IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2019, accepted March 24, 2019, date of current version April 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909073

No-Reference Stereoscopic Image Quality Assessment Based on Visual Attention and Perception

YAFEI LI, FENG YANG, WENBO WAN, JUN WANG, MIN GAO, JIA ZHANG, AND JIANDE SUN, (Member, IEEE)

School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

Corresponding authors: Feng Yang (yangfeng@sdnu.edu.cn) and Jiande Sun (jiandesun@hotmail.com)

This work was supported in part by the Natural Science Foundation of China under Grant U1736122, in part by the Natural Science Foundation of China under Grant 61601268, in part by the Natural Science Foundation of Shandong Province under Grant ZR2016FB12, and in part by the Shandong Provincial Key Research and Development Plan under Grant 2017CXGC1504.

ABSTRACT In recent years, the methods of no-reference stereoscopic image quality assessment (NR-SIQA) have been well investigated, but there still remain challenges due to the inaccurate extraction of binocular perception information. In this paper, we propose an NR-SIQA method based on visual attention and perception. We combine saliency and just noticeable difference (JND) to model visual attention and perception, respectively, and weight the global and local features extracted from the left and right views. Meanwhile, in order to obtain the accurate binocular perception information, the global structural features reflecting spatial correlation are extracted from the cyclopean map that is synthesized by the left and right views. Then, a regression model is learned based on a support vector machine regression (SVR) to evaluate the quality of stereoscopic images. The experiments on popular SIQA datasets demonstrate that the proposed NR-SIQA method has better and more reliable performance than the state-of-the-art methods.

INDEX TERMS Stereoscopic image quality assessment, saliency model, JND model, cyclopean map.

I. INTRODUCTION

With the rapid development of electronic and communication technology, 3D movies, 3D-TV, 3D digital cameras and other devices are producing countless stereo images every day. However, due to the complex structure of the stereo information, there usually exists image degradation due to the distortion during its transmission, storage, compression and so on. The degraded images need to be restored before use and stereoscopic image quality assessment (SIQA) indices can provide a quality evaluation for the restoration. SIQA aims to assess the quality of stereoscopic images just as the human does. Similar to the 2D-based IQA methods, most existing SIQA methods could be divided into three categories according to the usage of the original image, which are full-reference (FR) [1]–[4], reduced-reference (RR) [5]–[8], and no-reference (NR) [9]–[13] SIQA. In practical cases, we can not get the original vision of the distorted image, which makes the applications of FR and RR SIQA methods limited. Therefore, we focus on the no-reference SIQA (NR-SIQA) method in this paper.

The 2D-based NR-SIQA methods are proposed to process the left and right views separately, and then to construct a model for binocular fusion. Chen *et al.* [9] extracted natural scene statistics (NSS) features from the left and right views. Liu *et al.* [14] proposed to classify the distortions on the left and right views separately, and evaluate the quality according to the classification of distortion. Yang *et al.* [15] developed a NR-SIQA method by learning color characteristic, which is based on gradient dictionary. Ryu and Sohn [16] proposed to compute the perceptual blurriness and blockiness scores of the left and right views and then combined them into an overall quality evaluation. Tian *et al.* [17] extracted the monocular features from the left and right views via Gabor filtering, extracted the binocular features from the cyclopean view, and put these features into the depth belief network (DBN) to predict the quality score of the stereo image. Zhang *et al.* [13] proposed a convolutional neural network (CNN)-based method that can learn the complex mapping between the original image and its quality label.

The associate editor coordinating the review of this manuscript and approving it for publication was Chang-Tsun Li.

However, these binocular perception-based methods did not take the perceptual variation on different kinds of regions into consideration. In our method, we consider both visual attention and visual perception for NR-SIQA.

SIQA has a certain relationship with visual attention and perception. The visual attention is the selection mechanism of visual saliency regions. Visual saliency processes the important part and ignores the unimportant part of the visual information selectively [18], [19]. For quality assessment, the distortions presented in the salient regions draw more attention from the human viewers. In other words, the perceptual quality of the salient regions tends to represent the perceptual quality of the whole image [20]. Liu *et al.* [21] developed a 3D saliency model based on the absolute disparity map, which was used to weight the left and right views. The saliency model could assign appropriate weights to more perceptually important area. Yang *et al.* [22] proposed a novel saliency model in SIQA method for the cyclopean map called ''cyclopean saliency'', where the binocular combination characteristics with the cyclopean saliency are all considered. In this model, the saliency areas of the left and right views were first obtained and then synthesized to obtain the cyclopean map. The features of the saliency areas were used to obtain more effective binocular perception information. Since saliency map could indicate the relative importance of pixels in the spatial domain for left and right views, Wang *et al.* [23] proposed a quaternion representation based saliency model in stereopair, which comprises the image content, the inter-view disparity, and the difference map. In the model, the saliency maps were employed to pool the error maps produced by reference images and distortion images of the left and right views.

The just-noticeable-difference (JND) is defined as the largest perceptible distortion [24]. In other words, if the given distortion is below the JND threshold, it can not be perceived visually. JND can reflect the characteristics of visual perception. It is widely used in image/video encoding [25] and visual quality assessment [26]–[28]. And some JND-based methods are also used in SIQA. Shao *et al.* [26] proposed a JND-based SIQA method, in which JND model is used to calculate the visual sensitivity of binocular fusion and suppression regions. Fezza *et al.* [27] used the binocular JND to modulate the quality score of each region. And the method adjusted the 3D quality referring to the JND model without any cue about the distortion type and distribution. Fan *et al.* [28] proposed to assess quality by combining the quality of the JND-based cyclopean map with the quality of disparity map. The JND model was used to modulate the quality of the cyclopean map according to the visual importance of each pixel.

Visual saliency specifies the selected regions of interest and its degree of interest, while JND represents the visual perception threshold within the regions of interest. In this paper, saliency and JND model are combined as the weight factors for feature fusion of left and right views. To our best knowledge, there is no such work in the literatures where the combination of these two models are used for the quality assessment of the stereoscopic images. Additionally, the binocular perception of the visual cortex has a direct impact on the NR-SIQA. In this paper, we not only perform feature fusion on the basis of the monocular view, but also synthesize a cyclopean map based on the left and right views. Since the cyclopean map can well embody the visual scene perceived by the HVS, we extract the global structural features of the cyclopean map based on spatial correlation, which is another main feature for quality assessment. Finally, a regression model is learned based on a support vector machine regression (SVR) to evaluate the quality of stereoscopic images. The framework of the proposed method is shown in FIGURE 1. The main contributions are as follows:

1) Both visual attention and perception are considered. The saliency model and JND model are used for weighting the features in different views and regions.
2) Both global and local features are considered. The global structural features and local entropy features are extracted from the left and right views respectively. Since the cyclopean map, which is synthesized from left and right views, can represent the perception fusion of 3D scene, the global structural features are also extracted from the cyclopean map.
3) A regression model is learned to build the relationship between these extracted features and their subjective ratings (e.g., mean opinion score (MOS) or different mean opinion score (DMOS)) via support vector machine regression (SVR) as SVR is popular method used in previous works. Based on a large number of experiments, our method outperforms other state-of-the-art methods on four stereoscopic datasets.

The rest of the paper is structured as follows. In section II, the details of the proposed method is introduced. Section III provides the experimental process and performance of our proposed method. General conclusions and future works are given in section IV.

II. THE PROPOSED NR-SIQA METHOD

In this paper, we propose a NR-SIQA method based on the combination of visual attention and perception. Specifically, in this method, both multi-scale global and local features are considered, and the saliency model and JND model are used to weight these features. In order to obtain more accurate binocular perception information, the left and right views are synthesized to obtain binocular cyclopean map, and the global structural features are extracted from the binocular cyclopean map.

A. FEATURE FUSION BASED ON SALIENCY MODEL AND JND MODEL

Both monocular and binocular features are extracted from the left and right views. We extract the statistical features from the spatial domain and the structural features from the gradient domain in a global perspective. In order to obtain

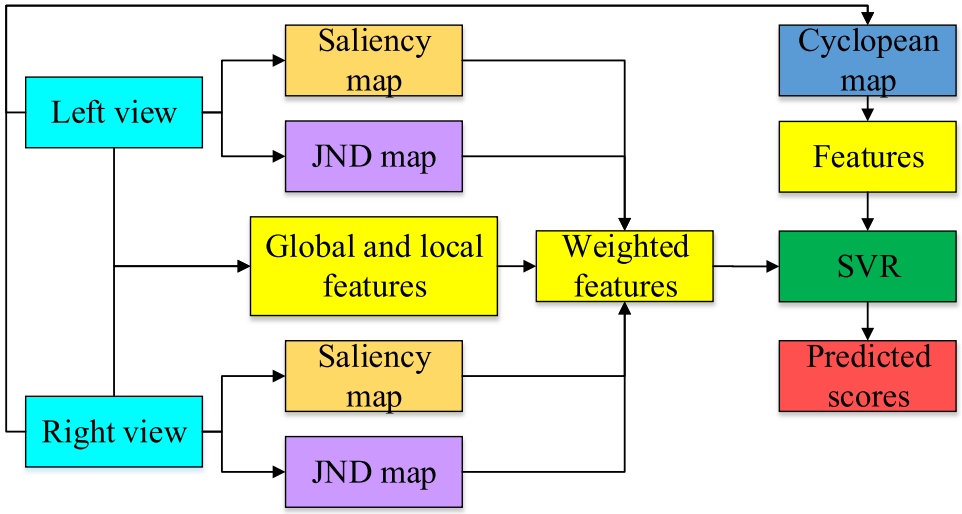


FIGURE 1. The framework of the proposed method for NR-SIQA.

more effective features, we extract the entropy information of the left and right views locally based on partitioning. Since the stereo image perceived by the HVS is fused, we use the saliency model and JND model to weight and fuse the features extracted from the left and right views to obtain more reliable quality assessment features.

1) GLOBAL FEATURE EXTRACTION

a: NATURAL SCENE STATISTICS

The regularity of NSS has been well established in the study of human vision, where the regularity has been demonstrated in both spatial domain [29] and wavelet domain [30]. We use the method in [31] to extract the statistical features of natural scenes as the global features of stereopairs. For a given image, the normalization process is

$$\begin{cases} I_M(i) = \frac{I(i)-\mu(i)}{\sigma(i)+C} \\ \mu(i) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(i) \\ \sigma(i) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i)-\mu(i))^2}, \end{cases} \tag{1}$$

where i is the pixel coordinates of the image. C is a constant which prevents the instability in case of denominator tending to zero. The transformed luminance $I_M(i)$ denotes the mean subtracted contrast normalized (MSCN) coefficients, whose statistical properties is changed by distortions. The quantization of these changes is helpful to predict the type of distortion that affects the perceptual quality of image. Moreover, it has been observed that the histogram of MSCN coefficients of a natural image exhibits a Gaussian like appearance [29]. In this paper, we use general gaussian distribution (GGD) and asymmetric GGD to extract parameter features to define quality prediction features. The GGD with zero mean is given by

$$f(n;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|n|}{\beta}\right)^{\alpha}\right), \tag{2}$$

where

$$\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}. \tag{3}$$

$f(n;\alpha,\sigma^2)$ is the normalized number of coefficients, and n is the MSCN coefficient. α and σ^2 reflect the image naturalness, and control the shape and variance of the distribution, respectively. β can be calculated via Eq. (3). $\Gamma(\cdot)$ is a gamma function. In this paper, the GGD model is deployed for MSCN distributions of the left and right views, and the parameters α and σ^2 extracted from two scales are regarded as the global quality-sensitive features f_{g1}.

b: STRUCTURE FEATURES IN GRADIENT DOMAIN

Image structure information is important to NR-SIQA. In this paper, local binary pattern on the gradient map (GLBP) is used as the structural feature. GLBP describes the relationship between pixels in the image neighborhood, and the microstructure pattern of this image can effectively capture the complex degradation caused by various distortions. Different GLBP patterns represent different local gradient patterns. The image distortions may transfer GLBP mode from one type to another, which will change the GLBP mode with their own characteristics [32]. In our study, gradient-weighted histogram of local binary pattern is calculated on the gradient map (GWH-GLBP) of stereopair. The GWH-GLBP is calculated by

$$h_{glbp}(z) = \sum_{i=1}^{N} \omega_i f\left(\mathrm{GLBP}_{P,R}(i), z\right), \tag{4}$$

where

$$f\left(GLBP_{P,R}\left(i\right),z\right)=\begin{cases}1, & \text{GLBP}_{P,R}\left(i\right)=z\\ 0, & \text{otherxise}\end{cases} \tag{5}$$

N denotes the number of image pixels, $z \in [0, Z]$ is the possible GLBP patterns, and ω_i is the weight assigned to the GLBP code. We use the gradient magnitude as the GLBP weight of each pixel in this paper. A locally rotation invariant uniform GLBP operator is defined as

$$GLBP_{P,R}^{\text{riu2}}=\begin{cases}\sum_{i=0}^{P-1}\ s(g_i-g_c), & if\left(u(GLBP_{P,R})\leq 2\right.\\ P+1, & \text{else}\end{cases} \tag{6}$$

with

$$\begin{aligned}u(GLBP_{P,R})=&\left\|s(g_{p-1}-g_c)-s\left(g_0-g_c\right)\right\|\\&+\sum_{i=0}^{p-1}\|s(g_i-g_c)-s(g_{i-1}-g_c)\|\end{aligned} \tag{7}$$

and

$$s(g_i-g_c)=\begin{cases}1, & g_i-g_c\geq 0\\ 0, & g_i-g_c<0,\end{cases} \tag{8}$$

where P is the number of neighbors and R is the radius of the neighborhood. u is the uniform measure, and superscript $riu2$ denotes the rotation invariant ''uniform'' patterns with $u \leq 2$. g_c and g_i are the gradient magnitudes at the center location and its neighbor, respectively. The thresholding function $s(\cdot)$ can be calculated with Eq. (8). The structural features extracted from three scales in gradient domain is taken as another global feature of quality prediction f_{g2}.

2) LOCAL FEATURE EXTRACTION

Image distortion usually affects the local entropy of the image. Since the influence caused by different types and degrees of distortion can be reflected by spatial entropy and spectral entropy [33], we choose the spatial entropy and spectral entropy to represent the local features of the stereopair. The spatial entropy is the probability distribution function of the local pixel value, which reflects the statistical characteristics of the pixel level. The spectral entropy is the probability distribution function of the local DCT coefficient value, which reflects the statistical characteristics of the DCT domain. We divide the stereopair into 8×8 blocks, and the spatial entropy and spectral entropy of each block can be expressed as follows

$$E_s=-\sum_j p(j)\log_2 p(j), \tag{9}$$

where j is the pixel value in a block and $p(j)$ is the probability density correspondingly.

$$E_f=-\sum_x\sum_y c(x,y)\log_2 c(x,y), \tag{10}$$

where $c(x, y)$ are the normalized DCT coefficients in a block.

Then, the mean and skewness of spatial entropy and spectral entropy are calculated in three scales as local features

$$f_l=\left[mean(E_s),\ skew(E_s),\ mean(E_f),\ skew(E_f)\right]. \tag{11}$$

3) SALIENCY-JND FEATURE FUSION MODEL

We have selected the existing models to weight the fusion of the features of the left and right views, which are briefly introduced as follows:

a: SALIENCY MODEL

In this paper, we adopt the relatively simple method proposed in [34] for the visual attention detection. FIGURE 2 (*b*) and (*e*) show the saliency maps obtained by [34]. The calculation of the saliency map S for a given image is described as

$$\begin{cases}A(t)=\Re(F(I(i)))\\ P(t)=\wp(F(I(i)))\\ L(t)=\log(A(t))\\ R(t)=L(t)-h_n(t)\cdot L(t)\\ S(i)=g(i)F^{-1}[\exp(R(t)+P(t))]^2,\end{cases} \tag{12}$$

where $I(i)$ is the given image. F and F^{-1} are the Fourier and Inverse Fourier Transform. $\Re$ is an operator to compute the amplitude information of $A(t)$. $\wp$ is an operator to calculate the phase information of $P(t)$. $L(t)$ is the log spectrum of the given image. $h_n(t)$ is a local average filter to approximate the shape of $A(t)$. $R(t)$ is the spectral residual of the image. $g(i)$ is a Gaussian filter.

b: JND MODEL

We use the JND model proposed in [35]. In the regular pattern, the interaction is relatively simple and the masking effect is limited. While in the irregular pattern, the interaction is complex and the masking effect is strong. FIGURE 2 (*c*) and (*f*) show the JND maps obtained by [35]. The direction presented in each pixel is taken as the basic element of the pattern, and the complexity of the pattern is calculated as the diversity of the direction of the local area. Finally, based on the model complexity and brightness contrast, the JND estimation model is obtained as

$$J(I(i))=L_A(I(i))+M_S(I(i))-C_1\min\left\{L_A(I(i)),M_S(I(i))\right\}, \tag{13}$$

where $L_A(I(i))$ is the luminance adaptation. $M_S(I(i))$ is the total spatial masking effect, which is defined as the maximum of contrast masking and pattern masking. C_1 is the gain reduction parameter determined by the overlapping between $L_A(I(i))$ and $M_S(I(i))$, and we set $C_1 = 0.3$ (the same as in [35]).

We use the saliency model and JND model for feature fusion to simulate the binocular rivalry. More specifically, we take full advantage of the saliency model and JND model to weight for each view's features. Since the features are extracted from multiple scales, the feature fusion is on multiple scales. And the saliency and JND maps are calculated

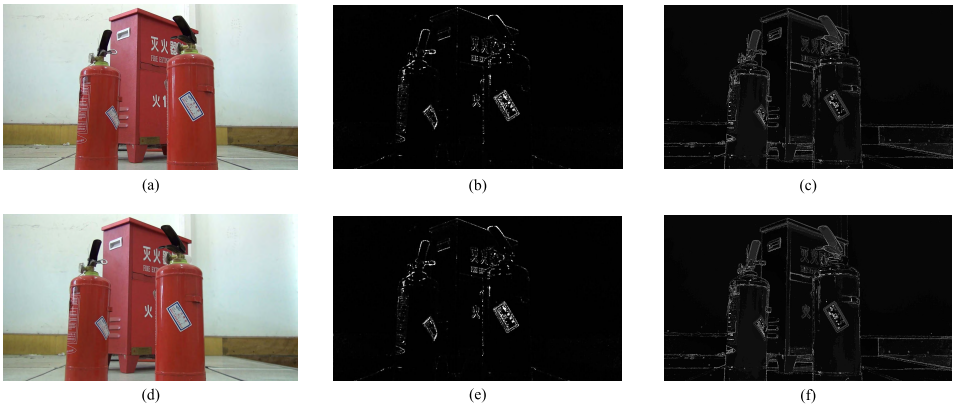


FIGURE 2. Saliency map and JND map of the left and right images. (a) Left image. (b) Saliency map of left image. (c) JND map of left image. (d) Right image. (e) Saliency map of right image. (f) JND map of right image.

FIGURE 3. Cyclopean map synthesized by left and right images. (a) Left image. (b) Right image. (c) Cyclopean map.

on multiple scales too. Therefore, we adopt f_g^m, f_l^m and f^m to represent the fused global, local and the final features separately. $f_{g\_L}^m$ and $f_{g\_R}^m$ denote the global features of the left and right view extracted from Eq. (2). $f_{l\_L}^m$ and $f_{l\_R}^m$ denote the local features of the left and right view extracted from Eq. (11). As for the calculation of weighting factors of saliency model and JND model, we use S_L^m and S_R^m to denote the sum of all elements in the left and right saliency maps, respectively. Similarly, J_L^m and J_R^m denote the sum of all elements in the left and right JND maps. Weighting factor ω can be calculated by Eq. (16)

$$f_g^m = \omega_1 \cdot \omega_2 \cdot f_{g\_L}^m + (1 - \omega_1 \cdot \omega_2) f_{g\_R}^m \tag{14}$$

$$f_l^m = \omega_1 \cdot \omega_2 \cdot f_{l\_L}^m + (1 - \omega_1 \cdot \omega_2) f_{l\_R}^m \tag{15}$$

$$\omega_1 = \frac{S_L^m}{S_L^m + S_R^m}, \quad \omega_2 = 1 - \frac{J_L^m}{J_L^m + J_R^m}. \tag{16}$$

Besides, in the local feature extraction, we also have JND weighted for each block. We choose the maximum value of JND on each block to normalize the JND values of this block. Finally, the normalized JND block is multiplied by the features extracted on each block to obtain the weighted local features. We use f_w to denote the weighted features, it can be defined as

$$f_w = [f_g^m, f_l^m]. \tag{17}$$

B. BINOCULAR CYCLOPEAN MAP SYNTHESIS AND FEATURE EXTRACTION

The difference between stereoscopic images and ordinary 2D images lies in binocular fusion. And the cyclopean map can represent the perception fusion of 3D scene. Besides, the spatial activity not only helps to determine whether there is binocular competition between the left and right views, but also it can be used to further assess the relative extent of the two-view binocular competition. Therefore, we use the binocular spatial activity model proposed in [36] to synthesize a cyclopean map

$$Cyc(i) = \frac{\{\varepsilon\,[S_L(i)]+C_2\} \cdot I_L(i) + \{\varepsilon\,[S_R(i+d_i)]+C_2\} \cdot I_R(i+d_i)}{\{\varepsilon\,[S_L(i)]+C_2\} + \{\varepsilon\,[S_R(i+d_i)]+C_2\}} \tag{18}$$

with

$$\varepsilon\,[S(i)] = \log_2\left[\sigma^2(i) + 1\right], \tag{19}$$

where $S(i)$ is a neighborhood centered at pixel point i. d_i is the disparity and $C_2 = 0.01$ is a small positive number to guarantee stability. The obtained cyclopean map is shown in FIGURE 3.

Cognitive neuroscience studies have shown that the arrangement of excitatory and inhibitory cortical cells produces orientation selectivity in the local receptor region, where the primary visual cortex extracts visual information for scene understanding [37], [38]. HVS is highly adaptable to the extracted scene-aware structural information. Accordingly, orientation selectivity based structure description is widely used in image quality assessment [39], [40], which can effectively extract the visual information of the image and reflect the image degradation caused by different types of distortions. To extract more spatial structure features from

cyclopean map, we adopt the structure descriptor based on gradient magnitude and orientation selectivity proposed in [41] to simulate the arrangement of excitatory/suppressor cells in the local receptive field. And we introduce a pattern based on direction selection to represent the spatial correlation of the structure. Finally, a structural histogram based on orientation selectivity is established as the quality-sensitive features to represent the content information of the cyclopean map.

For a given cyclopean map, the local structure intensity of each pixel can be demanded as its luminance change, and it is calculated as

$$M(i) = \sqrt{(G_h(i))^2 + (G_v(i))^2}, \tag{20}$$

where $G_h(i)$ and $G_v(i)$ are the gradient magnitudes alone the horizontal and vertical directions.

Then the response pattern $P(i)$ of a pixel is described as the arrangement of interactions between the central pixel i and its local neighbors $R(i)$ ($R(i) = \{i_1, i_2, ..., i_n\}$).

$$P(i) = A(I(i|i_1), I(i|i_2), ..., I(i|i_n)), \tag{21}$$

where

$$I(i|i_N) = \begin{cases} 1, & \text{if } |\theta(i) - \theta(i_N)| < T \\ 0, & \text{else,} \end{cases} \tag{22}$$

and

$$\theta(i) = \arctan \frac{G_v(i)}{G_h(i)}. \tag{23}$$

$I(i|i_N)$ is the interaction type between two pixels. The parameter T judges the interaction type, and in this work we set $T = 6$.

Finally, the quality sensitive feature f_c of the cyclopean map can be mapped into a structure based histogram

$$H_w(z) = \sum_{x=1}^{N} M(i)\delta(P_f(i), P_f^z) \tag{24}$$

and

$$\delta(P_f(i), P_f^z) = \begin{cases} 1 & \text{if } P_f(i) = P_f^z \\ 0 & \text{else,} \end{cases} \tag{25}$$

where P_f^z represents the z^{th} fundamental pattern form that pixel i belongs to.

Combined with the weighted features f_w, the overall quality sensitive feature f is obtained as

$$f = [f_w, f_c]. \tag{26}$$

C. SVR-BASED REGRESSION MODEL

In order to utilize the obtained quality sensitive features for NR-SIQA, in our experiment, SVR is adopted for learning in multiple datasets [42]–[47]. And it has been widely applied to image quality assessment. Based on the radial basis function (RBF) kernel, we implement SVR regression using a library for support vector machines (LIBSVM) package. By feeding the consolidated features into the trained SVR model with the subjective ratings (e.g., mean opinion score (MOS) or differential mean opinion score (DMOS)), the process of NR-SIQA is realized.

III. EXPERIMENTAL EVALUATION

A. SIQA DATABASE

In this section, to evaluate the performance of the proposed NR-SIQA metric, four datasets are used in our implement.

The LIVE 3D IQA Phase I (LIVE-I 3D) Dataset [34] contains 20 reference stereopairs and 365 symmetric distorted images. It consists of five distortion types, including JPEG, JPEG2000 (JP2K), WN, Gaussian blur (GBlur), and fast fading (FF). DMOS values are provided as subjective quality scores on the distorted ones.

The LIVE 3D IQA Phase II (LIVE-II 3D) Dataset [9] is more complex with both symmetric and asymmetric distorted images. It contains 8 reference stereopairs and 360 distorted ones. Among them, 120 pairs are symmetric distortions and 240 pairs are asymmetric distortions. The distortion types are the same as the LIVE 3D IQA Phase I. The DMOS is also provided for distorted ones.

The Waterloo-IVC 3D IQA Phase I (WIVC-I 3D) Dataset [48] consists of 6 reference stereopairs and 330 symmetrically and asymmetrically distorted images. The types of distortions are JPEG, WN and GBlur. MOS values are presented for subjective quality scores.

The Waterloo-IVC 3D IQA Phase II (WIVC-II 3D) Dataset [48] consists of 10 reference stereopairs and 460 symmetrically and asymmetrically distorted images. The types of distortions are the same as the WIVC-I 3D dataset. Note that it is the first dataset that contains mixed distortion types in asymmetrically distorted images. MOS values are also provided for distorted ones.

B. PERFORMANCE INDICATORS

In our implement, after the nonlinear regression with SVR, we compute Pearson linear correlation coefficients (PLCC) and Spearman rank-order correlation coefficients (SROCC) between subjective scores and predicted scores. PLCC and SROCC can evaluate prediction performance monotonicity and consistency respectively. In general, to get a good metric, the value between the subjective scores and predicted scores should be close to 1 for PLCC and SROCC. To reduce the nonlinearity in the regression prediction, a five-parameter logistic regression function is applied before the calculation of PLCC, which is defined as

$$s(p) = \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{\exp(\beta_2 \cdot (p - \beta_3)) + 1}\right) + \beta_4 \cdot p + \beta_5. \tag{27}$$

$s(p)$ is the fitted quality score, p is the objectively predicted score, and β_i ($i = 1, 2, 3, 4, 5$) denotes the parameter to be fitted.

In the training-testing procedure, we adopt five-fold cross-validation, and 80% of the dataset is for training and 20% is for testing. It is repeated 1000 times to eliminate the bias for training sets selection, and we take the mean value for the final result. In the performance test, the performance evaluation is carried out under five different distortions respectively, as JP2K, JPEG, WN, GBlur and FF in TABLE 1-7. And it is also carried out under the combination of all the distortions mentioned above, which is denoted as ALL in TABLE 1-7. We take the ALL performance as the main reference performance index, because in practice, there are cases where multiple distortions coexist.

TABLE 1. Comparison of fusion ways on LIVE-I 3D dataset.

		M_1	M_2	M_3	M_4
PLCC	JP2K	0.9517	0.9582	0.9570	0.9568
	JPEG	0.7991	0.8090	0.8096	0.8118
	WN	0.9581	0.9621	0.9594	0.9577
	GBlur	0.9541	0.9510	0.9547	0.9484
	FF	0.8574	0.8500	0.8628	0.8461
	ALL	0.9595	0.9597	0.9614	**0.9648**
SROCC	JP2K	0.9004	0.9095	0.9112	0.9102
	JPEG	0.7489	0.7520	0.7562	0.7604
	WN	0.9306	0.9355	0.9320	0.9299
	GBlur	0.8672	0.8627	0.8700	0.8642
	FF	0.7853	0.7887	0.8022	0.7894
	ALL	0.9504	0.9494	0.9528	**0.9531**

TABLE 2. Comparison between different saliency models on NR-SIQA.

		the GBVS method	the spectral residual method
PLCC	JP2K	0.9544	0.9568
	JPEG	0.7705	0.8118
	WN	0.9643	0.9577
	GBlur	0.9507	0.9484
	FF	0.8462	0.8461
	ALL	0.9608	**0.9648**
SROCC	JP2K	0.9102	0.9102
	JPEG	0.7090	0.7604
	WN	0.9378	0.9299
	GBlur	0.8775	0.8642
	FF	0.8005	0.7894
	ALL	0.9514	**0.9531**

TABLE 3. Effect of JND model on NR-SIQA.

		without JND model	with JND model
PLCC	JP2K	0.9517	0.9568
	JPEG	0.7991	0.8118
	WN	0.9561	0.9577
	GBlur	0.9421	0.9484
	FF	0.8373	0.8461
	ALL	0.9595	**0.9648**
SROCC	JP2K	0.9004	0.9102
	JPEG	0.7489	0.7604
	WN	0.9293	0.9299
	GBlur	0.8625	0.8642
	FF	0.7853	0.7894
	ALL	0.9504	**0.9531**

TABLE 4. Effect of cyclopean map on NR-SIQA.

		without cyclopean map	with cyclopean map
PLCC	JP2K	0.9547	0.9568
	JPEG	0.8252	0.8118
	WN	0.9569	0.9577
	GBlur	0.9424	0.9484
	FF	0.8303	0.8461
	ALL	0.9605	**0.9648**
SROCC	JP2K	0.9113	0.9102
	JPEG	0.7741	0.7604
	WN	0.9297	0.9299
	GBlur	0.8599	0.8642
	FF	0.7750	0.7894
	ALL	0.9518	**0.9531**

TABLE 5. Comparison between different features on NR-SIQA.

		f_{g1}	f_{g2}	f_g	f_l	$f_{g1}+f_l$	$f_{g2}+f_l$	f_g+f_l
PLCC	JP2K	0.9342	0.9331	0.9465	0.9326	0.9498	0.9475	0.9568
	JPEG	0.7634	0.7889	0.7926	0.6501	0.6849	0.8028	0.8118
	WN	0.9451	0.9465	0.9530	0.9362	0.9416	0.9453	0.9577
	GBlur	0.9344	0.9531	0.9404	0.9559	0.9421	0.9435	0.9484
	FF	0.7959	0.8230	0.8264	0.8419	0.8228	0.8363	0.8461
	ALL	0.9396	0.9537	0.9559	0.9393	0.9496	0.9588	**0.9648**
SROCC	JP2K	0.8671	0.8804	0.8900	0.8738	0.9090	0.8950	0.9102
	JPEG	0.7009	0.7458	0.7482	0.5605	0.6096	0.7387	0.7604
	WN	0.9052	0.9202	0.9213	0.9097	0.9010	0.9183	0.9299
	GBlur	0.8530	0.8633	0.8599	0.8992	0.8602	0.8511	0.8642
	FF	0.6997	0.7369	0.7534	0.7752	0.7608	0.7714	0.7894
	ALL	0.9264	0.9473	0.9479	0.9211	0.9323	0.9510	**0.9531**

C. EXPERIMENTS ON LIVE-I 3D DATASET

1) COMPARISON OF FUSION WAYS ON LIVE-I 3D DATASET

We compare the performance of the proposed method with four different fusion ways on LIVE-I 3D dataset. These four fusion ways are using saliency model to weight the left and right views (M_1), using JND model to weight the left and right views (M_2), combining the saliency model with JND model to weight the left and right views (M_3), and adding JND weighting to the local blocks based on M_3 (M_4). The results are shown in TABLE 1. From TABLE 1 we can conclude that the M_4 method shows the highest performance. Hence, we use M_4 in the following experiments.

2) COMPARISON BETWEEN DIFFERENT SALIENCY MODELS ON NR-SIQA

We select the spectral residual method from the computational model and the GBVS [49] method from the eye fixation model to obtain saliency map, and adopt the two models in the proposed method respectively for the performance comparison. TABLE 2 shows the comparison between the two saliency models on LIVE-I 3D dataset. By analyzing TABLE 2, we can see that the spectral residual model shows higher performance in our proposed method. Therefore, we use the spectral residual model in the following experiments.

3) EFFECT OF JND MODEL ON NR-SIQA

In our method, both saliency model and JND model are used. Therefore, in order to verify the effect of the JND model, we compare the performance of the methods with and without JND model on LIVE-I 3D dataset. The comparison results are listed in TABLE 3. We can conclude that the combination of saliency model and JND model can evaluate the quality of the stereopairs better.

TABLE 6. Performance on LIVE-I 3D dataset.

		Yang [15]	Yue [43]	Xu [44]	Ding [45]	Zhou [50]	Our Method
PLCC	JP2K	0.9112	0.9340	0.9510	0.9665	0.9848	0.9568
	JPEG	0.7670	0.7440	0.7380	0.9045	0.6260	0.8118
	WN	0.8899	0.9190	0.9480	0.9831	0.9250	0.9577
	GBlur	0.8960	0.9710	0.9660	0.9626	0.8990	0.9484
	FF	0.8083	0.8540	0.8480	0.9595	0.7070	0.8461
	ALL	0.9379	0.9373	0.9491	0.9401	0.9412	**0.9648**
SROCC	JP2K	0.9448	0.8320	0.9030	0.9636	0.8370	0.9102
	JPEG	0.7896	0.5950	0.6780	0.9102	0.6380	0.7604
	WN	0.9129	0.9320	0.9050	0.9830	0.9310	0.9300
	GBlur	0.7552	0.8570	0.9070	0.9458	0.8330	0.8642
	FF	0.8514	0.7790	0.8000	0.9520	0.6490	0.7894
	ALL	0.9363	0.9140	0.9342	0.9423	0.9215	**0.9531**

TABLE 7. Performance on LIVE-II 3D dataset.

		Yang [15]	Yue [43]	Xu [44]	Ding [45]	Zhou [50]	Our Method
PLCC	JP2K	0.8871	0.9860	0.9000	0.9805	0.6340	0.9504
	JPEG	0.8517	0.8430	0.8150	0.9363	0.6470	0.8789
	WN	0.8930	0.9860	0.9720	0.9881	0.9040	0.9744
	GBlur	0.8609	0.9730	0.9820	0.9861	0.9670	0.9905
	FF	0.9108	0.9230	0.8910	0.9734	0.8510	0.9561
	ALL	0.9145	0.9141	0.9265	0.9297	0.9236	**0.9549**
SROCC	JP2K	0.9116	0.9590	0.8610	0.9771	0.5530	0.9183
	JPEG	0.8989	0.7690	0.7710	0.9339	0.5930	0.8340
	WN	0.9088	0.9590	0.9360	0.9879	0.8930	0.9498
	GBlur	0.9333	0.8680	0.9220	0.9730	0.8690	0.9507
	FF	0.9378	0.9130	0.8580	0.9746	0.8280	0.9293
	ALL	0.9042	0.9063	0.9103	0.9238	0.9194	**0.9459**

4) EFFECT OF CYCLOPEAN MAP ON NR-SIQA

In order to verify the impact of the cyclopean map on the quality evaluation, we compare the performance of the methods with and without cyclopean map on LIVE-I 3D dataset. The comparison results are shown in TABLE 4. It can be seen from the table that the features of the cyclopean map are helpful to improve the performance of the proposed NR-SIQA method.

5) COMPARISON BETWEEN DIFFERENT FEATURES ON NR-SIQA

Since many features are used in our method, in order to verify the validity of the selected features, we compare the performance of the proposed method with different features in TABLE 5. These different features are f_{g1}, f_{g2}, f_g, f_l, $f_{g1} + f_l$, $f_{g2} + f_l$, and $f_g + f_l$. Among them, f_{g1} denotes the natural statistical features, f_{g2} is the gradient features, f_{g1} and f_{g2} are combined as global features f_g, and f_l denotes the local entropy features. The experimental results are shown in TABLE 5. It can be seen from TABLE 5 that when both global features are used and combined with local features for feature extraction, the accuracy of quality prediction is the highest, which further verifies the reliability of our method.

D. COMPARISON WITH OTHER NR-SIQA METHODS ON LIVE AND WIVC 3D DATASETS

1) PERFORMANCE ON LIVE 3D DATASETS

The performance of the proposed method on LIVE 3D datasets is evaluated by comparing with other state-of-the-art NR-SIQA methods including Yang *et al.* [15], Yue *et al.* [43], Xu *et al.* [44], Ding *et al.* [45], and Zhou *et al.* [50]. The comparisons on LIVE-I 3D and LIVE-II 3D dataset are presented in TABLE 6 and TABLE 7, respectively. It can be observed that on almost all of the distortions, the proposed method is superior to most of the compared NR-SIQA methods, except Ding *et al.* [45], which is a deep-learning-based NR-SIQA method. As we all know, PLCC and SROCC represent the consistency between the predicted and actual quality. In Table 6 and 7, the performance of the proposed method on the combination of multiple distortions is better than that of Ding *et al.* [45]. It means that the quality assessment obtained by the proposed method is more consistent with the ground-truth in this case. In addition, Yang *et al.* [15] and Yue *et al.* [43] perform the feature fusion by superimposing the monocular and binocular features, and Xu *et al.* [44] and Ding *et al.* [45] perform feature fusion by using saliency model. They do not consider the combination of saliency model and JND model, while we consider the combination of these two models, which can reflect the effect of visual attention and perception on quality assessment. Hence it is reasonable that the proposed method shows the best performance on the combination of multiple distortions, which is denoted as ALL in TABLE 6 and TABLE 7.

To further demonstrate the effectiveness of the proposed method, scatter diagrams of subjective scores (DMOS) and objective scores (predicted scores) on LIVE-I 3D and LIVE-II 3D datasets are given in FIGURE 4. In FIGURE 4 (*a*) and (*b*), each circle represents an image, and the distortion type is a circle marked with different colors. Besides, we fit the LIVE-II 3D dataset according to asymmetric and symmetric distortions, as shown in FIGURE 4 (*c*). We can observe that the circles closely gathered around the fitted curve, which means that the objective scores are consistent with the subjective scores showing consistency of the proposed method with human perception evaluation. Meanwhile, we calculate the root mean square error (RMSE) to evaluate the prediction error. In our method, the ALL performance of RMSE of the LIVE-I 3D dataset is 4.9306. In the LIVE-II 3D dataset, the ALL performance of RMSE is 3.4469, where the RMSE value of asymmetric distortion is 3.5929 and symmetric distortion is 4.2390.

2) PERFORMANCE ON WIVC 3D DATASETS

We also test the proposed method on WIVC 3D datasets and compare the proposed method with other methods including BLIINDS-II [51], Yue *et al.* [43], NR-DIIVINE [46], Chen *et al.* [9], and DECOSINE [47]. And the performance comparison is shown in TABLE 8. We can conclude that the proposed method not only performs well on WIVC-I 3D dataset compared with other NR-SIQA metrics, but also it is superior to all other methods on WIVC-II 3D dataset. It indicates that the proposed method can also deal with the mixed distortion types in asymmetrically distorted images.

3) CROSS-DATASET VALIDATION

In order to verify the generalization capability of the proposed method, cross-dataset tests are carried out on LIVE 3D and

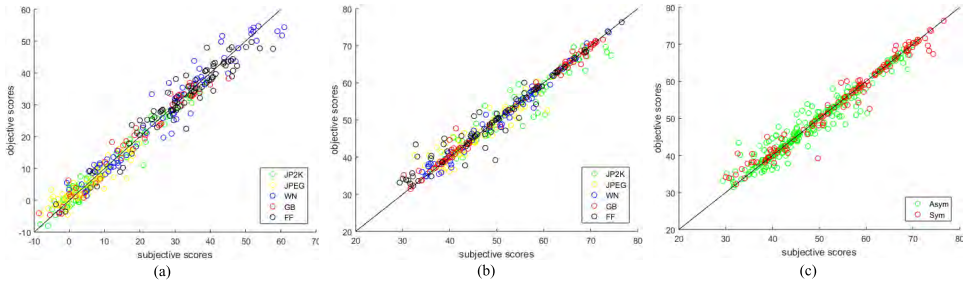


FIGURE 4. Scatter diagrams of subjective and objective scores on LIVE 3D datasets.

TABLE 8. Performance on WIVC 3D datasets.

Metrics	WIVC-I		WIVC-II	
	PLCC	SROCC	PLCC	SROCC
BLIINDS-II [51]	0.8765	0.8490	0.8341	0.7965
Yue [43]	0.9262	0.9192	0.9113	0.8951
NR-DIIVINE [46]	0.9265	0.9003	0.9098	0.8872
Chen [9]	0.9380	0.9220	0.8820	0.8840
DECOSINE [47]	0.9439	0.9246	0.9331	0.9143
Our Method	**0.9494**	**0.9373**	**0.9597**	**0.9519**

TABLE 9. Cross-dataset validation results.

	LIVE-I/LIVE-II	LIVE-II/LIVE-I	WIVC-I/WIVC-II	WIVC-II/WIVC-I
PLCC	0.8263	0.8605	0.8234	0.8602
SROCC	0.8175	0.8524	0.8119	0.8368

WIVC 3D datasets, and the results are shown in TABLE 9. LIVE-I 3D and LIVE-II 3D dataset provide DMOS values for subjective quality scores, while WIVC-I 3D and WIVC-II 3D provide MOS values. Therefore, cross-dataset experiments between LIVE 3D and WIVC 3D datasets are not appropriate because DMOS and MOS values are generated by different processes. Based on this, we conducted four experiments: i) the methods are trained on LIVE-I 3D dataset and tested on LIVE-II 3D dataset (LIVE-I/LIVE-II), ii) LIVE-II/LIVE-I, iii) WIVC-I/WIVC-II and iv) WIVC-II/WIVC-I. The prediction performance is degraded because there are image and distortion types in the test set that are not in the training set. Compared with the LIVE-I 3D dataset, LIVE-II 3D dataset contains asymmetric distortion images. And the difference between WIVC-I 3D and WIVC-II 3D datasets are that the asymmetric distortion image of WIVC-II 3D contains mixed distortion, that is, multiple distortions in one image. In addition, the reference images in four 3D datasets are different. Though the performance is degraded due to the differences in dataset and distortions between the training and testing sets, it still can reach a comparable level that of the BLINDS-II method in [51] when it is trained and tested in the same dataset.

IV. CONCLUSION

In this paper, we propose a NR-SIQA method based on the saliency model and JND model. Our method considers not only visual attention and visual perception, but also the combination of monocular features and binocular features. From our method, we can conclude that the combination of saliency model and JND model is better than only one of them. When feature extraction is performed, global features need to be combined with local features to improve the accuracy of the assessment. Finally, cyclopean map can well embody the visual scene perceived by the HVS. Experiments are conducted on four SIQA datasets and the results verify better and more reliable performance in comparison with other NR-SIQA methods.

Although the proposed method shows better performance, there are still problems need further consideration. The next step is to study the depth information related to the stereopair and design a more effective evaluation method for NR-SIQA. In addition, the proposed method is time consuming and needs to be further optimized.

REFERENCES

[1] S. K. Md, B. Appina, and S. S. Channappayya, "Full-reference stereo image quality assessment using natural stereo scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1985–1989, Nov. 2015.
[2] S. Ryu, D. H. Kim, and K. Sohn, "Stereoscopic image quality metric based on binocular perception model," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2013, pp. 609–612.
[3] X. Wang, S. Kwong, Y. Zhang, and Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. Vis. Commun. Image Process.*, Nov. 2011, pp. 1–4.
[4] Y. Zhang and D. M. Chandler, "3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3810–3825, Nov. 2015.
[5] C. T. E. R. Hewage and M. G. Martini, "Reduced-reference quality metric for 3D depth map transmission," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video*, Jun. 2010, pp. 1–4.
[6] Q. Xu, G. Zhai, M. Liu, and K. Gu, "Using structural degradation and parallax for reduced-reference quality assessment of 3D images," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2014, pp. 1–6.
[7] E. Yu, J. Sun, J. Li, X. Chang, X. Han, and A. Hauptmann, "Adaptive semi-supervised feature selection for cross-modal retrieval," *IEEE Trans. Multimedia*, to be published.
[8] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "Reduced-reference stereoscopic image quality assessment based on view and disparity zero-watermarks," *Signal Process., Image Commun.*, vol. 29, no. 1, pp. 167–176, 2014.
[9] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.
[10] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, "Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2059–2074, Mar. 2016.

[11] W. Wan *et al.*, "Pattern complexity-based jnd estimation for quantization watermarking," *Pattern Recognit. Lett.*, to be published. doi: 10.1016/j.patrec.2018.08.009.
[12] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685–1699, May 2015.
[13] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.
[14] T.-J. Liu, C.-T. Lin, H.-H. Liu, and S.-C. Pei, "Blind stereoscopic image quality assessment based on hierarchical learning," *IEEE Access*, vol. 7, pp. 8058–8069, 2019.
[15] J. Yang, P. An, J. Ma, K. Li, and L. Shen, "No-reference stereo image quality assessment by learning gradient dictionary-based color visual characteristics," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.
[16] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.
[17] W. Tian, F. Shao, G. Jiang, and M. Yu, "Blind image quality assessment for stereoscopic images via deep learning," *J. Comput.-Aided Des. Comput. Graph.*, vol. 28, no. 6, pp. 968–975, 2016.
[18] K. Herrmann, D. J. Heeger, and M. Carrasco, "Feature-based attention enhances performance by increasing response gain," *Vis. Res.*, vol. 74, pp. 10–20, Dec. 2012.
[19] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Ann. Rev. Neurosci.*, vol. 18, no. 1, pp. 193–222, 1995.
[20] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.
[21] Y. Liu, J. Yang, Q. Meng, Z. Lv, Z. Song, and Z. Gao, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Process.*, vol. 125, pp. 237–248, Aug. 2016
[22] J. Yang *et al.*, "Quality assessment metric of stereo images considering cyclopean integration and visual saliency," *Inf. Sci.*, vol. 373, pp. 251–268, Dec. 2016.
[23] X. Wang, L. Ma, S. Kwong, and Y. Zhou, "Quaternion representation based visual saliency for stereoscopic image quality assessment," *Signal Process.*, vol. 145, pp. 202–213, Apr. 2018.
[24] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.
[25] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.
[26] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.
[27] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Stereoscopic image quality metric based on local entropy and binocular just noticeable difference," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2002–2006.
[28] Y. Fan, M.-C. Larabi, F. A. Cheikh, and C. Fernandez-Maloigne, "Stereoscopic image quality assessment based on the binocular properties of the human visual system," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2037–2041.
[29] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[30] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.
[31] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[32] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
[33] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.
[34] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[35] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, and C.-C. J. Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2682–2693, Jun. 2017.
[36] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Process., Image Commun.*, vol. 58, pp. 287–299, Aug. 2017.
[37] R. Ben Yishai, R. L. Bar-Or, and H. Sompolinsky, "Theory of orientation tuning in visual cortex," *Proc. Nat. Acad. Sci. USA*, vol. 92, no. 9, pp. 3844–3848, 1995.
[38] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.
[39] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
[40] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–160, Mar. 2005.
[41] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
[42] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.
[43] G. Yue, C. Hou, Q. Jiang, and Y. Yang, "Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry," *Signal Process.*, vol. 150, pp. 204–214, Sep. 2018.
[44] X. Xu, Y. Zhao, and Y. Ding, "No-reference stereoscopic image quality assessment based on saliency-guided binocular feature consolidation," *Electron. Lett.*, vol. 53, no. 22, pp. 1468–1470, Oct. 2017.
[45] Y. Ding *et al.*, "No-reference stereoscopic image quality assessment using convolutional neural network for adaptive feature extraction," *IEEE Access*, vol. 6, pp. 37595–37603, 2018.
[46] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[47] J. Yang, K. Sim, X. Gao, W. Lu, Q. Meng, and B. Li, "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1314–1328, Mar. 2019.
[48] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.
[49] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 545–552.
[50] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognit.*, vol. 71, pp. 207–217, Nov. 2017.
[51] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

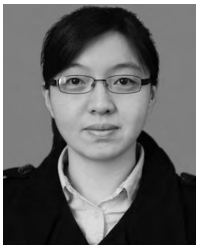
YAFEI LI received the bachelor's degree in communication engineering from the School of Information Science and Engineering, Shandong Normal University, Jinan, in 2017. She is currently pursuing the master's degree in communication and information systems with Shandong Normal University. Her research interests include stereoscopic image quality assessment and machine learning. She is a Student Member of the CCF.

FENG YANG is currently a Professor with the School of Information Science and Technology, Shandong Normal University. He has published more than 30 research papers. He holds seven granted invention patents. His research interests include digital image processing, information security, and applied electronics.

WENBO WAN received the Ph.D. degree from Shandong University, Jinan, China, in 2015. He is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University. His research interests include image/video processing and image/video watermarking.

JUN WANG received the bachelor's degree from the School of Physics and Electronics, Shandong Normal University, Jinan, Shandong, in 2017. He is currently pursuing the master's degree in communication and information systems with Shandong Normal University. His research interests include digital watermarking technology and deep learning.

MIN GAO received the bachelor's degree in communication engineering from the School of Information Science and Engineering, Shandong Normal University, Jinan, in 2016. She is currently pursuing the master's degree in communication and information systems with Shandong Normal University. Her research interests include computer vision, machine learning, and signal processing. She is a Student Member of the CCF.



JIA ZHANG received the B.E. and M.E. degrees from the School of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, in 2006 and 2009, respectively, and the Ph.D. degree in communication and information systems from the School of Information Science and Engineering, Shandong University, China, in 2013. She is currently a Lecturer with Shandong Normal University, Jinan, China. Her research interests include 5G radio techniques, joint resource allocation and optimization in multi-cell networks, dynamic programming, and interference coordination in heterogeneous cellular networks.

JIANDE SUN received the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2000 and 2005, respectively. From 2008 to 2009, he was a Visiting Researcher with the Institute of Telecommunications System, Technical University of Berlin, Berlin, Germany. From 2010 to 2012, he was a Postdoctoral Researcher with the Institute of Digital Media, Peking University, Beijing, China, and also with the State Key Laboratory of Digital-Media Technology, Hisense Group. From 2014 to 2015, he was a DAAD Visiting Researcher with the Technical University of Berlin and the University of Konstanz, Germany. From 2015 to 2016, he was a Visiting Researcher with the School of Computer Science, Language Technology Institute, Carnegie Mellon University, USA. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University. He has published more than 60 journal and conference papers. He is the coauthor of two books. His current research interests include multimedia content analysis, video hashing, gaze tracking, image/video watermarking, and 2D-to-3D conversion.