFSIM, (g) FSIMc, (h) SFF, (i) DSCSI, (j) IISIM, (k) HaarPSI, (l) VSI, (m) PSIM and (n) DSSIM.

computation process is demonstrated as follows: 1. compute the prediction residual of each method, where the residual is the absolute difference between the prediction value (the result of non-linear mapping) and the objective score. 2. for each method, compute its residual variance ratio with all other methods, the residual variance ratio is denoted with F. 3. compute $F_{critical}$, which is computed based on the residuals and a given confidence level. 4. if F is larger than $F_{critical}$, then the difference of the two methods is considered to be significant with that confidence level.

We set the confidence level as 95%, and the results of all methods on the four databases are demonstrated in Fig. 8. $+1$ in Fig. 8 means the method on the row is statistically superior than the method on the column, while -1 means that the method on the row is statistically inferior than the method on the column, meanwhile, 0 means the method on the row behaves relatively the same with the method on the column.

As we can see from Fig. 8, our proposed DSSIM outperforms most of the state-of-the-art. Specifically, for the database TID2013, DSSIM only behaves worse than PSIM, and behaves significantly better than the other methods. For the database TID2008, DSSIM behaves significantly better than the other methods except HaarPSI and PSIM. For the database LIVE, besides the VIF, FSIMc, and SFF, DSSIM behaves significantly better than the other methods.

G. COMPUTATIONAL COST

Computational cost is another important factor that determines whether an IQA method can be widely adopted. We conduct a series of experiments to test the computational efficiency. The experiments are conducted on a computer with 2.5GHz Intel core i5 processor, and 10 GB RAM. The program runs on Matlab R2014a. We use all the distorted

TABLE 8. SROCC value of IQA indices for each type of distortion in TID2008 and TID2013.

Distortion	PSNR	VSNR	VIF	GSM	SSIM	RFSIM	FSIMc	SFF	DSCSI	IISIM	HaarPSI	VSI	PSIM	DSSIM
AGN	0.9338	0.8270	0.8996	0.9063	0.8687	0.8877	0.9115	0.9070	0.8917	0.9319	0.9380	**0.9466**	**0.9418**	**0.9460**
ANC	0.8667	0.7266	0.8428	0.8175	0.7726	0.8476	0.8536	0.8166	0.8222	0.8584	0.8612	**0.8742**	**0.8766**	**0.8741**
SCN	0.9245	0.8024	0.8835	0.9158	0.8515	0.8821	0.8905	0.8998	0.8968	0.9157	0.9311	**0.9418**	**0.9476**	**0.9379**
MN	**0.8355**	0.7118	**0.8449**	0.7292	0.7766	**0.8366**	0.8093	0.8184	0.8074	0.7992	0.7855	0.7772	0.7833	0.7643
HFN	0.9182	0.8566	0.8972	0.8869	0.8633	0.9145	0.9058	0.9066	0.8840	0.9151	0.9060	**0.9185**	**0.9217**	**0.9215**
IN	**0.9000**	0.7343	0.8536	0.7964	0.7503	**0.9062**	0.8250	0.7870	0.8538	0.8072	0.8656	**0.8679**	0.8074	0.8600
QN	0.8754	0.8356	0.8161	0.8841	0.8657	**0.8968**	0.8806	0.8638	0.8745	**0.8932**	0.8893	0.8762	**0.9232**	0.8735
GB	0.9102	0.9469	0.9649	**0.9689**	0.9667	**0.9697**	0.9550	**0.9674**	0.9597	0.9594	0.9149	0.9567	0.9227	0.9546
DEN	**0.9503**	0.9104	0.9064	0.9432	0.9254	0.9359	0.9330	0.9090	0.9237	**0.9491**	0.9456	0.9480	**0.9577**	0.9470
JPEG	0.9217	0.9007	0.9191	0.9284	0.9200	0.9398	0.9386	0.9272	0.9374	**0.9527**	0.9512	**0.9550**	0.9508	**0.9556**
JP2K	0.8858	0.9273	0.9516	0.9601	0.9468	0.9518	0.9588	0.9575	0.9518	0.9622	**0.9704**	0.9685	**0.9706**	**0.9690**
JGTE	0.8060	0.8181	0.8441	0.8512	0.8493	0.8786	0.8610	0.8830	0.8783	0.8529	**0.8938**	**0.9042**	0.8719	**0.8869**
J2TE	0.8905	0.8407	0.8760	0.9181	0.8828	0.9102	0.8918	0.8707	0.8895	0.9104	**0.9204**	**0.9225**	0.9151	**0.9215**
NEPN	0.6791	0.6652	0.7719	0.8130	0.7821	0.7704	0.7936	0.7667	0.8013	0.8042	**0.8154**	0.7972	**0.8277**	**0.8140**
Block	0.3297	0.1771	0.5306	**0.6418**	0.5720	0.0338	0.5531	0.1785	0.6148	**0.6222**	0.4970	**0.6448**	0.5461	0.4209
MS	0.7571	0.3632	0.6275	**0.7874**	**0.7751**	0.5547	0.7486	0.6653	0.7685	0.7321	0.7152	**0.7787**	0.7671	0.6798
CTC	0.4466	0.3319	**0.8523**	0.4856	0.4314	**0.5591**	0.4755	0.4902	0.4319	0.4951	**0.4964**	0.4697	0.3937	0.4125
CCS	0.6388	0.3676	0.3099	0.3573	0.4141	0.0204	**0.8358**	**0.8268**	0.7217	0.2522	0.6661	0.8217	0.7810	**0.8460**
MGN	0.8831	0.7644	0.8466	0.8347	0.7803	0.8487	0.8569	0.8434	0.8103	0.8955	0.8903	**0.9076**	**0.9090**	**0.9038**
CN	0.8413	0.8690	0.8948	0.9124	0.8565	0.8917	0.9135	0.9007	0.8834	0.9219	**0.9275**	0.9243	**0.9288**	**0.9318**
LCNI	0.9155	0.8821	0.9229	0.9562	0.9057	0.9009	0.9485	0.9261	0.9316	0.9544	**0.9622**	0.9550	**0.9592**	**0.9631**
LCQD	**0.9201**	0.8695	0.8463	0.8972	0.8542	0.8959	0.8815	0.8794	0.8863	**0.9065**	0.8953	0.8997	**0.9240**	0.8937
CHA	0.8797	0.8644	0.8848	0.8822	0.8774	**0.8990**	**0.8925**	0.8788	0.8592	0.8701	0.8599	**0.8952**	0.8854	0.8769
SSR	0.9108	0.9364	0.9371	**0.9667**	0.9460	0.9325	0.9576	0.9536	0.8950	0.9646	0.9651	0.9622	**0.9678**	**0.9672**
AGN	0.9070	0.7727	0.8797	0.8606	0.8106	0.8415	0.8758	0.8731	0.8958	0.9106	0.9177	**0.9219**	**0.9275**	**0.9274**
ANC	0.8994	0.7793	0.8750	0.8090	0.8029	0.8621	0.8930	0.8625	0.8786	0.9016	0.8982	**0.9089**	**0.9117**	**0.9041**
SCN	0.9169	0.7664	0.8709	0.8941	0.8143	0.8475	0.8718	0.8951	0.9156	0.9017	0.9271	**0.9379**	**0.9416**	**0.9286**
MN	**0.8515**	0.7294	**0.8683**	0.7452	0.7794	**0.8533**	0.8263	0.8365	0.7757	0.8213	0.7959	0.7896	0.8200	0.7837
HFN	**0.9270**	0.8800	0.9074	0.8945	0.8773	0.9181	**0.9233**	0.9187	0.9082	0.9153	0.9155	**0.9227**	0.9217	0.9196
IN	**0.8723**	0.6471	**0.8464**	0.7234	0.6732	**0.8805**	0.7719	0.7483	0.8002	0.7392	0.8269	0.8234	0.7362	0.8027
QN	0.8696	0.8261	0.8816	0.8799	0.8530	**0.8950**	0.8725	0.8471	**0.8903**	0.8855	0.8842	0.8722	**0.9103**	0.8678
GB	0.8684	0.9330	0.9540	**0.9599**	**0.9544**	0.9408	0.9471	**0.9623**	0.9462	0.9467	0.9001	0.9470	0.9068	0.9401
DEN	0.9416	0.9299	0.9182	**0.9724**	0.9529	0.9399	0.9618	0.9383	0.9660	0.9701	0.9701	0.9701	**0.9795**	**0.9707**
JPEG	0.8717	0.9174	0.9167	0.9393	0.9251	0.9385	0.9385	0.9322	0.9394	0.9486	**0.9523**	**0.9573**	0.9398	**0.9601**
JP2K	0.8131	0.9515	0.9709	0.9761	0.9629	0.9487	0.9780	0.9764	0.9695	0.9748	**0.9860**	**0.9828**	**0.9849**	0.9781
JGTE	0.7565	0.8113	0.8585	0.8790	0.8677	0.8534	0.8756	0.8567	**0.9018**	0.8686	0.8921	**0.9062**	0.8910	**0.8991**
J2TE	0.8308	0.7909	0.8500	0.8935	0.8576	0.8591	0.8554	0.8385	0.8909	0.8776	**0.8963**	**0.9073**	0.8884	**0.8974**
NEPN	0.5814	0.5715	0.7619	0.7386	0.7107	0.7274	0.7514	0.6969	0.7683	0.7354	**0.8010**	0.7543	**0.7891**	**0.7755**
Block	0.6192	0.1926	0.8320	**0.8862**	0.8462	0.6258	0.8470	0.5368	0.8331	**0.8731**	0.8026	0.8288	**0.8478**	0.7919
MS	**0.7107**	0.3714	0.5095	**0.7190**	**0.7230**	0.4335	0.6553	0.5224	0.6344	0.6518	0.6051	0.7055	0.6631	0.5885
CTC	0.6042	0.4746	**0.8188**	**0.6691**	0.4411	0.5431	**0.6509**	0.4635	0.3156	0.6434	0.4810	0.6487	0.5463	0.6159

TABLE 9. Time cost of each FR-IQA Index.

Method	VIF	PSNR	VSNR	GSM	SSIM	RFSIM	FSIMc	SFF	DSCSI	IISIM	HaarPSI	VSI	PSIM	**DSSIM**
Time cost (second per image)	1.3476	0.0022	0.6236	0.0456	0.0231	0.1767	0.4733	0.1163	0.9718	0.8624	0.1021	0.2483	0.0537	**0.0813**

images in the TID2013 database, and record the average processing time in Table 9.

As demonstrated in Table 9, the PSNR is the fastest method. However, they achieve fairly worse evaluation performance than DSSIM and the other methods. Compared with the VSI, the performance of the proposed DSSIM is slightly lower on the TID2013. But the DSSIM only spends 0.0813 second per image, i.e. the DSSIM runs three times

	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.
01.PSNR	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
02.VSNR	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
03.VIF	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
04.SSIM	1	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
05.GSM	1	1	1	1	0	1	-1	-1	-1	0	-1	-1	-1	-1
06.RFSIM	1	1	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
07.FSIMc	1	1	1	1	1	1	0	1	1	1	-1	-1	-1	-1
08.SSF	1	1	1	1	1	1	0	0	0	1	-1	-1	-1	-1
09.DSCSI	1	1	1	1	1	1	0	1	0	1	-1	-1	-1	-1
10.IISIM	1	1	1	1	1	1	0	0	0	0	-1	-1	-1	-1
11.HaarPSI	1	1	1	1	1	1	1	1	1	1	0	0	-1	0
12.VSI	1	1	1	1	1	1	1	1	1	1	1	0	0	1
13.PSIM	1	1	1	1	1	1	1	1	1	1	1	1	0	1
14.DSSIM	1	1	1	1	1	1	1	1	1	1	1	0	-1	0

(a). TID2013

	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.
01.PSNR	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
02.VSNR	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
03.VIF	1	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
04.SSIM	1	1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
05.GSM	1	1	1	1	0	-1	-1	-1	1	-1	-1	-1	-1	-1
06.RFSIM	1	1	1	1	1	0	0	-1	1	-1	-1	0	-1	-1
07.FSIMc	1	1	1	1	1	1	0	0	1	0	-1	0	-1	0
08.SSF	1	1	1	1	1	1	1	0	1	0	-1	1	-1	0
09.DSCSI	1	1	1	1	0	-1	-1	-1	0	-1	-1	-1	-1	-1
10.IISIM	1	1	1	1	1	1	1	1	1	0	-1	1	-1	0
11.HaarPSI	1	1	1	1	1	1	1	1	1	1	0	1	0	1
12.VSI	1	1	1	1	1	1	1	1	1	1	0	-1	0	-1
13.PSIM	1	1	1	1	1	1	1	1	1	1	1	1	0	1
14.DSSIM	1	1	1	1	1	1	1	1	1	1	1	-1	1	0

(b). TID2008

	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.
01.PSNR	0	0	0	1	1	0	1	-1	-1	-1	-1	1	-1	-1
02.VSNR	1	0	1	1	1	1	1	-1	-1	-1	-1	1	-1	-1
03.VIF	1	0	0	1	1	1	1	-1	-1	-1	-1	1	-1	-1
04.SSIM	0	0	0	0	1	0	1	-1	-1	-1	-1	1	-1	-1
05.GSM	0	0	0	0	0	0	1	-1	-1	-1	-1	1	-1	-1
06.RFSIM	1	0	0	1	1	0	1	-1	-1	-1	-1	1	-1	-1
07.FSIMc	0	0	0	0	0	0	0	-1	-1	-1	-1	0	-1	-1
08.SSF	1	1	1	1	1	1	1	0	1	1	1	1	1	1
09.DSCSI	1	1	1	1	1	1	1	-1	0	-1	-1	1	-1	-1
10.IISIM	1	1	1	1	1	1	1	-1	1	0	0	1	-1	1
11.HaarPSI	1	1	1	1	1	1	1	-1	1	1	0	1	-1	1
12.VSI	0	0	0	0	0	0	1	-1	-1	-1	-1	0	-1	-1
13.PSIM	1	1	1	1	1	1	1	0	1	1	1	1	0	1
14.DSSIM	1	1	1	1	1	1	1	-1	1	-1	-1	1	-1	0

(c). CSIQ

	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.
01.PSNR	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
02.VSNR	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
03.VIF	1	1	0	1	1	1	0	0	1	1	-1	1	1	1
04.SSIM	1	1	-1	0	-1	1	-1	-1	1	-1	-1	0	-1	-1
05.GSM	1	1	-1	1	0	1	-1	-1	1	0	-1	1	-1	0
06.RFSIM	1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1
07.FSIMc	1	1	1	1	1	1	0	0	1	1	-1	1	1	1
08.SSF	1	1	1	1	1	1	1	0	1	1	-1	1	1	1
09.DSCSI	1	1	-1	0	-1	1	-1	-1	0	-1	-1	0	-1	-1
10.IISIM	1	1	0	1	1	1	-1	-1	1	0	-1	1	0	1
11.HaarPSI	1	1	1	1	1	1	1	1	1	1	0	1	1	1
12.VSI	1	1	-1	1	0	1	-1	-1	1	-1	-1	0	-1	0
13.PSIM	1	1	0	1	1	1	0	-1	1	1	-1	1	0	1
14.DSSIM	1	1	-1	1	1	1	-1	-1	1	0	-1	1	-1	0

(d). LIVE

FIGURE 8. Statistical significance test on evaluation databases. '1' at location (i, j) indicates method i is significantly better than j with 95% confidence level. '−1' indicates the opposite and '0' indicates there is no significant difference between two methods.

faster than the VSI, which means our proposed DSSIM would be more appropriate for real time applications.

V. CONCLUSION

In this paper, we demonstrate that the HVS perceives distortions depending on intra-block structure and inter-block texture. Based on this insight, we adopt the image gradient magnitude to measure the changes of the intra-block structure. Meanwhile, we devise a novel visual feature called region contrast based on the normalized image diffusion speed to characterize the inter-block texture. Additionally, we adopt the normalized diffusion speed, the gradient magnitude, and the chrominance information to evaluate the local image quality. In the pooling stage, we employ the normalized diffusion speed as a weighting coefficient to derive the final similarity score. The experimental results on six public benchmark databases confirm that our proposed DSSIM achieves a superior performance than the state-of-the-art methods.

REFERENCES

[1] 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.

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

[3] D. S. Ly, S. Beucher, M. Bilodeau, S. Persa, K. J. Damstra, R. Pot, and J. V. Rooy, "Automatic color correction: Region-based approach and performance evaluation using full reference metrics," *J. Electron. Imag.*, vol. 24, no. 6, Dec. 2015, Art. no. 061207.

[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] D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Jan. 2010, Art. no. 011006.

[6] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2200–2211, May 2019.

[7] X. Yang, F. Li, and H. Liu, "A survey of DNN methods for blind image quality assessment," *IEEE Access*, vol. 7, pp. 123788–123806, 2019.

[8] C. Deng, S. Wang, A. C. Bovik, G.-B. Huang, and B. Zhao, "Blind noisy image quality assessment using sub-band kurtosis," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1146–1156, Mar. 2020.

[9] W. Zhu, G. Zhai, X. Min, M. Hu, J. Liu, G. Guo, and X. Yang, "Multi-channel decomposition in tandem with free-energy principle for reduced-reference image quality assessment," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2334–2346, Sep. 2019.

[10] P. Paudyal, F. Battisti, and M. Carli, "Reduced reference quality assessment of light field images," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 152–165, Mar. 2019.

[11] Y. Fang, J. Liu, Y. Zhang, W. Lin, and Z. Guo, "Reduced-reference quality assessment of image super-resolution by energy change and texture variation," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 140–148, Apr. 2019.

[12] J. Ross and H. D. Speed, "Contrast adaptation and contrast masking in human vision," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 246, pp. 61–70, Oct. 1991.

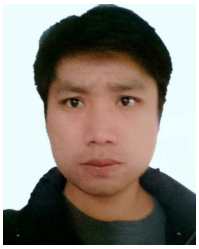
[13] S. J. Daly, "Application of a noise-adaptive contrast sensitivity function to image data compression," *Opt. Eng.*, vol. 29, no. 8, pp. 977–987, 1990.

[14] J. Lubin, "A human vision system model for objective picture quality measurements," in *Proc. Int. Broadcasting Conv.* Edison, NJ, USA: IET, 1997, pp. 498–503.

[15] 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.

[16] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, 2004, pp. 1398–1402.

[17] 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.
[18] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 2929–2932.
[19] C.-L. Yang, W.-R. Gao, and L.-M. Po, "Discrete wavelet transform-based structural similarity for image quality assessment," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 377–380.
[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, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 321–324.
[22] 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.
[23] 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.
[24] J. Yang, Y. Lin, B. Ou, and X. Zhao, "Image decomposition-based structural similarity index for image quality assessment," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 31, Dec. 2016.
[25] 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.
[26] S. Cakir and A. E. Cetin, "Image quality assessment using two-dimensional complex mel-cepstrum," *J. Electron. Imag.*, vol. 25, no. 6, Aug. 2016, Art. no. 061604.
[27] H. Ziaei 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.
[28] 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.
[29] 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.
[30] S. Bosse, D. Maniry, K.-R. Muller, 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.
[31] 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.
[32] 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.
[33] F. Zhou, Z. Lu, C. Wang, W. Sun, S.-T. Xia, and Q. Liao, "Image quality assessment based on inter-patch and intra-patch similarity," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0116312.
[34] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.*, vol. 21, no. 3, pp. 409–418, Jan. 1981.
[35] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.
[36] 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.
[37] T. Brox and J. Weickert, "A TV flow based local scale measure for texture discrimination," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2004.
[38] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón, "Minimizing total variation flow," *Differ. Integral Equ.*, vol. 14, no. 3, pp. 321–360, 2001.
[39] J. Weickert, B. M. T. H. Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 398–410, Mar. 1998.
[40] J.-M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 12, pp. 1338–1350, Dec. 2001.
[41] 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.
[42] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[43] H.-W. Chang, H. Yang, Y. Gan, and M.-H. Wang, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4007–4018, Oct. 2013.
[44] 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.
[45] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2013, pp. 106–111.
[46] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.
[47] D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Jan. 2010, Art. no. 011006.
[48] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[49] P. Le Callet and F. Autrusseau. (2005). *Subjective Quality Assessment IRCCYN/IVC Database*. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/
[50] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 1693–1697.
[51] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2499–2513, May 2018.

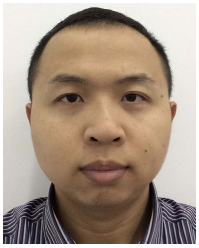


JUNFENG YANG received the B.S. degree in electronic information engineering from the Hunan University of Arts and Science, China, in 2006, and the M.S. and Ph.D. degrees in software engineering from Hunan University, China, in 2010 and 2018, respectively. Since 2017, he has been a Lecturer with the College of Computer and Information Engineering, Hunan University of Technology and Business, Changsha, China. His current research interests include image processing and computer vision.

WEI ZHANG received the B.S. degree in computer science from Hunan University, China, in 2011, and the Ph.D. degree with the College of Information Science and Engineering, Hunan University, in 2016. He was also a Visiting Student with Temple University, Philadelphia, PA, USA. His research interests include cloud computing, network security, computer vision, and data mining.

XIAOLONG LI received the B.E. degree from the Harbin Institute of Technology, Harbin, China, in 2003, and the Ph.D. degree from Hunan University, Changsha, China, in 2008. Since July 2017, he has been with the Hunan University of Technology and Business, Changsha. He is currently a Professor. His research interests include sensor networks, M2M networks, and wireless networks. In 2013, he won the Best Paper Award at the ChinaCom Conference.

TING ZHOU received the B.S. degree from the University of South China, in 2011, and the M.S. degree from the College of Information Science and Engineering, Hunan University. Her research interests include cloud computing, computer vision, and the Internet of Things.



BO OU received the B.S. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2008 and 2014, respectively. He is currently an Associate Professor with Hunan University. His current research interests include image processing and digital watermarking.