IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 August 2023, accepted 30 August 2023, date of publication 5 September 2023, date of current version 11 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312126

RESEARCH ARTICLE

Blind Quality Assessment of Stereoscopic Images Considering Binocular Perception Based on Shearlet Decomposition

DONGHUI WAN[1,2], XIUHUA JIANG[1,3], AND QING SHEN[4]

[1]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China
[2]School of Electronic Information, Huzhou College, Huzhou 313000, China
[3]Peng Cheng Laboratory, Shenzhen 518000, China
[4]School of Information Engineering, Huzhou University, Huzhou 313000, China

Corresponding author: Donghui Wan (wandonghui@zjhzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61802123; and in part by the Primary Research and Development Plan of Zhejiang Province, China, under Grant 2020C01097.

ABSTRACT Due to the deficient knowledge of binocular vision properties, how to effectively evaluate stereoscopic images still remains a challenging task. Inspired by multichannel processing of human visual system (HVS), we propose a blind method for stereoscopic image quality assessment (SIQA) by extracting quality related features in sub-bands of the image. First of all, we introduce the shearlet transform to decompose the left- and right-view images into multiple sub-bands content with diverse combinations of scales and orientations, and obtain the combined view based on energy-weighted summation of the corresponding sub-bands of two eye views. Then, natural scene statistics (NSS) of the original left and right images are obtained as quality-sensitive features, followed by extracting NSS features of the sub-bands of left, right and combined views. Moreover, we calculate the gradient similarity between each sub-band pair to denote the asymmetric distortion and disparity information. Finally, all the extracted features are mapped into a quality score by support vector regression (SVR). experimental results on multiple benchmark databases verify the superiority of our method.

INDEX TERMS Blind quality assessment, human visual system, natural scene statistics, shearlet transform, stereoscopic image.

I. INTRODUCTION

In recent decades, the technology of stereoscopic three dimensions (3D) imaging has gained tremendous attention due to its ability of providing more immersive viewing experience compared to the 2D counterpart. The stereoscopic method captures two slightly different images which are set to be respectively viewed by the left and right eyes at the same time. As a result of the viewing disparity from two eyes, we may perceive 3D depth by binocular fusion. Along with the development of 3D image applications, stereoscopic image quality assessment (SIQA), which is extremely critical for quality optimization related to 3D image services, has attracted widespread research interest. Compared to its 2D counterpart, the task of SIQA is more challenging according to the comprehensive viewing integration of an image pair from the left and right eyes simultaneously [1], [2].

Similar to 2D image quality assessment [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], objective methods of SIQA can be classified as three categories i.e., full-reference (FR), reduced-reference (RR), and no-reference (NR). The FR model makes quality assessment by accessing the full original high-quality reference image pair for comparison. The RR algorithm computes the score with limited access to the reference image pair, in order to reducing the burden of saving or transmitting the redundant information. And the NR method, which is most challenging, just evaluates the tested image pair without any access to

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

original pristine information. As we know, high-quality reference stereoscopic content is usually unavailable in practical terms, hence the NR method is a much more competitive way for application.

Stereoscopic image distortion may come from camera artifacts, coding compression, channel errors and noise, etc., and a stereoscopic pair may be symmetrically or asymmetrically distorted. From previous research, the mainly distinctive inferior Quality of Experience (QoE) caused by stereoscopic image corruption may be binocular rivalry, wrong depth sensation and visual discomfort and fatigue, which are direct combining results on human visual system (HVS) from a different-perspective image pair [20], [21], [22]. Some SIQA algorithms utilize traditional 2D IQA methods to evaluate the image quality of left and right views separately, and then the two rating values are weighted averaged for the final stereoscopic image quality score. This type of method has a clear idea and is easy to implement, but the performance is barely satisfactory. To achieve effective results, articles [28], [29] demonstrate that disparity information, the origination of 3D depth, should be further considered. Finally, some other kinds of methods fuse the image pair to obtain a cyclopean image by simulating HVS integration, and use 2D IQA of the cyclopean image as ultimate result [30], [31]. However, there lies huge difficulty of how to combine binocular perception with the characteristics of stereoscopic images owing to the complex visual mechanism.

From above, it is known that accurately evaluating 3D image quality arising from binocular perception poses new challenges for the community. To correlate SIQA highly with human subjective evaluation, it makes sense of building computing models based on well understanding the visual processing [23], [24], [25]. Binocular image information is assumed to be processed through two visual routes, dorsal pathway and ventral pathway, from low-level to high-level areas in human visual cortex [26]. Furthermore, physiological studies have shown that human visual cortex has different sensitivities to different stimulus frequency, and these different-frequency stimuli are processed in different channels of the visual system to achieve the best visual effect through interaction [52]. To represent the multi-frequency-channel characteristic of visual procedure, the image wavelet transform is recognized as a feasible way [53]. However, the classical wavelet transform is not effective for image singularity detection due to the scarce direction representation. Hence, some directional wavelet based methods, such as curvelet transform [54], contourlet transform [55], shearlet transform [56], [57], [58], [59], etc., have been proposed for performance enhancement. mathematically, the shearlet transform has advantages in better direction selectivity and smaller compact support compared to the other directional wavelet based algorithms. As there are two images viewed the same time, it is significant to consider their integrated perception. Among numerous binocular combination methods, energy-based Gain-Control model is commonly utilized to accurately expresses the integrating behavior of two eyes watching the stereoscopic image pair [60].

Based on these explorations, we propose a blind (NR) method for SIQA by considering unique experience deterioration caused by distortion in the stereoscopic pair. As there are two images i.e., the left-eye and right-eye images, are offered for assessment, we first introduce the shearlet transform to the two images respectively, simulating the multi-channel processing of human eyes. In each channel, natural scene statistics (NSS) are calculated as quality related features of individual images. Then, the left- and right-view channel content are integrated base on a Gain-Control model. In an effort to evaluate the binocular perception, we calculate the NSS features of integrated content of every channel as well. Considering that the difference between the stereoscopic image pair contains the information of disparity and asymmetric distortion, we also compute the mean of gradient similarity map in each channel between the stereo pair to represent the unique 3D properties. Finally, all these features are fused to predict the quality score by a learned support vector regression (SVR) [62].

On the whole, the contributions of this paper are listed as follows:

- We are the first to introduce the shearlet transform to SIQA and effectively simulate the multi-channel property of visual cortex processing stereoscopic images.
- We calculate the NSS in every sub-frequency channel to extract quality-related features from the individual images of the stereoscopic pair and integrated content as well.
- We compute the mean of the gradient similarity map in each channel between the image pair to indicate the influence caused by asymmetric distortion or unsatisfying disparity.

The rest of this paper is organized as follows. Section II reviews related work. In Section III, the proposed method is detailly represented. Next, our method is experimentally compared with numerous state-of-the-art (SOTA) quality assessment methods on multiple databases in section IV. Finally, the conclusions of this work are drawn in Section V.

II. REALATED WORK

Subjective evaluation needs to build a standard watching room and recruit some graders. Due to the large amount of manpower and material resources consumed, subjective evaluating methods are not suitable for embedding into real-time image processing systems for viewing optimization [18], [19]. However, subjective scores, deemed to be most accurate of representing viewing experience, can be regularly used as benchmarks for developing objective evaluation algorithms. Towards automatically assessing image quality, objective 2D IQA has been widely investigated and many methods have been proposed. Over the years, along with 3D applications universally involved in daily life, SIQA has been emerged as a hot research topic.

In the early stage, SIQA methods are exploited directly based on traditional 2D algorithms. Campisi et al. [27] Introduced four full-reference or reduced-reference 2D image quality evaluation methods for SIQA. Specifically, the final SIQA is derived from combining the left- and right-view image quality scores, respectively calculated from the same 2D algorithm. You et al. [28] explored the effectiveness of more than ten 2D methods utilized in SIQA, and made conclusion that simply integrating the two scores of one stereoscopic pair cannot achieve desirable results. Furthermore, they indicated the specially disparity information generated by watching stereoscopic 3D images should be taken into account for quality evaluation. Benoit et al. [29] employed the SSIM algorithm to first estimate image quality scores of the left and right views, respectively, and averaged the two scores. After that, they computed the disparity score, combined with the previous averaged one, to acquire the final quality score. Based on the binocular perception characteristics, Lin et al. [30] integrated the stereoscopic image pair into one cyclopean image and its score, calculated by classical 2D IQA methods, is obtained for the final SIQA. Chen et al. [31] linearly fused the stereoscopic image pair into a cyclopean image by taking consideration of binocular rivalry. Khan et al. [35] presented a FR method, which combined saliency maps, gradient maps and inner gradient maps with depth perception edges to generate the final SIQA value. Shao et al. [37], [38] suggested that the combining weights should be adapted to distinct distortion types and employed the sparse feature distribution to calculate them. Liu et al. [39] formed a SIQA model considering the impact factors of binocular fusion, rivalry, suppression, and reverse saliency. Yue et al. [40] extracted the naturalness features of the left, right and cyclopean views, together with the quantified similarity and difference between the stereoscopic image pair, for quality prediction through SVR. Shen et al. firstly generated the cyclopean image, rivalry map, depth map and weight map, and then collected three types of features relating to image distortion, depth perception and binocular disparity for quality assessment. Jiang et al. [42] Proposed a unified quality evaluation model for singly and multiply distorted stereoscopic images by learning visual primitives based on a supervised dictionary framework to encode quality related features. Messai et al. [43], [44], [45] created cyclopean images in the first stage, followed by predicting scores based on machine learning or convolutional neural network (CNN). Oh et al. [46] built a deep CNN for blind SIQA trained through two-step regression, where the first step is responsible for automatically extracting local features, and the second part aggregates the local features into global features. Zhou et al. [47] designed a generic deep learning approach called StereoQA-Net. It contains two sub-networks of left and right views, and provides interconnections between the two networks in certain layers. Xu et al. [48] employed the encoder-decoder architecture oriented by binocular rivalry to recover the distorted content and then obtained the perceptual score by using a regression network to the fusion image. Sim et al. [49] utilized a pre-trained VGGNet [50] to collect features for semantic evaluation, and derived handcrafted features for direct image quality assessment, where these two evaluation scores are weighted to form an overall score. *Si* et al. [51] proposed a hierarchical no-reference stereoscopic image quality assessment network simulating human cortex processing of binocular interaction and fusion.

Although the substantial success of above methods has been made, it is still a challenging topic for ameliorating SIQA performance. Currently, deep learning techniques have been widely used in numerous researching fields of image processing, and have shown huge potential of enhancing SIQA. The further development of deep learning methods for SIQA is yet severely hindered by insufficient labeled data available for training an ideal model. Comparatively, traditional SIQA methods can meet needs with a relative very small tagged dataset, and requires extremely less computational power. As human eyes have highly complex and comprehensive procedure of perceiving stereoscopic images, how to manually pick up features correlated well with QoE can be difficult. To effectively represent IQA, extracting quality related features based on well understanding visual characteristics is recognized as a necessity. Also, the two images of a stereoscopic image pair may be distorted asymmetrically, simple quality pooling metrics of the left and right views, such as averaging two scores, are obviously not practical. From the above, we can find that the exiting traditional SIQA algorithms either may not fully consider the distortion characteristics of stereoscopic images, or may neglect delving into the specific visual mechanics of watching 3D image pairs. Towards building a more effective model for SIQA, we first introduce the shearlet transform to SIQA, simulating the multi-channel information process of the cortex. And then, viewing integration are made to mimic binocular vision based on channel content, followed by NSS features of the image pair and channel content of the left image, right image, and integrated view obtained. Finally, channel-based similarity comparisons are conducted between the image pair to inspect the asymmetric distortion or unsatisfying disparity.

III. PROPOESED METHOD

The flowchart of our algorithm is shown in Fig. 1. The luminance maps of the left- and right-view images are first decomposed through shearlet transform, respectively. Then, NSS features of the decomposed images and the original images, which can indicate the left and right images quality individually, are extracted. To measure the combining effects of two eyes, the two images' channel contents are integrated in every channel, and NSS features are calculated in all these synthesized channels. Next, considering the disparity information and asymmetric image distortion, we compute similarity related features. Finally, all these features are fed into SVR for predicting scores of SIQA.

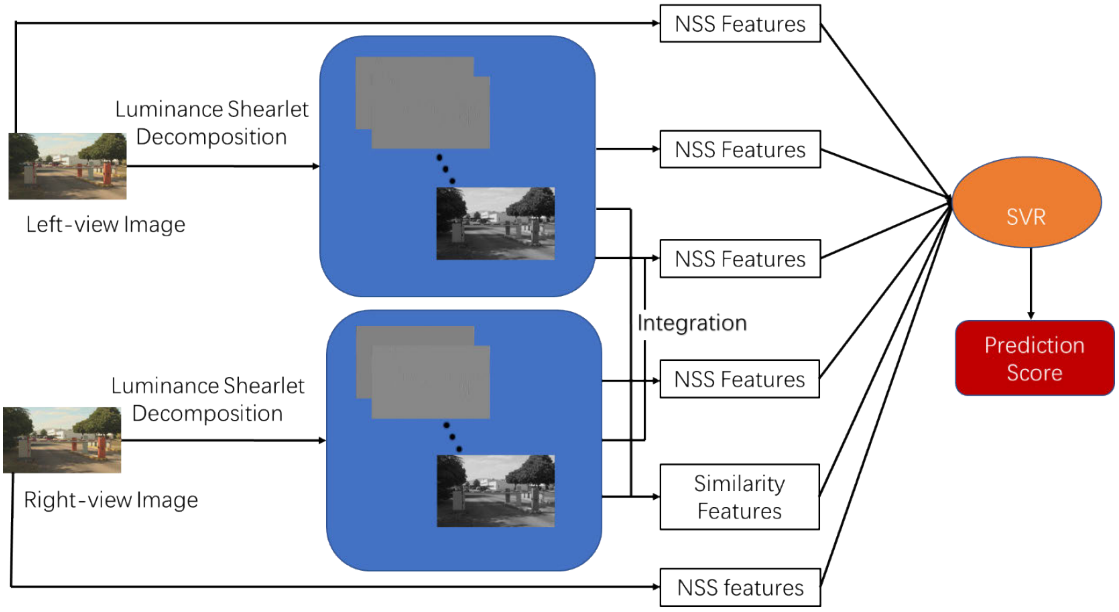

FIGURE 1. Flowchart of the proposed method.

A. SHEARLET TRANSFORM

It is known that HVS performs multiscale and multidirectional [49]. The wavelet transform [50], simulating the multi-channel characteristic, has been used effectively in IQA. However, the wavelet can only provide three directions in each scale, which is not sufficient for representing anisotropic features, such as curves in pictures. Towards representing more directions, various approaches of post wavelet analysis have been proposed, including the curvelet transform [51], the contourlet transform [52] and the shearlet transform [53], [54], [55], [56], etc. The shearlet tansform is built based on affine transformation of scaling, shearing and translation, drawing on the successful experience of the curvelet and contourlet transforms. Owing to the mathematical properties of multiscale, multidirection and anisotropy, the shearlet transform can be employed to well simulate the multi-channel processing mechanism of HVS.

The two-dimensional shearlet transform is defined by

$$SH_f(\alpha, \beta, \ell) = < f, \psi_{\alpha,\beta,\ell} > \tag{1}$$

where $f \in L^2(\mathbb{R}^2)$ is a function to be transformed, $\alpha \in \mathbb{R}_{>0}$, $\beta \in \mathbb{R}$, and $\ell \in \mathbb{R}^2$ denote scaling, shearing and translation parameters, respectively. $\psi \in L^2(\mathbb{R}^2)$ is the generating function, and the shearlet $\psi_{\alpha,\beta,\ell}$ can be represented by

$$\psi_{\alpha,\beta,\ell} = \alpha^{-\frac{3}{4}} \psi \left(\mathrm{A}^{-1}\mathrm{B}^{-1} (X - \ell) \right) \tag{2}$$

with

$$\mathrm{A} = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{\frac{1}{2}} \end{pmatrix}, \mathrm{B} = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \tag{3}$$

Therefore, the shearlets are diversified with the anisotropic scaling and shearing, having the ability of optimally covering the multidimensional singularities.

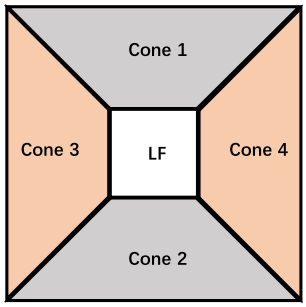

FIGURE 2. The separation of cone-adapted shearlet system in frequency-domain.

To reduce the number of applying the shearlet matrix, the cone-adapted shearlet system is practically adopted, in which the Fourier-domain is separated into two horizontal (Cone 1 and Cone 2 in Fig. 2), two vertical (Cone 3 and Cone 4 in Fig. 2) and a low-frequency (LF in Fig. 2) zones. ψ, $\hat{\psi}$ are introduced as the generating functions of the horizontal and vertical zones, respectively, and ϕ as the scale function of low-frequency zone. Then, the cone-adapted shearlets are given as

$$\psi_{\alpha,\beta,\ell} = \alpha^{-\frac{3}{4}} \psi \left(\mathrm{A}^{-1}\mathrm{B}^{-1} (X - \ell) \right) \tag{4}$$

$$\hat{\psi}_{\alpha,\beta,\ell} = \alpha^{-\frac{3}{4}} \hat{\psi} \left(\hat{A}^{-1}\mathrm{B}^{-T} (X - \ell) \right) \tag{5}$$

$$\phi_\ell = \phi (X - \ell) \tag{6}$$

with

$$\hat{A} = \begin{pmatrix} \alpha^{\frac{1}{2}} & 0 \\ 0 & \alpha \end{pmatrix} \tag{7}$$

where $\alpha \in (0, 1]$, $|\beta| \leq 1 + \alpha^{\frac{1}{2}}$ and $\ell \in \mathbb{R}^2$.

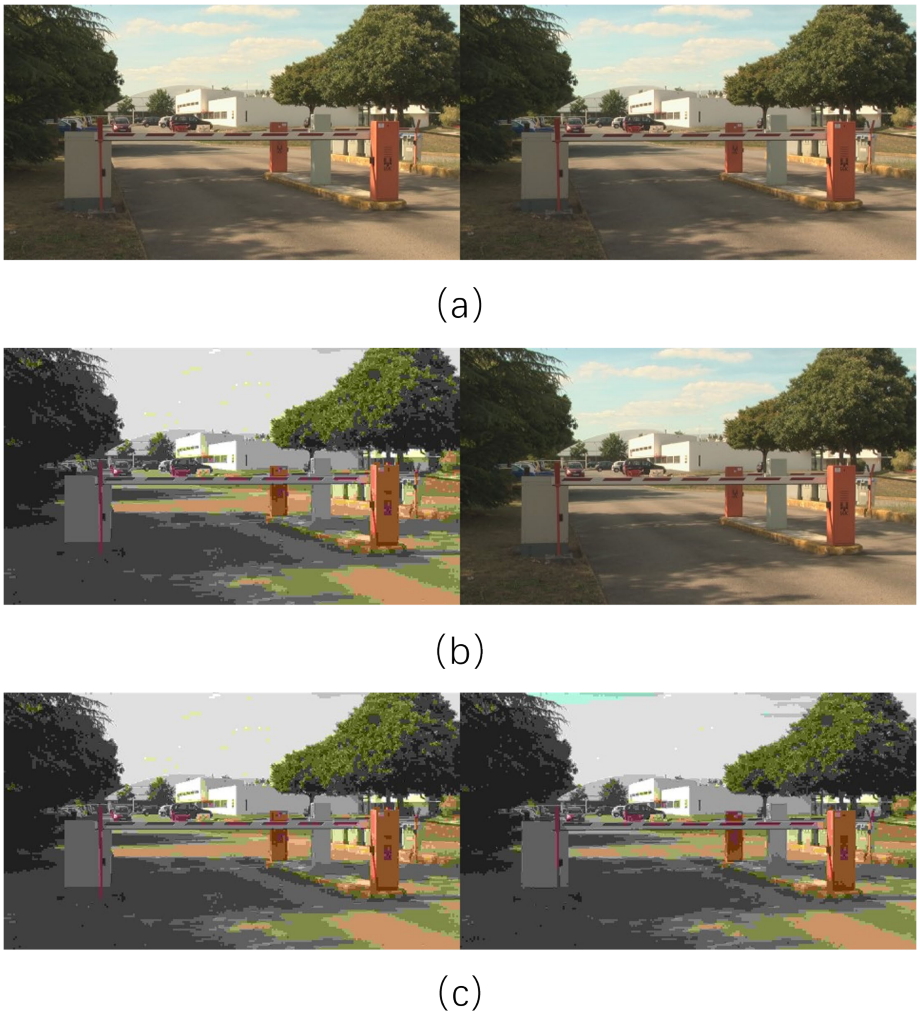

FIGURE 3. Three stereoscopic image pairs of the same content on Waterloo IVC Image Quality database Phase II. (a) is the pristine pair, the left view of (b) is JPEG compressed, and both images of (C) are distorted by JPEG compression.

In [61], it is demonstrated that compactly supported generators ψ, $\hat{\psi}$ and ϕ can be obtained. And, as regards sparsely representing a 2D image, the N-term optimal approximation error of using shearlets follows

$$\mathcal{L}_{\mathbb{N}} = \mathbb{N}^{-2}(log\mathbb{N})^3 \tag{8}$$

which is better than the performance of the wavelet transform and where N denotes the number of the largest shearlet coefficients involved. Afterwards, Lim proposed a discrete shearlet system, i.e., the discrete nonseparable shearlet transform (DNST), based on compactly supported shearlets [56]. DNST is competent in offering local and directional selectivity, and sparsely encoding 2D or 3D data. Therefore, we employ DNST to imitate the multi-channel processing of HVS viewing stereoscopic images.

B. NSS FEATURES OF INDIVIDUAL IMAGES

When viewing stereoscopic content, two images are simultaneously displayed to our left and right eyes, respectively. Fig. 3 shows three stereoscopic image pairs of the same content from Waterloo IVC Image Quality database Phase II. Here, (a) is a reference image pair, the left one of (b), the both images of (c) are compressed by JPEG, and the right image of (b) is identical with the right image of (a). The individual left- and right-view MOS values of (a), (b) and (c) are 95.47 and 95.47, 27.79 and 95.47, and 27.79 and 27.79, respectively. And, the final 3D quality scores of (a), (b) and (c) are 93.22, 38.01, and 21.19, respectively. Hence, it's easy to spot that the individual quality of the left or right view affects the overall experience. In previous research work, NSS were frequently introduced for 2D IQA [33], [34]. To evaluate the naturalness of images, we first calculate the values by the mean subtraction and divisive normalization (MSCN) as defined below

$$\hat{L}(x, y) = \frac{L(x, y) - \mu_L(x, y)}{\sigma_L(x, y) + 1} \tag{9}$$

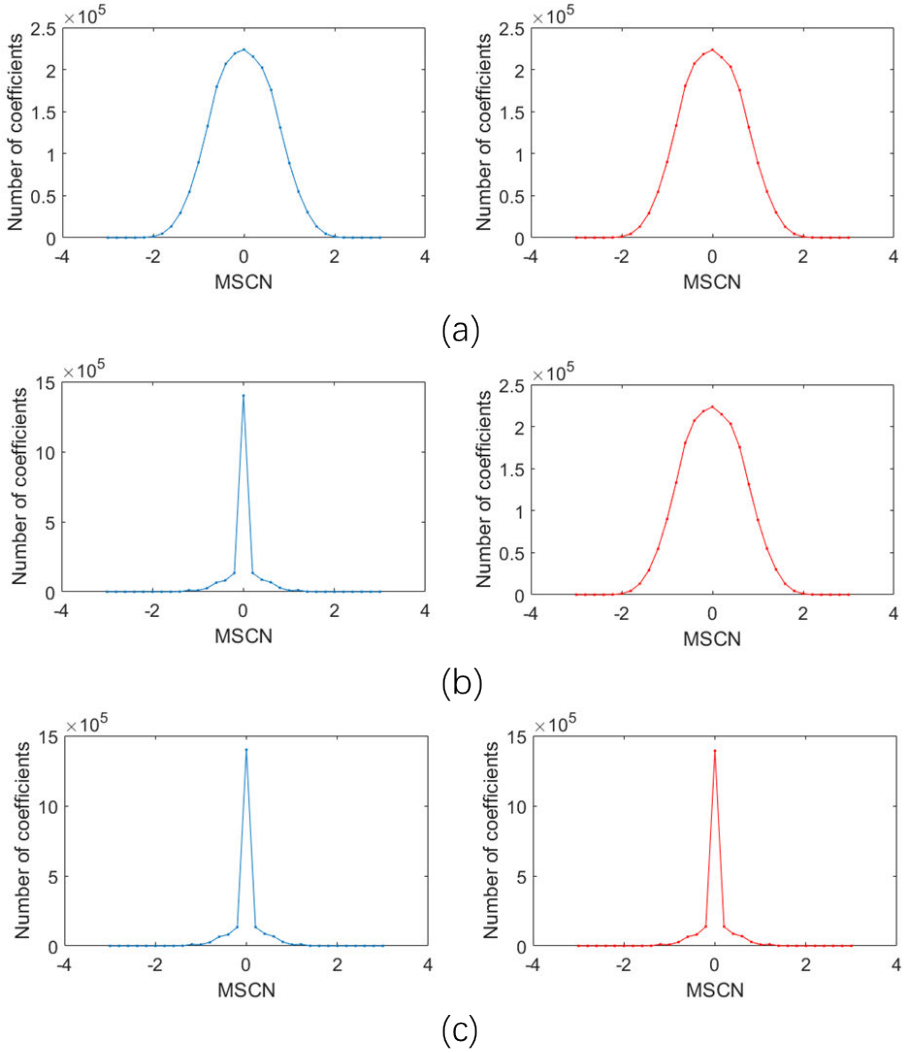

FIGURE 4. The MSCN distributions of the luminance maps of Fig.3. The left and right distributions of (a) correspond to the left- and right-view images of Fig.3 (a), respectively. And, (b) and (c) represent the Fig. 3 (b) and Fig. 3 (c) similarly.

with

$$\mu_L(x,y) = \sum_{m=-M}^{m=M}\sum_{n=-N}^{n=N} w_{m,n} L(x+m, y+n) \tag{10}$$

$$\sigma_L(x,y) = \sqrt{\sum_{m=-M}^{m=M}\sum_{n=-N}^{n=N} w_{m,n}\left[L(x+m,y+n) - \mu_L(x,y)\right]^2} \tag{11}$$

where $L(x,y)$ means the luminance value of an image, $\mu_L(x,y)$ is the local mean and $\sigma_L(x,y)$ is the local standard deviation of the luminance map, and $w_{m,n}$ is the weight value of a circularly symmetric 2D $(2M+1)\times(2N+1)$ Gaussian kernel. Here, we set $M=7, N=7$.

Fig. 4 shows the MSCN distributions of the six individual images of Fig. 3. The distributions of the luminance of the pristine images, such as the both of Fig. 3 (a) and the right one of Fig. 3 (b), are Gaussian-like. However, the shapes of the other luminance distributions significantly deviate from Gaussian appearance. we can utilize a generalized Gaussian distribution (GGD) to fit the MSCN distribution, and the shape and scale parameters of the GGD model are able to represent the quality. The model can be mathematically expressed by

$$f\left(x,\vartheta,\Upsilon^2\right) = \frac{\vartheta}{2\beta\Gamma(1/\vartheta)}\exp\left[-\left(\frac{|x|}{\varsigma}\right)^{\vartheta}\right] \tag{12}$$

with

$$\varsigma = \Upsilon\sqrt{\frac{\Gamma(1/\vartheta)}{\Gamma(3/\vartheta)}} \tag{13}$$

$$\Gamma(\vartheta) = \int_0^{\infty} t^{(\vartheta-1)}e^{-t}dt, \vartheta > 0 \tag{14}$$

where parameters ϑ and ς respectively signify the shape and scale of the distribution, and $\Gamma(\vartheta)$ is a gamma function.

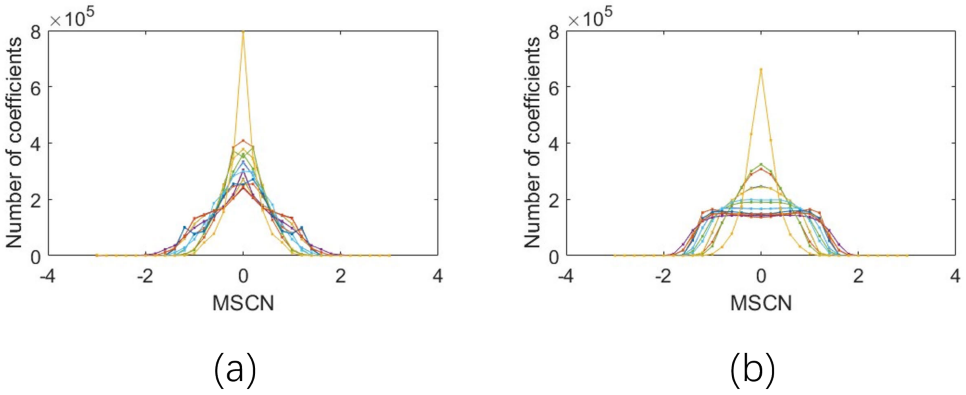

FIGURE 5. The MSCN distributions of sub-band responses, decomposed by DNST. (a) corresponds to the left image of Fig. 3 (b), and (b) are the responses of the left image of Fig. 3 (a).

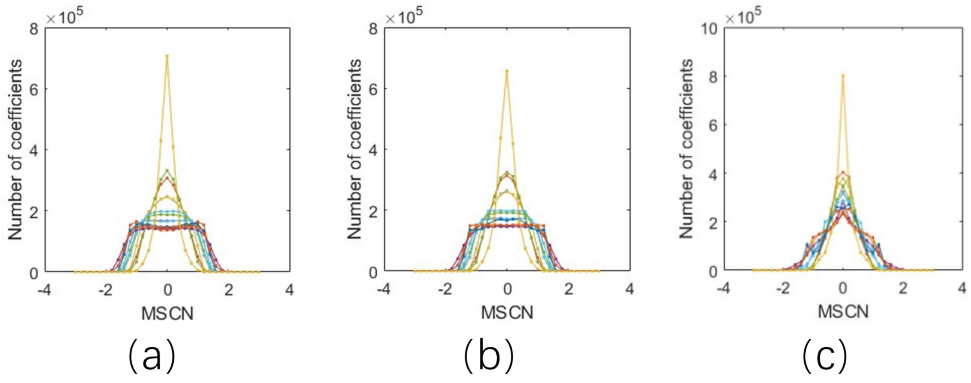

FIGURE 6. The sub-band MSCN distributions of combined vision. (a) , (b) and (c) correspond to Fig. 3(a), Fig. 3(b) and Fig. 3(c).

It has been verified that natural scenes' sub-band responses of wavelet transform are apt to follow heavy-tailed distribution, which can be parameterized by GGD. Inspired by this, we decompose the luminance by DNST and sub-band MSCN distributions are shown in Fig. 5. One luminance map is decomposed into 17 sub-bands, and the Fig. 5(a) and Fig. 5(b) present the sub-band MSCN distributions of the left images of Fig. 3(b) and Fig. 3(a), respectively. We can see that the sub-band distributions of Fig. 3(b), distorted by JPEG compression, appear more centralized to zero than that of the pristine image of Fig. 3(a). hence, we yet employ GGD to obtain statistical features of the sub-bands to represent image quality. Altogether, 72 features attributed to Class F_I are gotten in this part.

C. NSS OF COMBINED VISION

As we know, HVS will combine the views of two eyes to obtain the final visual experience. Accordingly, it is essential to evaluate the combining effects. Due to the various visual sensitivity of sub-bands, we consider to combine the content in each sub-band, instead of just integrating two views to generate a cyclopean image. In each sub-band, we fuse the left-view luminance map with the corresponding right-view one by

$$C(x,y) = W_L(x,y) \times I_L(x,y) + W_R(x,y) \times I_R((x + d(x,y), y) \tag{15}$$

with

$$W_L(x,y) = \frac{\mathbb{C}_L(x,y)}{\mathbb{C}_L(x,y) + \mathbb{C}_R((x + d(x,y)), y)} \tag{16}$$

$$W_L(x,y) = \frac{\mathbb{C}_R((x + d(x,y)), y)}{\mathbb{C}_L(x,y) + \mathbb{C}_R((x + d(x,y)), y)} \tag{17}$$

$$\mathbb{C}_L(x,y) = |I_L(x,y)|^2 \tag{18}$$

$$\mathbb{C}_R(x,y) = |I_R(x,y)|^2 \tag{19}$$

where, I_L and I_R are the sub-band luminance maps of left and right images, respectively, W_L and W_R are their relevant weighted values, d presents the disparity map of two images, and $\mathbb{C}_L$ and $\mathbb{C}_R$ are the energies.

The MSCN distributions of the combined 17 sub-bands are illustrated in Fig. 6. Fig. 6 (a), Fig. 6 (b) and Fig. 6 (c) show the combined luminance distributions of Fig. 3 (a), Fig. 3 (a) and Fig. 3 (c), respectively. We can find those distributions are distinct for images of different MOS values. Therefore, we also introduce GGD to each sub-band for feature extraction, and obtain 34 features grouped as F_C.

D. SIMILARITY FEATURES BETWEEN TWO VIEWS

Two images of a pair may asymmetrically suffer different types or degrees of distortion. In Fig. 3(b), albeit the high-quality right image, the overall 3D quality score of 38.01 is not gratifying induced by the low-quality left image. Furthermore, the disparity in the content presented between two eyes can generate depth perception, which is critical for stereoscopic vision. To catch the specific influence caused by the difference between the left- and right-view images, we calculate the gradient similarity in each sub-band of DNST. The gradient map of a left sub-band and the corresponding right sub-band can be calculated by

$$G_L(x,y) = \sqrt{(I_L(x,y) * g_h(x,y))^2 + (I_L(x,y) * g_v(x,y))^2} \quad (20)$$

$$G_R(x,y) = \sqrt{(I_R(x,y) * g_h(x,y))^2 + (I_R(x,y) * g_v(x,y))^2} \quad (21)$$

with

$$g_h(x,y) = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, g_v(x,y) = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (22)$$

where I_L is the luminance map of a left sub-band, I_R is the luminance map of the corresponding right sub-band, and $g_h(x,y)$ and $g_v(x,y)$ denote the horizontal and vertical filter kernels of Sobel filter. Next, the gradient similarity in a sub-band is expressed as

$$SIM_G(x,y) = \frac{2G_L(x,y)G_R(x,y) + \varepsilon}{G_L^2(x,y) + G_R^2(x,y) + \varepsilon} \quad (23)$$

where ε is a small constant to avoid instability of the equation. In each sub-band, the mean value of the gradient similarity map is computed as a feature, totally 17 features classified as F_S.

E. IMAGE QUALITY EVALUATION

Based on the above explorations, we have obtained 123 quality-related features. The next step is to map these features to the associated MOS values through learning a prediction metric. There are several learning methods have been proposed, such as K-Nearest Neighbor (KNN) [63], Random Forest (RF) [64], SVR [62], Neural Network (NN) [65], etc. Practically, SVR is extensively applied as a learning method [40], [44], [46] to predict the quality score for SIQA due to the fast implementation and high accuracy. Here, we adopt LIBSVM package [66] to learn a SVR model. SVR [62] is formulated as

$$\begin{aligned} \min_{w,b,\zeta\zeta'} \frac{1}{2}||w||_2^2 + \lambda \sum_{i=1}^{\Omega} (\zeta_i + \zeta_i') \\ s.t. \quad w^t \phi(X_i) + b - y_i \le \eta + \zeta_i \\ y_i - w^t \phi(X_i) - b \le \eta + \zeta_i' \\ \zeta_i, \zeta_i' \ge 0, i = 1, 2, \cdots, \Omega \end{aligned} \quad (24)$$

where X_i is the ith input feature vector, y_i is the associated quality score. $K(x_i, x_j) = \phi(X_i)^T \phi(X_j)$ is the kernel function, and $\phi(X_i)$ maps X_i into a higher dimension. In our method, a widely used radial basis function kernel (RBF) is employed, which is defined as:

$$K(X_i, X_j) = \exp\left(-\mathrm{P}||X_i - X_j||^2\right) \quad (25)$$

where P is the kernel parameter. Then, 80% of a specific image dataset are used for training a model, and the other 20% remainder for testing.

IV. EXPERIMENTAL RESULTS

A. DATABASES AND EVALUATION CRITERIA

1) DATABASES

We evaluate the performance of the proposed SIQA on four major databases, such as LIVE 3D Phase I [61], LIVE 3D Phase II [31], Waterloo IVC SIQA database Phase I [62] and Waterloo IVC SIQA database Phase II [62].

LIVE 3D Phase I: The database consists of 20 reference image pairs and 365 distorted image pairs, which are created by inducing Gaussian White Noise (GN), Gaussian Blur (GB), Raleigh Fast Fading (FF), JPEG, or JPEG 2000 (JP2k) to pristine pairs. And of the all 365 distorted pairs, there are 45 for GB, and 80 pairs each for FF, JPEG and JP2K. Besides, the subjective scores are provided in the term of Differential Mean Opinion Score (DMOS), ranging from 0 to 80.

LIVE 3D Phase II: 8 reference image pairs and 60 distorted image pairs, along with their corresponding DMOS values ranging in [9, 0], are offered in this database. The distortion types are the same as LIVE 3D Phase I. However, unlike LIVE 3D Phase I, one image pair may be asymmetrically corrupted by different distortion levels in LIVE 3D Phase II, which is more consistent with realistic factors and increases the difficulty for SIQA. Totally, there are 120 symmetric and 240 asymmetric image pairs.

Waterloo IVC SIQA database Phase I: The database is created from 6 pristine stereoscopic image pairs by introducing three types of distortions, including GN, BB, and JPEG, in four levels. Altogether, there are 252 asymmetrically and 78 symmetrically distorted image pairs with their corresponding MOS values on a scale of 0 to 100.

Waterloo IVC SIQA database Phase II: Compared toWaterloo IVC SIQA database Phase I, this database employs more diverse image content and contains 10 reference image pairs. And, the pristine pairs are corrupted by the same types and degrees of distortions as Waterloo IVC SIQA database Phase I. As a result, there are totally 460 distorted stereoscopic image pairs with MOS values from 0 to 100.

2) EVALUATION CRITERIA

To avoid bias, the train-test process of our learning method is repeatedly executed 1000 times and the median values are obtained as the final results. In each of the 1000 trials, the whole database is randomly partitioned into two parts

TABLE 1. Comparison results on the LIVE 3D Phase I and LIVE 3D Phase II databases. The best results are bolded.

Metrics	Types	LIVE 3D PHASE-I			LIVE 3D PHASE-II		
		RMSE	PLCC	SRCC	RMSE	PLCC	SRCC
SSIM	2D, FR	7.879	0.877	0.877	6.727	0.803	0.792
VSI	2D, FR	8.288	0.863	0.865	7.262	0.766	0.748
NIQE	2D, NR	8.854	0.812	0.801	8.855	0.740	0.708
BRISQUE	2D, NR	6.793	0.910	0.901	7.038	0.770	0.770
Benoit	3D, FR	8.201	0.866	0.856	8.465	0.638	0.662
Khan	3D, FR	-	0.927	0.916	-	0.932	0.922
StereoQUE	3D, NR	6.598	0.917	0.911	7.297	0.854	0.888
Chen	3D, NR	7.247	0.895	0.891	5.102	0.895	0.880
Yue	3D, NR	5.692	0.937	0.914	4.449	0.914	0.906
Messai	3D, NR	5.905	0.911	0.928	4.629	0.932	0.909
Sim	3D, NR	**3.9411**	**0.970**	**0.962**	**3.042**	**0.962**	**0.955**
Proposed	3D, NR	5.454	0.938	0.922	4.392	0.937	0.915

TABLE 2. Performance comparison on the waterloo IVC SIQA database Phase I AND II. The best metrics are marked in bold.

Metrics	Type	Waterloo-IVC PHASE-I			Waterloo-IVC PHASE-II		
		RMSE	PLCC	SRCC	RMSE	PLCC	SRCC
SSIM	2D, FR	12.164	0.634	0.490	16.749	0.484	0.389
VSI	2D, FR	9.664	0.789	0.737	14.556	0.644	0.601
NIQE	2D, NR	13.535	0.510	0.229	18.993	0.124	0.157
Khan	3D, FR	-	0.934	0.925	-	0.910	0.905
Yue	3D, NR	4.610	0.926	0.919	7.739	0.911	0.895
Sim	3D, NR	**4.192**	**0.963**	**0.957**	**4.598**	**0.970**	**0.970**
Proposed	3D, NR	5.591	0.929	0.916	6.959	0.933	0.910

corresponding to image content: 80% samples as the training set, and the remaining 20% as the testing set.

And then, we adopt three commonly used criteria for performance evaluation, such as Root Mean Squared Error (RMSE), Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SRCC), which are all recommended by the Video Quality Experts Group (VQEG) [69]. RMSE, PLCC and SRCC are used to respectively denote the consistency, accuracy, and monotonicity between objective and subjective values.

The PLCC and SRCC range from 0 to 1, and higher PLCC and SRCC but lower RMSE mean better performance. Before calculating PLCC and RMSE, it is necessary to remove the nonlinearity of objective scores by a logistic regression, which is defined as:

$$\mathrm{O} = a_1\left[\frac{1}{2} - \frac{1}{1+e^{a_2(\mathrm{O}_p - a_3)}}\right] + a_4\mathrm{O}_p + a_5 \qquad (26)$$

where O_p is the input objective prediction score, and a_1, a_2, a_3, a_4 and a_5 are the parameters to be fitted by nonlinear regression.

B. PERFORMANCE

To conduct performance evaluation, we make experimental comparisons between the proposed method and some state-of-the-art (SOTA) methods, including SSIM [3], VSI [32], NIQE [9], BRISQUE [9], Benoit's method [29], Khan's method [35], StereoQUE [36], Chen's method [31], Yue's method [40], Messai's method, and Sim's method [49]. For 2D IQA methods, such as SSIM, VSI, NIQE and BRIQUE, the average score of the left- and right-view images is taken to be the predicted score of a stereopair. The other competing algorithms are specifically designed for SIQA, and Messai's method, as well as Sim's method, extracted deep features.

The comparison results on the LIVE 3D Phase I and LIVE 3D Phase II databases are shown in TABLE 1. And, TABLE 2 demonstrates the results on the Waterloo IVC SIQA database Phase I and Waterloo IVC SIQA database Phase II. From the data in the two table, we can get that: First of all, the 2D IQA methods i.e., SSIM, VSI, NIQE and BRISQUE, cannot obtain satisfactory performance on SIQA, which may be caused by ignoring the depth information of the stereoscopic images. But we also see an interesting phenomenon that 3D algorithm, such as Benoit's metric, is inferior to BRISQUE method on the two databases. This indicates the individual quality evaluation of the left- and right-view images is yet significant for SIQA. Finally, our indices can nearly all be ranked into top three on the four databases of the two laboratories, with the exception of the SRCC value on Waterloo IVC SIQA database Phase I. Furthermore, the comprehensive performance, by integratedly considering PLCC, SRCC and RMSE values, of the proposed method outperform all the other competing methods except Sim's algorithm.

In TABLE 3 and TABLE 4, the comparison results of different distortions on the LIVE 3D Phase I and LIVE 3D Phase II databases are presented. From the table, the proposed method has maximum number of ranked in the top three. From data comparison, our method inferiorly assesses the JPEG distortion with PLCC = 0.754, but the PLCC of assessing the GB distortion can reach 0.970. the performance

TABLE 3. PLCC performance of different distortions on the LIVE 3D Phase I and LIVE 3D Phase II Databases. We bold the best results.

Metrics	Type	LIVE 3D PHASE-I					LIVE 3D PHASE-II				
		GB	WN	FF	JPEG	JP2K	GB	WN	FF	JPEG	JP2K
SSIM	2D, FR	0.919	0.944	0.754	0.448	0.875	0.849	0.931	0.861	0.666	0.726
VSI	2D, FR	0.819	0.945	0.739	0.498	0.876	0.777	0.958	0.841	0.669	0.684
NIQE	2D, NR	0.919	0.980	0.579	0.589	0.748	0.969	0.604	0.760	0.712	0.654
BRISQUE	2D, NR	0.926	0.941	0.853	0.615	0.847	0.862	0.846	0.935	0.769	0.593
Benoit	3D, FR	0.948	0.925	0.747	0.640	0.939	0.887	0.861	0.847	0.533	0.647
Khan	3D, FR	0.959	0.947	0.858	0.711	0.951	**0.978**	0.970	0.899	**0.893**	0.927
StereoQUE	3D, NR	0.881	0.919	0.758	0.806	0.938	0.878	0.920	0.836	0.829	0.867
Chen	3D, NR	0.917	0.917	0.735	0.695	0.907	0.900	0.950	0.933	0.867	0.867
Yue	3D, NR	**0.971**	0.962	0.854	0.744	0.934	0.973	**0.986**	0.923	0.843	**0.986**
Messai	3D, NR	0.967	0.936	0.887	0.811	0.905	0.951	0.931	0.851	0.689	0.944
Sim	3D, NR	0.995	**0.994**	**0.987**	**0.957**	**0.994**	0.944	0.908	0.867	0.409	0.902
Proposed	3D, NR	0.970	0.966	0.831	0.754	0.923	**0.978**	0.965	**0.942**	0.832	0.896

TABLE 4. SRCC performance of different distortions on the LIVE 3D Phase I and LIVE 3D Phase II Databases. The best ones are bolded.

Metrics	Type	LIVE 3D PHASE-I					LIVE 3D PHASE-II				
		GB	WN	FF	JPEG	JP2K	GB	WN	FF	JPEG	JP2K
SSIM	2D, FR	0.879	0.938	0.586	0.436	0.858	0.838	0.922	0.834	0.678	0.704
BRISQUE	2D, NR	0.860	0.940	0.784	0.569	0.812	0.862	0.846	**0.935**	0.769	0.862
Benoit	3D, FR	0.931	0.930	0.699	0.603	0.910	0.455	0.923	0.773	**0.867**	0.751
Khan	3D, FR	0.930	0.938	0.809	0.606	0.907	0.885	**0.958**	0.865	0.840	0.833
StereoQUE	3D, NR	0.865	0.910	0.666	0.782	0.917	0.846	0.932	0.860	0.839	0.864
Chen	3D, NR	0.878	0.919	0.652	0.617	0.863	0.900	0.950	0.933	**0.867**	0.867
Messai	3D, NR	0.924	0.925	0.799	0.666	0.921	0.900	0.928	0.880	0.809	**0.909**
Sim	3D, NR	**0.993**	**0.992**	**0.980**	**0.953**	**0.987**	**0.930**	0.888	0.816	0.315	0.815
Proposed	3D, NR	0.928	0.935	0.878	0.779	0.920	0.922	0.941	0.864	0.823	0.858

TABLE 5. SRCC performance on symmetrically and asymmetrically distorted images on the LIVE 3D Phase II database.

Metrics	SYMM	ASYMM
SSIM	0.828	0.733
BRISQUE	0.849	0.667
Benoit	0.860	0.671
Chen	0.918	0.834
StereoQUE	0.857	0.872
Messai	0.921	**0.909**
Proposed	**0.923**	0.875

TABLE 6. Comparison results of the different combination of feature groups.

Features			LIVE 3D PHASE-I		LIVE 3D PHASE-II	
F_I	F_C	F_S	PLCC	SRCC	PLCC	SRCC
√	×	×	0.892	0.834	0.801	0.767
×	√	×	0.793	0.761	0.763	0.742
√	√	×	0.897	0.841	0.823	0.790
√	×	√	0.904	0.893	0.875	0.867
×	√	√	0.907	0.894	0.883	0.859
√	√	√	**0.928**	**0.915**	**0.917**	**0.914**

gap means our method has room for upgrading in terms of JPEG compression. However, we can see Sim's metric perform outstanding on almost every distortion type except the JPEG distortion. Especially notable is that both the PLCC and SRCC values of Sim's method for JPEG distortion on LIVE 3D Phase II are minimum among those of all the methods.

How to evaluate asymmetric distortion is a challenging task for SIQA. We do comparative experiments on the LIVE 3D Phase II database. As we can see from TABLE 5, the proposed method has the biggest SRCC value of 0.923 on symmetric distortions, and has the second biggest SRCC of 0.875, merely below the SRCC value of Messai's method, on asymmetric distortions. Our method also represents the SRCC index on asymmetrically distorted images is close to that on symmetrically distorted images with a small gap of 5.2%, but the gap between the two distortions of Benoit's algorithm can even reach 22.0%. From these comparative results, our method shows the competing power in evaluating asymmetric distortions as well as in assessment of symmetric distortions.

C. ABLATION ANALYSIS

As introduced in Section III, our method fully considers HVS processing of Stereoscopic images, and extracts three types of features for SIQA based on multi-scale and multi-direction decomposition by shearlet transform. The NSS features of individual images, F_I, are extracted for evaluating

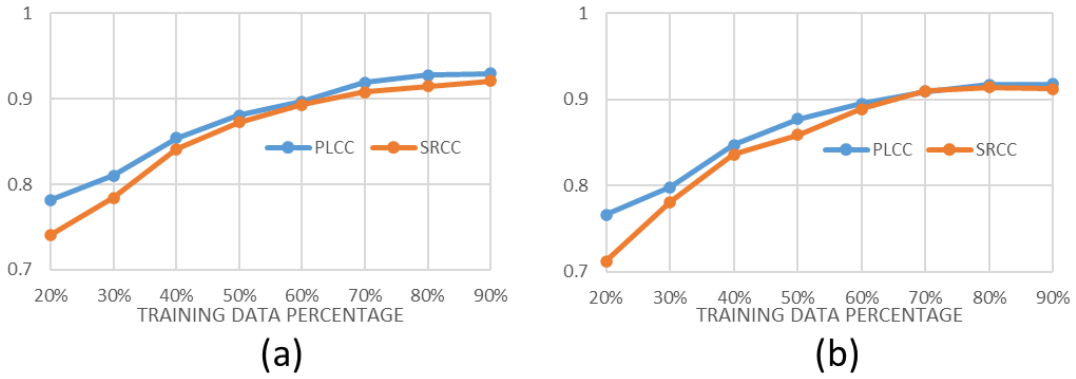

FIGURE 7. Performance results with different training data percentage. (a) shows the PLCC and SRCC values on the LIVE 3D Phase I database and (b) demonstrates the corresponding results on the LIVE 3D Phase II database.

2D image quality. Meanwhile, the NSS features of synthesized sub-bands, F_C, are used to assess the integrated visual quality between two eyes. We also calculate the similarity features, F_S, to measure the disparity information or asymmetric distortion. In order to configure the contribution of each feature group, we implement different combinations of features for ablation analysis. the ablation results are shown in TABLE 5, and it is evident that the combinations can boost the performance. For instance, the PLCC of combining F_C and F_S is 0.904, about 14% higher than that of only applying F_C, and the combination of all three categories of features obtains the highest PLCC value of 0.928 on the LIVE 3D Phase I. And, the other three columns of indices in the table show the similar property as well. accordingly, we can infer the three feature groups complement each other and combining them all are desirable.

Since the proposed method is a train-test metric, it is necessary to figure out how to divide a database into training and testing sets. We make experiments by increasing the percentage of training set from 10% to 90%, and 1000 non-overlap random splits and tests are done for each percentage to obtain median results. From the trend lines in Fig. 7, we can see the performance gradually improved with the increased size of training set. Yet, the 90% of training set has flat rises in terms of both PLCC and SRCC values on the both LIVE databases. To avoid potential overfitting due to excessive percentage of training data, we employ 80% of all data for training and the remaining 20% for testing as in [40].

As mentioned above, the will-be-evaluated images firstly require multi-scale and multi-direction shearlet decomposition, which imitates HVS processing. The number of scales and orientations may have influence on the performance. To explore the correlations between the number settings and quality values, we do experiments by setting the decomposing modes shown in TABLE 7. the SRCC performance and mean computational times of one image are shown on Fig. 8. The SRCC value of ST3 is slightly smaller than the SRCC values of ST4 and ST5, but the average time consumption of ST3 is

TABLE 7. Different settings of the shearlet decomposition. The first number within a pair of brackets indicates the number of orientations for the first scale, the second number corresponds to the second scale, and so on.

Decomposition	Scales	Orientations
ST1	1	[4]
ST2	2	[4,4]
ST3	3	[4, 4, 4]
ST4	3	[4, 4, 8]
ST5	4	[4, 4, 4, 4]

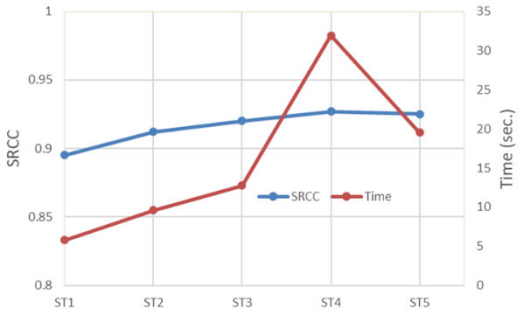

FIGURE 8. SRCC values and mean computational times of the varying decomposing modes on the Waterloo-IVC PHASE-I database.

substantially lower than those of both ST4 and ST5. Specifically, the average times of ST4 and ST5 are respectively 150.7% and 53.4% higher than this of ST3. Hence, we adopt ST3 in our method for the consideration of balancing effectiveness and efficiency.

D. CROSS-DATASET EVALUATION

To evaluate the generalization ability of our method, several cross-dataset experiments are conducted. We train the proposed method on LIVE 3D PHASE-I or Waterloo-IVC

TABLE 8. Cross data-set performance evaluation.

Train	Test	PLCC	SRCC
LIVE 3D PHASE-I	LIVE 3D PHASE-II	**0.854**	**0.850**
LIVE 3D PHASE-I	Waterloo-IVC PHASE-I	0.773	0.761
LIVE 3D PHASE-I	Waterloo-IVC PHASE-II	0.714	0.711
Waterloo-IVC PHASE-I	LIVE 3D PHASE-I	0.801	0.783
Waterloo-IVC PHASE-I	LIVE 3D PHASE-II	0.756	0.748
Waterloo-IVC PHASE-I	Waterloo-IVC PHASE-II	**0.815**	**0.798**

PHASE-I database, and test on the other three databases, respectively. The indices on TABLE 8 show superior generalization performance, with the lowest PLCC and SRCC values both bigger than 0.7. Besides, we can find that closer relationship between the training and testing datasets obtains better results. For example, training on LIVE 3D PHASE-I performs best on predicting LIVE 3D PHASE-II database, and the same outcome can be discovered on using Waterloo-IVC PHASE-I database as training set.

E. FURTHER DISCUSSION

we have developed an effective SIQA algorithm based on shearlet decomposition mimicking the multi-channel processing of HVS. After the decomposition, features extracted from the sub-bands and the original stereoscopic pairs are utilized to represent the quality through SVR. The Experimental results show the competitiveness of our algorithm compared to the other state-of-the-art methods. Albeit the advantages our method presents, there are some aspects deserve to be concerned as follows:

1) As seen from TABLE 1 and TABLE 2, Sim's method almost performs best in all the criteria. since Sim's method is deep-learning based, it has a comparatively large and complex network to effectively represent quality-related features. Instead of directly learning a network model, Sim's method applies two identical parallel pre-trained DNNs to extract 4096 semantic features from a stereoscopic pair, and yet the authors noted the possible overfitting tendency due to the relatively too small sizes of databases on cross-database experiments. Also, it's interesting to find both the PLCC and SRCC values of Sim are smallest on JPEG distortion on the Waterloo IVC SIQA database Phase II from TABLE 3 and TABLE 4. Therefore, our work is still of considerable actual value. In the future work, the primary task should be developing a large-scale database for SIQA based on deep neural network (DNN).
2) All the experiments are done on a laptop with a i5 CPU @ 2.5GHz and an 8GB RAM. The operating system is Windows 10 and working software is MATLAB 2016b. Limited by computing power, our method spends about 12 seconds on each stereopair of the Waterloo-IVC PHASE-I database. Among all the working steps of the proposed method, the shearlet decomposition consumes the biggest part of time, more than 10 seconds. It is known that shearlet decomposition can be done in a parallel manner, and introducing advanced hardware that supports parallel computing may further reduce the time spent. On the other hand, we can see even the SRCC of ST1 can offer a desirable value, slightly smaller than 0.9. Hence, we also can reduce the time consumption through the reduction of scales and orientations, at the expense of comparatively small decrease in performance.
3) Our method extracts NSS features in the sub-bands of left-view, right-view and combined images, and the effectiveness of these features have been validated in TABLE 5. Nevertheless, we can see from TABLE 3 and 4 that the performance on each type of distortion is inconsistent. Specially, the results on JPEG distortion are worst, all indices near 0.8, compared to the data on the other distortions. This indicates those NSS features may lean to accurately represent some certain types of distortions. To boost the performance, some hand-picked features, such as structure and texture features, can be added for tries.
4) We mimic the multi-channel property of HVS based on shearlet decomposition for SIQA. And, the two-view image pair is integrated by a simple sub-band based energy weighting. Since the processing mechanism of HVS is quite sophisticated, the study of it is still on the initial stage. Hence, the simulation method should maintain amelioration for better evaluating SIQA with the deeper understanding of the HVS processing. Another solution to precisely simulating the multi-channel processing is that we can apply a DNN to simulate a channel and interconnect all channels for combination. The parameters of those branch networks can be automatically set by training on the SIQA database. This multi-branch DNN will be inevitably fulfilled at the expense of costing massively more computing resources.

V. CONCLUSION

Considering the binocular characteristics of HVS, we propose a new blind quality assessment method for evaluating stereoscopic 3D image pairs. Firstly, the individual left- and right-view images are decomposed into multi-channel content based on shearlet transform to simulate the HVS processing. The NSS statistics of the sub-bands and the original ones from individual left and right images are calculated as quality-related features. After that, we combine the sub-band pairs based on energy weighting and extract the NSS features of integrated sub-bands as well. Then, we calculate the gradient similarity between the image pair in each sub-band to denote the asymmetric distortion and disparity information. Finally, SVR is applied to fuse all the extracted features into the subjective score. The experimental results demonstrate the outperformance of our method compared to the state-of-art SIQA metrics on the benchmark databases.

REFERENCES

[1] C.-C. Su, A. K. Moorthy, and A. C. Bovik, "Visual quality assessment of stereoscopic image and video: Challenges, advances, and future trends," in *Visual Signal Quality Assessment*. Cham, Switzerland: Springer, 2015, pp. 185–212.

[2] J. Schild, J. LaViola, and M. Masuch, "Understanding user experience in stereoscopic 3D games," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2012, pp. 89–98.

[3] 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.

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

[5] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

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

[7] 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.

[8] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[9] 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.

[10] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2035–2048, Aug. 2018.

[11] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[12] D. Lee and K. N. Plataniotis, "Toward a no-reference image quality assessment using statistics of perceptual color descriptors," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3875–3889, Aug. 2016.

[13] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, Apr. 2017.

[14] Y. Zhou, L. Li, J. Wu, K. Gu, W. Dong, and G. Shi, "Blind quality index for multiply distorted images using biorder structure degradation and nonlocal statistics," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.

[15] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[16] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[17] 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.

[18] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.

[19] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[20] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, International Telecommunication Union, 2015.

[21] M. Urvoy, M. Barkowsky, and P. L. Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Ann. Telecommun.*, vol. 68, nos. 11–12, pp. 641–655, Dec. 2013.

[22] R. Patterson, "Human factors of 3-D displays," *J. Soc. Inf. Display*, vol. 15, no. 11, pp. 861–871, 2007.

[23] R. Blakeab and H. Wilson, "Binocular vision," *Vis. Res.*, vol. 51, no. 7, pp. 754–770, Apr. 2011.

[24] I. P. Howard and B. J. Rogers, *Seeing in Depth*. New York, NY, USA: Oxford Univ. Press, 2008.

[25] G. Mather, *Foundations of Sensation and Perception*. Oxon, U.K.: Psychology Press, 2008.

[26] D. Stidwill and R. Fletcher, *Normal Binocular Vision: Theory, Investigation and Practical Aspects*, 1st ed. New York, NY, USA: Wiley, 2010.

[27] P. Campisi, P. L. Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 2110–2114.

[28] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, Scottsdale, AZ, USA, 2010, pp. 1–6.

[29] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 389–392.

[30] Y. Lin, J. Yang, W. Lu, Q. Meng, Z. Lv, and H. Song, "Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 89–101, Feb. 2017.

[31] 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.

[32] 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.

[33] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, no. 1, pp. 167–192, Jan. 2008.

[34] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, Jul. 1994.

[35] S. Khan 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.

[36] B. Appina, S. Khan, and S. S. Channappayya, "No-reference stereoscopic image quality assessment using natural scene statistics," *Signal Process., Image Commun.*, vol. 43, pp. 1–14, Apr. 2016.

[37] F. Shao, K. Li, W. Lin, G. Jiang, and Q. Dai, "Learning blind quality evaluator for stereoscopic images using joint sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2104–2114, Oct. 2016.

[38] 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, May 2016.

[39] 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, Oct. 2017.

[40] 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.

[41] L. Shen, R. Fang, Y. Yao, X. Geng, and D. Wu, "No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 1, pp. 59–72, Feb. 2019.

[42] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1866–1881, Apr. 2019.

[43] O. Messai, F. Hachouf, and Z. A. Seghir, "AdaBoost neural network and cyclopean view for no-reference stereoscopic image quality assessment," *Signal Process., Image Commun.*, vol. 82, Mar. 2020, Art. no. 115772.

[44] O. Messai, A. Chetouani, F. Hachouf, and Z. A. Seghir, "No-reference stereoscopic image quality predictor using deep features from cyclopean image," *Electron. Imag.*, vol. 297, no. 9, pp. 1–9, Jan. 2021.

[45] O. Messai, A. Chetouani, F. Hachouf, and Z. A. Seghir, "3D saliency guided deep quality predictor for no-reference stereoscopic images," *Neurocomputing*, vol. 478, pp. 22–36, Mar. 2022.

[46] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4923–4936, Oct. 2017.

[47] W. Zhou, Z. Chen, and W. Li, "Dual-stream interactive networks for no-reference stereoscopic image quality assessment," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3946–3958, Aug. 2019.

[48] J. Xu, W. Zhou, Z. Chen, S. Ling, and P. L. Callet, "Binocular rivalry oriented predictive autoencoding network for blind stereoscopic image quality measurement," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[49] K. Sim, J. Yang, W. Lu, and X. Gao, "Blind stereoscopic image quality evaluator based on binocular semantic and quality channels," *IEEE Trans. Multimedia*, vol. 24, pp. 1389–1398, 2022.
[50] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014, *arXiv:1405.3531*.
[51] J. Si, B. Huang, H. Yang, W. Lin, and Z. Pan, "A no-reference stereoscopic image quality assessment network based on binocular interaction and fusion mechanisms," *IEEE Trans. Image Process.*, vol. 31, pp. 3066–3080, 2022.
[52] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.
[53] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 7, pp. 674–693, Jul. 1989.
[54] E. J. Candès and D. L. Donoho, "Curvelets—A surprising effective nonadaptive representation for objects with edges," in *Curves and Surfaces*. Nashville, TN, USA: Vanderbilt Univ. Press, 2000, pp. 105–120.
[55] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[56] W.-Q. Lim, "Nonseparable shearlet transform," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2056–2065, May 2013.
[57] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *Proc. SPIE*, vol. 5914, Sep. 2005, Art. no. 59140U.
[58] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Anal.*, vol. 39, no. 1, pp. 298–318, Jan. 2007.
[59] G. Kutyniok, W.-Q. Lim, and R. Reisenhofer, "ShearLab 3D: Faithful digital shearlet transforms based on compactly supported shearlets," *ACM Trans. Math. Softw.*, vol. 42, no. 1, pp. 1–42, Jan. 2016.
[60] J. Ding and G. Sperling, "A gain-control theory of binocular combination," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 4, pp. 1141–1146, Jan. 2006.
[61] G. Kutyniok and W.-Q. Lim, "Compactly supported shearlets are optimally sparse," *J. Approximation Theory*, vol. 163, no. 11, pp. 1564–1589, Nov. 2011.
[62] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[63] L. Breiman and R. Ihaka, "Nonlinear discriminant analysis via scaling and ACE," Dept. Statistics, Univ. California, Berkeley, CA, USA, Tech. Rep. 40, 1984. [Online]. Available: https://digitalassets.lib.berkeley.edu/sdtr/ucb/text/40.pdf
[64] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2021.
[65] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," *Psychol. Learn. Motivat.*, vol. 24, pp. 109–165, Jan. 1989.
[66] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
[67] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Sep. 2013.
[68] 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.
[69] VQEG. (Mar. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.vqeg.org/

DONGHUI WAN received the B.S. degree in communication engineering from the Guilin University of Electronic Technology, Guilin, China, in 2002, and the M.S. degree in communication and information systems from Soochow University, Suzhou, China, in 2005. He is currently pursuing the Ph.D. degree in communication and information systems with the Communication University of China, Beijing, China. He is a Lecturer with the School of Science and Engineering, Huzhou College. He has been working on image/video quality assessment. His research interests include image processing, machine learning, and communication signal processing.

XIUHUA JIANG received the M.S. degree from Shandong University, Jinan, China, in 1982. She is currently a Professor and a Ph.D. Tutor with the School of Information and Communication Engineering, Communication University of China. Her research interests include image quality assessment, image processing, and video compression, where she has filed four patents, authored six books, and published 14 journal articles.

QING SHEN was born in Shanxi, China, in 1982. She received the B.S. degree in computer science and technology and the M.S. degree in computer applications technology from the North University of China, in 2004 and 2007, respectively. She is currently a Professor with Huzhou University, Huzhou, China. She is the author or coauthor of more than 30 articles in refereed international journals. Her current research interests include image processing, intelligent information processing, and swarm intelligence.