Blind Stereoscopic Image Quality Assessment Accounting for Human Monocular Visual Properties and Binocular Interactions

YUN LIU[ID]1, WEIQING YAN2, ZHI ZHENG3, BAOQING HUANG1, AND HONGWEI YU1

1College of Information, Liaoning University, Shenyang 110036, China
2School of Computer and Control Engineering, Yantai University, Yantai 264005, China
3Department of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Yun Liu (yunliu@lnu.edu.cn)

ABSTRACT Human visual perceptual model is a key factor for evaluating stereoscopic image quality. This paper focuses on the contributions of monocular and binocular properties on quality perception and proposes a novel blind stereoscopic image quality assessment model by comprehensively digging the relationship between visual features and quality perception. The statistical quality-aware monocular features are extracted from both left view and right view to reveal monocular quality perception, including the color statistical features which are missed in most previous models, while the multiple features of the summation signal and the entropy features of the difference signal are extracted to quantify the binocular quality perception. Finally, support vector regression (SVR) is utilized to train a regression model based on the extracted features and the subjective scores. Three public databases, LIVE 3D Phase I, LIVE 3D Phase II, and MCL 3D Database, are adopted to prove the effectiveness of the proposed model. Experimental results demonstrate that the proposed model is superior to other existing state-of-the-art quality metrics.

INDEX TERMS Stereoscopic image quality, monocular feature, binocular feature, human visual system.

I. INTRODUCTION

Over recent years, stereoscopic contents have grown explosively, including stereoscopic film, virtual reality, 3D-TV, and so on, which drive the demand and development of stereoscopic image quality assessment (SIQA). Consequently, the quality assessment of stereoscopic contents aroused much attention and became an important research topic. Different from two-dimensional (2D) image, the three-dimensional (3D) image has the depth information, which makes the stereoscopic image quality assessment more difficult. Many types of distortions, such as JPEG compression (JPEG) distortion, JPEG2000 compression (JP2K) distortion, additive white noise (WN) distortion, the fast fading (FF) distortion, and the Gaussian blur (Blur) distortion, may be introduced in 3D images to affect not only visual perception quality but also depth perception.

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang[ID].

In general, SIQA method can be divided into three categories based on the availability of the reference stereoscopic image: full-reference (FR) method, reduced-reference (RR) method and no-reference (NR) method. The FR method, which requires all information of the reference image, usually achieves the best results among the above three methods. At the earlier stage, FR SIQA models are proposed based on 2D image quality assessment (IQA) model[1]–[3]. The overall performance of the 2D extended type models on 3D image is not as good as that on 2D image. Since human is the final judgment subject, the characteristics of the human visual system (HVS) should be considered. Many works about how human eyes perceive 3D images and get the depth perception are proposed in SIQA area. These works proved that there exists a visual channel to process the information received from human eyes, and there is one combined image (called ''cyclopean image'') formed in human brain to perceive the depth perception. Chen *et al.* [4] proposed an FR SIQA method to model the whole procedure and evaluate 3D image quality. Bensalma and Larabi [5] then considering

this binocular theory, built an effective quality index based on the complex wavelet transform (CWT). By investigating human visual cortex and binocular visual properties, Lin and Wu [6], Shao *et al.* [7], and Shao *et al.* [8] presented many good models, which further proved the effectiveness of human visual characteristics on SIQA area.

The RR method, proposed based on partial information of the reference image, is generally second to the FR method, which mainly applied to access the distortion of the image in the transmission process. Because the reference image is not always available in practical cases, therefore, the NR method, without referring any information of the reference image, receives increasing attention. The NR method is also referred to as the blind SIQA method. Inspired by human binocular perception model, Ryu and Sohn [9] proposed a binocular perception metric by measuring the blurriness and the blockiness of the image views. Zhou *et al.* [10] proposed an NR SIQA metric based on two binocular combination models: Eye-Weighing model and the Contrast Gain-Control model. Considering the importance of the depth perception, Chen *et al.* [11] proposed a cyclopean image combination model by utilizing 3D cues in the disparity map to estimate the 3D image quality, which yield better performance than the models without considering depth information. With the development of machine learning technology, researchers focus on modeling human binocular visual combination behavior to solve the SIQA problem. Zhou *et al.* [12] built a blind SIQA metric by utilizing the self-similarity of binocular features and applied the support vector regression (SVR) model to drive the overall quality score. The above models are built based on the visual single-channel theory, while the double-channel model is proved to be a more plausible way, which has been applied and proved its effectiveness in our previous SIQA works [13], [14]. Then based on the model of human double-channel theory, Yang *et al.* [15] modeled human visual summation and difference channels to extract image features and utilized the SVR method to predict the quality of the stereoscopic images. These models have better performance than previous works, which provide us a way to extract the quality-sensitive visual features to build the SIQA method.

Since one stereoscopic image contains two views (left and right views), the monocular view quality and the interactions between these two views inevitably affect the overall quality perception. Various factors should be considered, including monocular image quality, binocular interaction (binocular rivalry, suppression, etc.), and depth information. To cope with this challenge, many works have been proposed. Shao *et al.* [16] proposed a blind SIQA method using joint sparse representation, in which stereoscopic image quality is predicted based on the weighted monocular quality score. To handle the asymmetric distortion, Shao *et al.* [17] designed an NR metric by utilizing both monocular and binocular properties for quality assessment. Likewise, Liu *et al.* [18] combined the monocular and binocular information and

proposed a blind SIQA model by using the SVM method to fuse the quality score.

Considering the promising prospect of the NR SIQA model in practical application, we propose a novel blind SIQA model by comprehensively digging the intrinsic relationship between human visual properties and quality perception. Motivated by the above works, the monocular and binocular visual features are both extracted, which are correlate well with human subjective observation, especially for evaluating the asymmetric degradation on 3D images. Finally, SVR is utilized to predict the final quality score. The advantages of this work are as follows:

(1) Towards the monocular features, three types of quality-aware features are extracted to reveal image quality under different types and degrees of distortions, including the color statistical features, which have a significant influence on quality perception but ignored in most SIQA models.

(2) Multi-scale and multi-orientation visual properties are deployed to address the summation visual features and quantify the naturalness of the stereoscopic images, which can improve the accuracy of the proposed model.

(3) The difference signal implicitly contains the disparity and depth information, in which the pixels exhibit strong dependence that affects visual perception. Two types of entropies are adopted to estimate the dependence level between pixels and the depth degradation, which is a simple way to get rid of depth information and obtains a low computational complexity.

The rest of this paper is organized as follows. Section II presents the motivations of this work. Section III describes the proposed framework. The experimental results and discussion are declared in Section IV, and the conclusions are drawn in Section V.

II. MOTIVATIONS

For the problem of SIQA, the most intuitive way is to simulate the human visual perception characteristics to build a subjective assessment-like model. Therefore, numbers of computational models have been proposed to explain the possible mechanism of the human visual system [6], [19]–[21]. However, the above models mainly focus on the binocular interaction in the human brain, which is not enough to simulate human visual perception in image quality assessment. Seuntiens *et al.* [22] reported that the overall 3D image quality is not only affected by the binocular information, but also by the information of the left view and the right view (especially for the asymmetrically distorted 3D image). Lv *et al.* [23] summarized that 3D image quality assessment should be account for two types of distortions related to 3D perception quality, that is monocular distortion and binocular distortion. Each view of one 3D image suffers from the distortion called monocular distortion, which will cause monocular quality degradation, while both views suffer from symmetric or asymmetric distortion called binocular distortion, which will cause binocular confusion, visual discomfort and so on. Wang *et al.* [24]

also found that the monocular cues have a strong impact on the depth perception based on subjective experiments. Jiang *et al.* [25] addressed both monocular and binocular quality issues and built an effective NR quality evaluation method for the stereoscopic image using a three-column non-negativity constrained sparse auto-encoder framework, which proves that the 3D image quality perception is determined not only by the qualities of each view but also their interactions.

Based on the neural processing visual pathways in the visual cortex, Shao *et al.* [26] modeled the perceived stereoscopic image quality ς as the posterior probability with the left view I_L and right view I_R. By simulating human visual cortex, they divided the quality index into two independent parts, as follows:

$$P(\varsigma|I_L, I_R) = P_M - k * \frac{P(I_L \cap I_R)}{P(I_L, I_R)} * P_B \qquad (1)$$

where P_M and P_B are the monocular perception index and binocular perception index, respectively. The work [26] proved the reasonable and effectiveness of the combination of monocular and binocular information on 3D image quality evaluation. Therefore, we follow the above conclusions and build a blind SIQA model based on two indexes: monocular and binocular perception indexes.

A. MONOCULAR PERCEPTION

Lots of effective assessment methods for single view image have been proposed based on HVS [27]–[29]. With the development of the conventional neural network (CNN), many NR IQA models have been built by learning image features and regression model [30]–[32]. However, the color information, which plays an important role in visual perception, is ignored in the above models. Zhang *et al.* [33] extended their FR IQA framework on color information and found that the later model with color information has better overall performance than the original model. Wang *et al.* [34], considering the importance of the color information in quality perception, studied an NR-IQA model using the natural color statistic features. Later, Ziaei Nafchi *et al.* [35] and Sun *et al.* [36] took image color information as an important part in the IQA model, and obtained good results in IQA area. Besides, color related artifacts also play an important role in visual discomfort estimation for stereoscopic image quality perception, which should be considered in stereoscopic image quality assessment [37], [38]. The above works reveal that the color information embodies important quality information and can be effectively used to learn useful quality-sensitive features, which is adopted to address the monocular quality in our model. Since the left view is similar to the right view, we extract the monocular color features from the left view to as one set of the monocular features.

The natural scene statistics (NSS) features of the image can reflect the naturalness of the distorted images, which have been demonstrated its effectiveness on IQA area [39]–[43]. This paper takes the NSS features of each view's luminance information as the other two sets of monocular features,
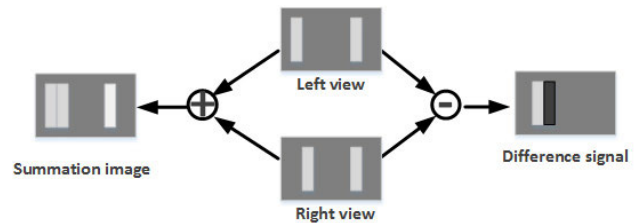

FIGURE 1. The double channel visual model.

which contain important asymmetric degradation information on quality perception.

1) BINOCULAR PERCEPTION

As the supplement of the monocular perception, the binocular information is considered to study binocular perception quality in our proposed model. Many biological visual models have been proposed to explain human binocular behavior, which can be divided into two categories: single-channel model and double-channel model. The former predicts that human eyes collect the individual eye view separately, and then combine them to generate the unique stereopsis in the human brain [44]. An alternative way, the double-channel model, suggests that there are two separate visual signals in the human brain to obtain the stereopsis: the summation signal and the difference signal [45], [46], shown in Fig.1. Kingdom [47] proved the existence of these two adaptable binocular signals in human binocular perception. May *et al.* [48] and May and Zhaoping [49] later gave strong support to this theory and proved that human brain contains the summation channel and the difference channel that adapt to prevailing binocular statistics, which provide an optimal way for visual system to transmit the binocular information.

At the physiological level, it proved that V1 neurons receive the multiplexed signals from the summation and difference channels, so that they can tune to different disparities [50]. Besides, the summation and difference signals have been applied to evaluate the stereoscopic image quality and achieved better performance than previous models [43], [51]. Considering the binocular difference channel is missing from the single-channel model, we take the double-channel model to simulate human binocular interaction behavior.

For the summation signal, some biological models have been proposed to describe it [52]–[54]. Ding and Sperling [55] proposed a Grain-Control model, which can well explain an early stage of human binocular visual behavior, such as Fechner's paradox [56] and cyclopean perception (including binocular fusion and binocular rivalry). Thus, we utilize the Gain-Control cyclopean model to synthesize the summation signal. Since the image local structural information is sensitive to the distortion, and the image gradient has a strong ability to represent the image structure distortion. Thus, we obtain the gradient-related information of the cyclopean signal to get the binocular features.
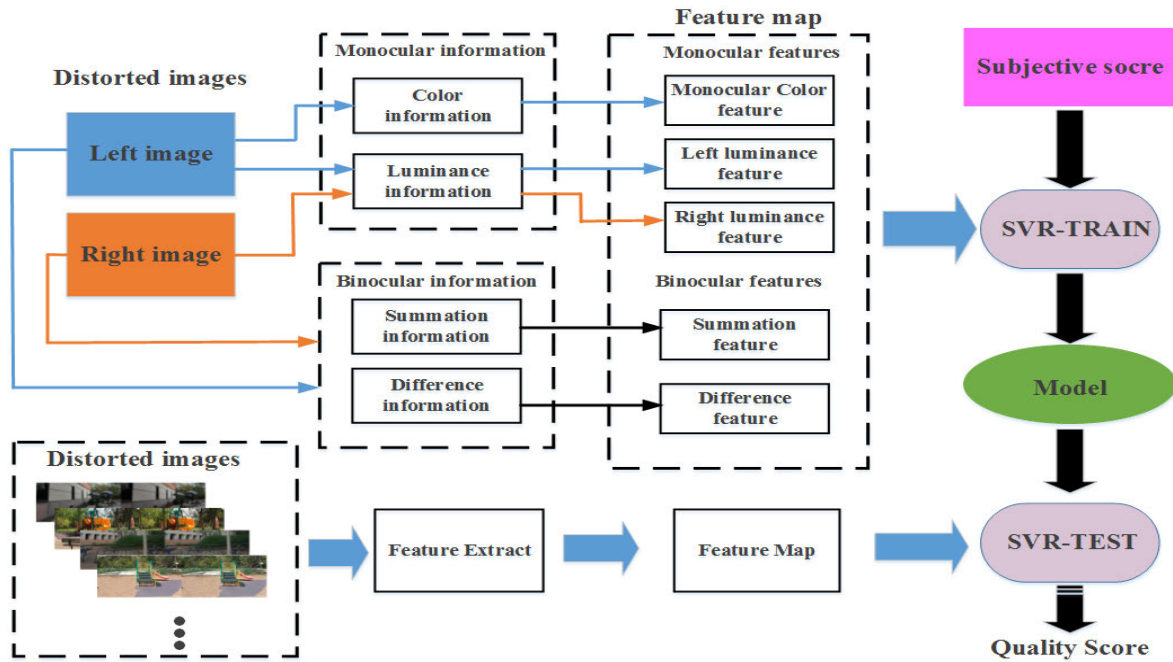
FIGURE 2. The flowchart of the proposed stereoscopic image quality assessment model.

The difference signal is given by subtracting the right image from the left image, which has been proved that the difference information can well explain disparity perception in SIQA area [15], [57]. Works [58], [69] revealed that the difference signal is sensitive to the disparity and carries information critical for stereo perception. While image entropy, a tool for measuring the amount of information, captures the statistical information over scales [60], which is significantly sensitive to quality degradation and distortion [39]. There are two types of image entropies: spatial entropy and spectral entropy. The former one reveals the statistical characteristics of the local pixel, while the spectral entropy reflects that of the local DCT coefficients. Both of them are sensitive to the types and degrees of distortions [43], [61]. We hypothesize that quantifying the changes of these two entropies will make it possible to predict the type and degree of distortions that affecting the difference signal. Thus, we address the features of the difference signal from the spatial entropy and the spectral entropy.

III. PROPOSED MODEL

In this paper, we build a blind SIQA model by considering both monocular and binocular interaction features. The whole framework is shown in Fig.2. Firstly, luminance NSS features of two views and the color features of the left view are extracted to estimate the monocular quality. Secondly, the binocular quality index is computed based on the multi-scale and multi-orientation features extracted from the summation signal (the cyclopean image) and the entropy features from the difference signal. Finally, the overall quality

score is obtained based on a two-stage regression method: training stage and testing stage. In the training stage, a regression model is learned based on the above features to build the relationship between features and subjective scores via SVR. In the testing stage, the final quality score is estimated by input all the extracted features into the trained model.

A. MONOCULAR FEATURES
For the monocular features, we firstly extract the luminance NSS features of two views in the spatial domain. Take the left view as an example, we operate a divisive normalization to its luminance information I_L, as follows:

$$\hat{I}_L(i,j) = \frac{I_L(i,j) - \mu_L(i,j)}{\sigma_L(i,j) + 1} \quad (2)$$

where $\mu_L(i,j)$ and $\sigma_L(i,j)$ are the local mean and the standard deviation of the left view's luminance information, which are computed by using a 2D Gaussian kernel with the size of $3*3$ as in [62]. As we all know, the coefficients of Eq.2 follow a Gaussian distribution, which can be modeled by a zero-mean generalized Gaussian distribution (GGD), as follows:

$$GGD(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp(-(\frac{x}{\sigma}\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}})^\alpha) \quad (3)$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt, x > 0$. α and σ^2 reflect the shape and the variance of the distribution, respectively. When the signal suffers distortion, this distribution will change. These changes can be quantified by using these two parameters: α, σ^2, which related to the distortion types and degrees, and can be used to predict the image quality [63]. Here,

we apply the locally mean subtracted and contrast normalized (MSCN) coefficients to obtain these two parameters to build the first set of luminance NSS features $f_{L1} = (\alpha, \sigma^2)$.

Besides, we extract the other set of luminance NSS features of the left view from the distribution of the products of pairs of adjacent MSCN coefficients, which also contains image quality information [63]. The adjacent MSCN coefficients are obtained along four orientations: horizontal, vertical, main-diagonal and secondary-diagonal orientations. Similar to the MSCN coefficients, the adjacent MSCN coefficients can be modeled by a zero mode asymmetric generalized Gaussian distribution (AGGD) [64], as follows:

$$
\begin{aligned}
&AGGD\,(x; \gamma, \beta_l, \beta_r) \\
&= \begin{cases} \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp(-(\dfrac{-x}{\beta_l})^\gamma) & \forall x \le 0 \\ \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp(-(\dfrac{-x}{\beta_r})^\gamma) & \forall x \ge 0 \end{cases}
\end{aligned} \tag{4}
$$

The mean of the AGGD distribution can be obtained by:

$$
\eta = (\beta_r - \beta_l) \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)} \tag{5}
$$

The parameters $(\gamma, \beta_l, \beta_r, \eta)$ are extracted as the second set of the luminance NSS features of the left view f_{L2}. Then, 16 parameters (4 parameters * 4 orientations) of AGGD are extracted. Considering the multi-scale information in image, we extract the left view monocular features at two scales: the original image scale and a reduced scale down-sampling the image by a factor of 2. Thus, a total of 36 features are used to build the left view monocular feature map. Similarly, we extract the above two sets of luminance features (f_{R1}, f_{R2}) from the right view to get the right view monocular feature map.

As pointed out in [34], the color statistical properties can't directly be applied to get the above two types of NSS features. Hurvich [65] suggested that there are three types of cones in human retina, and each of them has two opponent color members. Later, Ruderman *et al.* [66] indicated that the logarithmic-scale color information follows a Gaussian probability model, which provide us a way to capture the color statistic features. We extract the color features from the left view in RGB space based on the logarithmic scale transformation as follows:

$$
\begin{cases} R' = \log R - \langle \log R \rangle \\ G' = \log G - \langle \log G \rangle \\ B' = \log B - \langle \log B \rangle \end{cases} \tag{6}
$$

where $\langle X \rangle$ is the mean of X. In fixing the axes of the new logarithmic space, an orthogonal transformation is conducted as follows:

$$
\begin{cases} C_1 = (R' + G' + B') * 1/\sqrt{3} \\ C_2 = (R' + G' - 2B') * 1/\sqrt{6} \\ C_3 = (R' - G') * 1/\sqrt{2} \end{cases} \tag{7}
$$

The coefficients of these unit direction vectors (C_1, C_2, C_3) follow a Gaussian probability law, which can be modeled by

a Gaussian model as follows:

$$
GM(x; \varsigma, \rho^2) = \frac{1}{\sqrt{2\pi}\rho} \exp(\frac{-(x - \varsigma)^2}{2\rho^2}) \tag{8}
$$

The parameters (ς, ρ^2) of three color components (C_1, C_2, C_3) at two scales are extracted as the color statistic features f_C (3 color components * 2 parameters * 2 scales = 12 color features). Then the monocular visual features extraction is finished (total $36 * 2 + 12 = 84$ features).

B. BINOCULAR FEATURES

As pointed out in [49], [50], the binocular features are extracted from the summation signal and the difference signal, both of which play an important role in binocular interaction and quality perception. Firstly, we compute the cyclopean image based on the Gain-Control model [55] to represent the summation signal in the human brain, shown as follows:

$$
S = \frac{1 + E_L}{1 + E_L + E_R} * I_L + \frac{1 + E_R}{1 + E_L + E_R} * I_R \tag{9}
$$

where E_L and E_R are the normalized energy responses of the left and right views, respectively, which are computed by the local energy of the log-Gabor filter response. More details can refer to [33]. The 2D log-Gabor filter is given by:

$$
G_{s,o}(\omega, \theta) = exp[-\frac{(log(\omega/\omega_s))^2}{2\sigma_s^2}] \times exp[-\frac{(\theta - \theta_0)^2}{2\sigma_o^2}]. \tag{10}
$$

where ω and ω_s are the radial frequency and the center frequency, respectively. θ and θ_o are the orientation angles of the filter. σ_s and σ_o determine the filter's radial bandwith and the angular bandwith. By deploying the log-Gabor filter to filter the monocular view, a set of responses on scale s and orientation o, denote as $[\eta_{s,o}, \varsigma_{s,o}]$, can be got. The local energy E_o along orientation o is given by:

$$
E_o = \sqrt{\sum_s \eta_{s,o} + \sum_s \varsigma_{s,o}} \tag{11}
$$

The local energy of the log-Gabor filter responses is then obtained as follows:

$$
E = \sum_o E_o \tag{12}
$$

Since the distortion affects the image quality across multiple scales and multiple orientations, we apply the multi-scale log-Gabor filter to extract the binocular NSS features, which can well explain the spatial-frequency response of visual cortical cells in each eye's receptive filed [67], [68]. Given that gradient information has a strong ability to reflect the local structural distortion, the binocular features of the summation signal are extracted from the gradient map of the cyclopean image. By deploying the log-Gabor filter of 3 different center frequencies and 4 different orientations to the cyclopean image, 12 response maps can be obtained. The related parameters are set as follows: $\sigma_s = 0.6$ and $\sigma_s = 0.71$.
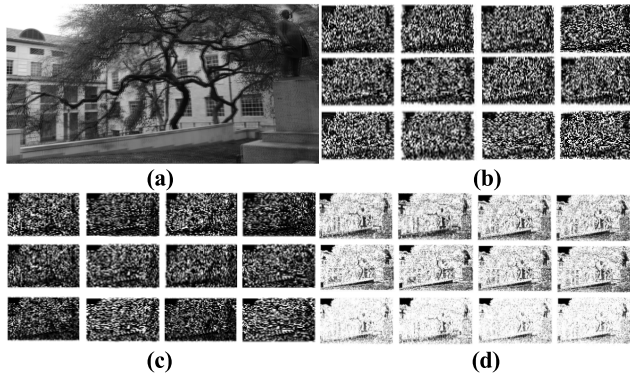
FIGURE 3. (a) Cyclopean image. (b) Log-Gabor response image. (c) Directional gradient image. (d) Gradient magnitude image.

FIGURE 4. Four distorted images. (a) d1. (b) d2. (c) d3. (d) d4.

Define the response of the real part as $R_{m,n}$, and that of the imaginary part as $IM_{m,n}(m = 0, 1, 2; n = 0, 1, 2, 3)$. We first adopt the GGD model to extract the best-fit model parameters (α and σ^2) of the above two response maps and that of the smoothed directional gradient components of the above two response maps as the first set of NSS features of the summation signal f_{S1}, and a total of 144 features are extracted.

Then a Weibull distribution is applied to model the gradient magnitude of the above response maps to obtain the additional NSS features of the summation signal, which is given by:

$$WD(x; a, b) = \begin{cases} \dfrac{a}{b^a} x^{a-1} \exp(-(\dfrac{x}{b})^a) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (13)$$

The parameter α reflects the texture of the gradient magnitude map, while b reflects its local contrast. Hence we take these two best-fit model parameters as the additional summation NSS features $f_{S2} = (a, b)$, and a total of 48 features are extracted.

To illustrate the above features, the responses of 3 scales and 4 orientations filter banks of a cyclopean image are shown in Fig.3. It needs to mention that different content images have the same behavior, and we only present one image as an example. Fig.3(a) is the cyclopean image, and (b)-(d) are the log-Gabor responses, the directional gradient components and the gradient magnitudes, respectively. It reveals that the responses of different scales and orientations differ from each other, and different components have different information, which prove the reasonable and plausible of extracting the multi-scale and multi-orientation features to generate the quality-aware features.

The difference signal represents the difference between two views, and the amount of information within it reveals the depth perception and the asymmetric distortion. The entropy is an effective way to compute the amount of information, and high sensitive to the degree and type of image distortion [43], [61], which provides us a way to extract the features of the difference signal. Different from the NSS features based on pixels relation, we compute the statistic features of
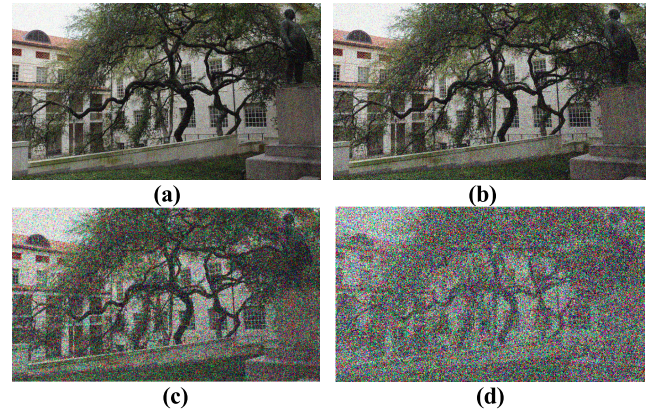
the difference signal based on the local region (8*8 image patch), which can represent the local image structural information and quality perception. Spatial and spectral entropies are both considered in this paper. In order to visualize the behavior of the above two entropies against the types and the degrees of distortions, images affected with different degrees and different types of distortions and the corresponding histogram distributions of two entropies are shown in Fig.4 and Fig.5. It needs to mention that different content images have the same behavior, and we only present one image as an example. Fig.4 shows the left view of four 3D images affected by four different degrees of WN distortion from LIVE 3D Phase I [69]: (a) Figure d1 with the minor degree, (d) Figure d4 with the most severe degree, and (b)(c) with the middle degree. Fig.5(a) is the spatial entropy histograms of the original difference signal and of Fig.4(a)-(d). Fig.5(b) is the spectral entropy histograms of the original difference signal and of Fig.4(a)-(d). Fig.5(c)(d) is the spatial entropy histograms and the spectral entropy histograms of the original difference signal and of five difference signals with five types of distortions, respectively. Fig.5(e)(f) is the spatial entropy histograms and the spectral entropy histograms of the original difference signal and of two distorted difference signals affected by symmetrically and asymmetrically distortions.

Fig.5(a) reveals that the distortions in Figures d1, d2, and d3 increase the mean and skew the spatial entropy histogram to the left, while the distortion in Figure d4 reduces the mean and skews the histogram to the right. In Fig.5(b), the distortions in Figures d1 and d2 increase the mean and skew the spectral entropy histogram to the left, while the distortions in Figures d3 and d4 reduce the mean and skew the histogram to the right. It can be concluded that the spatial entropy and the spectral entropy are sensitive to the degree of distortion. Fig.5(c)(d) reveals that the types of distortions are related to the distributions of the spatial and spectral entropies. The mean and the skew vary with the type of distortion. For example, WN increases the mean and skew the spatial entropy histogram to the left, while the other four types of distortions reduce the mean and skew the histogram to the right. The
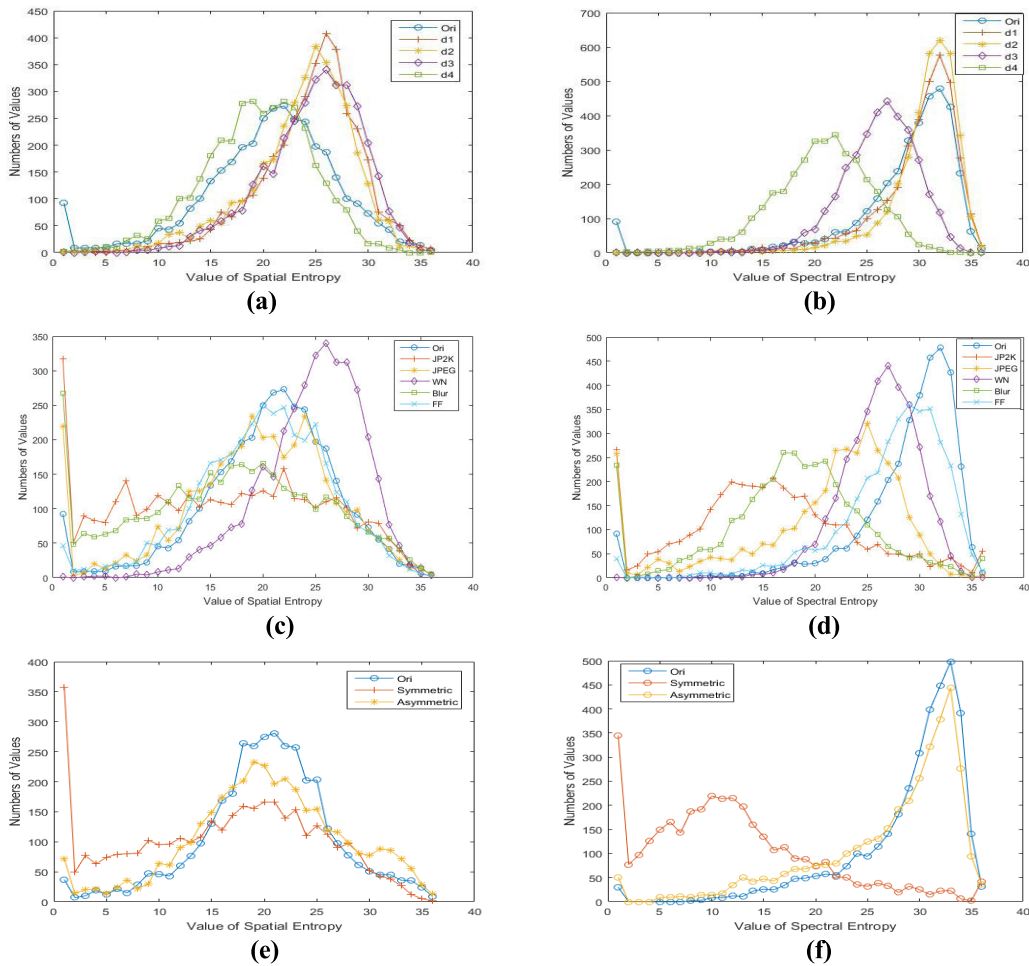
FIGURE 5. (a) Spatial entropy under different degrees of WN. (b) Spectral entropy under different degrees of WN. (c) Spatial entropy under five types of distortions. (d) Spectral entropy under five types of distortions. (e) Spatial entropy under symmetric and asymmetric distortions. (f) Spectral entropy under symmetric and asymmetric distortions.

spectral entropy of the original difference image has a bigger mean and skew than that of all the types of distortions. Fig.5(e)(f) indicates that the distributions of the spatial and spectral entropies can reflect the distortion type: symmetric distortion or asymmetric distortion. Overall, spatial entropy and spectral entropy are both sensitive to the type and the degree of the distortion, which can be utilized to extract the features of the difference signal.

Since the mean and the skew of the histogram change with the type and the degree of the distortion, we extract the mean and the skew of the difference signal as its quality features. Specifically, the difference signal is firstly partition into 8∗8 block as the basic computation unit to compute the entropy. The spatial entropy is given by:

$$EN = -\sum p(x) * \log_2 p(x) \tag{14}$$

where $p(x)$ is the probability density of a block, and x is the pixel values of one block. Then we compute the spectral entropy of each block of the difference signal based on the

normalized DCT coefficients matrix c as follows:

$$EN_s = -\sum \sum \frac{c(i,j)^2}{\sum \sum c(i,j)^2} * \log_2 \frac{c(i,j)^2}{\sum \sum c(i,j)^2} \tag{15}$$

where $1 \leq i, j \leq 8$ and $i \neq j$. Then we take the mean and skew of the spatial and spectral entropies at two scales as the features of the difference signal f_D, and a total of 8 features are extracted. Overall, all quality-aware visual features are extracted, as tabulated in Table 1.

C. QUALITY PREDICTION

After extracting all the above features, the LIBSVM package is utilized to solve the quality prediction problem [70]. In this paper, we adopt SVR with a radial basis function (RBF) kernel to train the prediction function, which is effectively used in other NR image quality assessment models [16], [43], [51]. We firstly train a regression model based on the extracted features and the corresponding subjective scores. Each database is divided randomly into two non-overlapping parts: 80% training sample, and 20% testing sample, as other SIQA works. 1000 times iterations are performed, and the mean

TABLE 1. Summary of extracted features.

Feature vector	Feature description	Total number
$f_{L1}(f_{R1})$	MSCN coefficients of the left view (right view)	4(4)
$f_{L2}(f_{R2})$	Adjacent MSCN coefficients of the left view (right view)	32(32)
f_C	Distribution measurements of the color features	12
f_{S1}	GGD model parameters of Log-Gabor response and gradient feature of summation signal	144
f_{S2}	Distribution measurements of gradient magnitude features of summation signal	48
f_D	The mean and skew of spatial entropy and spectral entropy of the difference signal	8

value is reported for the parameter selection during the training procedure. At the testing stage, we use the regression model to map the extract features into the final 3D quality objective score.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATABASES AND PERFORMANCE CRITETIA

In this section, three publicly available 3D IQA databases are utilized to verify the performance of the proposed model: LIVE 3D Phase I [69], LIVE 3D Phase II [11] and MCL Database [71]. All these databases provide both reference images and distorted images with corresponding subjective scores.

- LIVE 3D Phase I contains 365 distorted images with a co-registered human score in the form of the difference mean opinion score (DMOS). They created from 20 reference images and affected symmetrically by five types of distortions: JP2K (80), JPEG(80), WN (80), FF (80), and Blur (45).
- LIVE 3D Phase II is created from 8 reference images. Every reference stereopair was processed to create 3 symmetric distorted images and 6 asymmetric distorted images. Altogether, there are 360 distorted images with co-registered human scores in the form of DMOS.
- MCL 3D Database consists of 20 reference images and 648 symmetrically distorted stereoscopic images affected by 6 types of distortions (JPEG, JP2K, GB, WN, down-sampling blur (SB), and transmission error (TE)) at four distortion levels.

In this paper, only the distorted images are used, and three commonly used performance criteria are utilized to verify the performance of the proposed model: the Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), and root mean square error (RMSE). The higher the values of PLCC and SROCC, the better the proposed model. While the smaller the value of RMSE, the better the model. PLCC=SROCC=1 and RMSE=0 indicate the perfect performance.

B. OVERALL PERFORMANCE COMPARISON

We compare the proposed model with three 2D IQA models (SSIM, BRISQUE [62] and ADD-GSIM [72]) and eleven

3D metrics: Shao2016 [16], Shao2017 [73], Zhou2017 [10], Zhou'2017 [12], Liu2018 [74], Wang2018 [75], Ma2018 [20], Yue2018 [40] Shao2018 [17], Chen2019 [21], and Liu2019 [18], on LIVE 3D Phase I and LIVE 3D Phase II. Works conducted on MCL 3D Database are relatively fewer, so here we take eight works compared with our proposed model on MCL 3D Database: SSIM, BRISQUE, Shao2016, Zhou2017, Liu2018 [74], Yang'2018 [15], Chen2019 and Liu2019. The comparison results are listed in Table 2 and Table 3, and the top two performances are highlighted in boldface. "NA" is marked as the unavailable data.

From Table 2, the 2D-extended metrics, especially the ADD-GSIM model, hold the competitive performance on the LIVE 3D Phase I, but they fail to predict the quality of the asymmetric distorted images due to the missing binocular information. The proposed model ranks the top two on all the three databases, which proves its effectiveness on predicting 3D image quality. Specifically, in LIVE 3D Phase I, the proposed model has the best PLCC and RMSE performance, and obtains the competitive SROCC performance compared to Ma2018's work. And in LIVE 3D Phase II, the proposed model has better overall performance than Ma2018's work, which concludes that the proposed model performs better for asymmetrically distorted images than the symmetric distorted images. Besides, the proposed model has better performance than Liu2019's work in LIVE 3D Phase I, but yields worse performance than Liu2019' model in LIVE 3D Phase II. It needs to mention that the performance values in Phase II of the proposed work are very close to that of the top one: Liu2019's work, which concludes that the proposed model shows high correlations with human perception for symmetrically distorted images. Based on the above analysis, the proposed model correlates well with human subjective perception for both symmetrically and asymmetrically distorted images.

For the MCL 3D Database, the proposed model ranks the first among all the models in terms of PLCC and SROCC, and the value of RMSE is very close to the top one. The proposed model presents a robust and accurate performance and further proves its effectiveness on quality assessment of the symmetrically distorted images. Overall, the proposed model correlates well with human subjective perception by

TABLE 2. The overall performance comparison in terms of PLCC, SRCC and RMSE on LIVE 3D image quality database.

Metric	LIVE 3D Phase I			LIVE 3D Phase II		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
SSIM	0.873	0.877	8.006	0.803	0.792	6.471
BRISQUE	0.910	0.901	6.793	0.771	0.770	7.038
ADD-GSIM	0.939	0.933	7.534	0.736	0.730	7.642
Shao2016	**0.953**	0.944	NA	0.903	0.885	NA
Shao2017	0.891	0.885	NA	0.784	0.805	NA
Zhou2017	0.941	0.921	5.540	0.923	0.919	4.262
Zhou'2017	0.934	0.904	5.838	0.905	0.890	4.747
Liu2018	0.924	0.948	5.564	0.930	0.932	4.120
Wang2018	0.924	0.916	6.272	0.744	0.714	7.536
Ma2018	0.951	**0.950**	**5.074**	0.911	0.905	4.652
Yue2018	0.937	0.914	5.652	0.914	0.906	4.449
Shao2018	0.885	0.867	NA	0.910	0.872	NA
Chen2019	0.939	0.923	NA	0.865	0.859	NA
Liu2019	0.941	0.936	5.548	**0.950**	**0.949**	**3.526**
Proposed	**0.959**	**0.949**	**4.574**	**0.936**	**0.933**	**3.804**

TABLE 3. The overall performance comparison in terms of PLCC, SRCC and RMSE on MCL 3D image quality database.

Metric	PLCC	SROCC	RMSE
SSIM	0.765	0.783	1.677
BRISQUE	0.810	0.786	1.500
Shao2016	**0.914**	0.904	1.023
Zhou2017	0.885	0.852	1.177
Liu2018	0.896	0.891	1.185
Yang'2018	0.912	0.905	1.062
Chen2019	0.905	0.896	**0.835**
Liu2019	0.900	**0.908**	1.141
Proposed	**0.916**	**0.908**	0.945

considering monocular and binocular visual properties, and can well evaluate the quality of symmetrically and asymmetrically distorted stereoscopic images, which indicates that the proposed metric is a consistent, stable and accurate model that has better application in reality than other works.

C. PERFORMANCE ON INDIVIDUAL DISTORTION TYPE

To validate the effectiveness of the proposed model on predicting the specific type of distortion, we test the proposed model and other representative SIQA models on different distortion types. Since the MCL 3D Database, like LIVE 3D Phase I, is a symmetric database, and the LIVE 3D Phase I and II contain five commonly different distortion types, we test all the works on LIVE 3D databases to conduct the performance comparison experiments. To save space, we only present the results of PLCC and SROCC. The results are listed in Table 4 and Table 5, in which the top results are highlighted in boldface.

The proposed model yields the best PLCC performance on JP2K and WN in LIVE 3D Phase I, and on Blur and FF in LIVE 3D Phase II. Although it doesn't perform the best than other models on the rest distortion types, our model shows competitive robust and stability of prediction accuracy across all the distortion types. For the SROCC performance, our model yields the best performance on JP2K and WN in
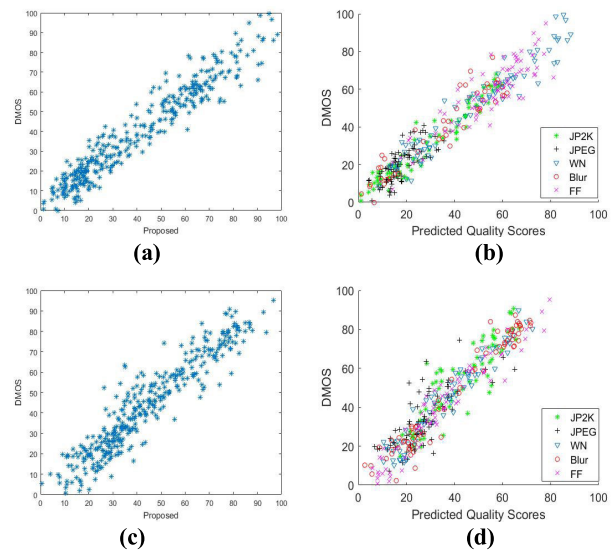

(a)

(b)

(c)

(d)

FIGURE 6. Scatter plots of predicted objective scores versus subjective scores in (a)(b)LIVE 3D Phase I and (c)(d)LIVE 3D Phase II.

LIVE 3D Phase I and on FF in LIVE 3D Phase II, and holds good performance on other types of distortions, which proves the monotonicity of the proposed model. Overall, the proposed model holds the best performance 7 times compared with Zhou2017's model(1 time), Wang2018's model (0 time), Yang'2018's model (0 time), Yue2018's model (4 times), Chen2019's model (5 times) and Liu2019's model(3 times), which concludes that the proposed model can predict the 3D image quality under different types of distortions, including symmetrically and asymmetrically distortions.

Fig.6 shows the scatter plots of subjective scores versus the objective scores based on our proposed model, and the degree of convergence represents the overall performance: better convergence to the better performance. The vertical axis denotes the provided DMOS values and the horizontal axis represents the predicted objective quality scores. Fig.6(a)(b) is the overall performance and the distortion performance of

TABLE 4. The Performance comparison on five types of distortion in terms of PLCC.

	LIVE 3D Phase I					LIVE 3D Phase II				
	JP2K	JPEG	WN	Blur	FF	JP2K	JPEG	WN	Blur	FF
Zhou'2017	0.945	0.748	0.953	**0.975**	0.838	0.835	0.860	0.973	0.977	0.915
Wang2018	0.909	0.617	0.949	0.951	0.778	0.957	0.790	0.829	0.784	0.868
Yang'2018	0.947	0.819	0.957	0.952	0.876	0.936	0.862	0.952	0.975	0.935
Yue2018	0.934	0.744	0.962	0.971	0.854	**0.986**	0.843	**0.986**	0.973	0.923
Chen2019	0.905	**0.870**	0.952	0.939	**0.941**	0.836	0.858	0.927	0.950	0.892
Liu2019	0.925	0.713	0.929	0.948	0.857	0.919	**0.868**	0.966	0.973	0.941
Proposed	**0.948**	0.768	**0.967**	0.910	0.852	0.942	0.788	0.973	**0.986**	**0.949**

TABLE 5. The Performance comparison on five types of distortion in terms of SROCC.

	LIVE 3D Phase I					LIVE 3D Phase II				
	JP2K	JPEG	WN	Blur	FF	JP2K	JPEG	WN	Blur	FF
Zhou'2017	0.856	0.670	0.918	0.929	0.768	0.812	0.773	0.946	0.901	0.860
Wang2018	0.881	0.513	0.944	0.931	0.686	0.953	0.773	0.826	0.770	0.831
Yang'2018	0.903	0.739	0.927	0.862	0.799	0.908	0.793	0.920	0.892	0.908
Yue2018	0.832	0.595	0.932	0.857	0.779	**0.959**	0.769	**0.959**	0.868	0.913
Chen2019	0.835	0.830	0.923	**0.946**	**0.907**	0.832	0.850	0.906	**0.949**	0.875
Liu2019	0.901	**0.962**	0.926	0.891	0.834	0.923	**0.856**	0.954	0.947	0.924
Proposed	**0.912**	0.743	**0.953**	0.901	0.845	0.921	0.788	0.954	0.936	**0.939**

TABLE 6. The overall performance comparison in terms of PLCC, SRCC and RMSE on 3D image quality databases.

	LIVE 3D Phase I			LIVE 3D Phase II			MCL		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Scheme A	0.951	0.939	4.749	0.935	0.930	4.440	0.884	0.876	1.114
Scheme B	0.950	0.933	5.345	0.909	0.901	4.570	0.891	0.882	1.170
Scheme C	0.955	0.942	4.806	0.933	0.928	4.233	0.914	0.903	1.087
Proposed	**0.959**	**0.949**	**4.574**	**0.936**	**0.933**	**3.804**	**0.916**	**0.908**	**0.945**

LIVE 3D Phase I, and (c)(d) is that of LIVE 3D Phase II. Fig.6 presents that the proposed model has a good convergence on both databases and on five types of distortions, which shows the consistence between the subjective score and the objective score. In summary, the proposed model can effectively and accurately evaluate the stereoscopic image affected by different types of distortions, which proves that the proposed metric is independent of the distortion types.

D. COMPARISON AMONG DIFFERENT FEATURES
In our paper, the monocular and binocular visual features are both extracted to build the NR SIQA model. To understand the role of features in the proposed model, we design three different schemes for performance comparison. For scheme A, only the monocular features are considered. For scheme B, only the binocular features are considered. For scheme C, we delete the monocular luminance NSS features of the left view. The corresponding results are listed in Table 6. It can be concluded that only using the monocular or binocular features cannot achieve the best performance. The monocular and binocular features are complementary and can provide a more reasonable and accurate quality evaluation model. Besides, scheme C yields the lower performance

than the proposed model, which indicates that the monocular luminance features of each view play an important role in quality assessment.

E. CROSS-DATABASE PERFORMANCE EVALUATION
Besides, we perform the cross-database evaluation for the proposed model to validate its generality and stability. Since the subjective scores of the LIVE 3D databases are provided in the form of DMOS, and the subjective scores of the MCL 3D database are provided in the form of MOS. So cross-database experiments between LIVE 3D database and MCL 3D database is not appropriate. Therefore, in this paper, we only conducted the cross-test based on LIVE 3D databases. The experiment training samples and testing samples are from different databases, and the experiment procedure is the same as the previous representation in Section III.C. Experimental results are shown in Table 7, and the top two performances are highlighted in boldface.

Table 7 reveals that the overall performances of all the models are lower than the former performances that the training and testing samples are from the same database. Although Yang2018's model, trained on LIVE II and tested on LIVE I, denote LIVE II/LIVE I, achieves better perfor-

TABLE 7. Results of cross-database experiment.

Train/Test	Model	PLCC	SROCC
LIVE I/LIVE II	BRISQUE	0.4961	0.4920
	Zhou2017	NA	0.827
	Yang'2018	0.8289	0.8167
	Liu2019	**0.8669**	**0.8378**
	Proposed	0.8476	0.8281
LIVE II/LIVE I	BRISQUE	0.8428	0.8325
	Zhou2017	0.899	----
	Yang'2018	**0.9096**	**0.9053**
	Liu2019	**0.9192**	**0.9136**
	Proposed	0.8958	0.8943

mance than others, it performs not good at the strategy of LIVE I/LIVE II. Although Liu2019's model holds a good performance in both two cross-database tests, the proposed model holds a competitive and stable performance across the two strategies. Besides, the proposed model under the condition LIVE II/LIVE I performs better than it under LIVE I/LIVE II. The probable reason is that LIVE II has symmetrical and asymmetrical distorted images, and the training sample contains two types of distortions but not in the testing sample. Overall, the results indicate that the proposed model is low dependent on image contents, and presents good robustness and generality.

V. CONCLUSION

This paper proposes a novel stereoscopic image quality assessment model. Human monocular and binocular properties are both considered. Especially, color quality-aware features are extracted to complement the monocular quality index, which is ignored in most previous works. Our binocular visual model adopts a more reasonable visual model to obtain the binocular features, in which the binocular multiple gradient features are first extracted for 3D quality assessment. The difference signal and its features provide a simple way to explain the depth perception with an acceptable computational complexity. Experiments conducted on three popular and public used databases, LIVE 3D Phase I, Phase II and MCL 3D Image Quality Databases, demonstrate that the proposed model yields more robust and consistent with human subjective perception than previous state-of-art models, which proves its effectiveness on evaluating the stereoscopic image quality. In the future, we will focus on utilizing deep learning tools to learn more effective visual perception quality-aware features to predict the stereoscopic image quality.

ACKNOWLEDGMENT

REFERENCES

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

[2] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE, Int. Soc. Opt. Eng.*, vol. 6803, Feb. 2008, Art. no. 680305.

[3] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 1–13, 2008.

[4] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1143–1155, Oct. 2013.

[5] R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, Jun. 2013.

[6] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.

[7] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.

[8] F. Shao, W. Chen, G. Jiang, and Y.-S. Ho, "Modeling the perceptual quality of stereoscopic images in the primary visual cortex," *IEEE Access*, vol. 5, pp. 15706–15716, 2017.

[9] 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, Aug. 2014.

[10] 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.

[11] 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.

[12] W. Zhou, S. Zhang, T. Pan, L. Yu, W. Qiu, Y. Zhou, and T. Luo, "Blind 3D image quality assessment based on self-similarity of binocular features," *Neurocomputing*, vol. 224, pp. 128–134, Feb. 2017.

[13] J. Yang, Y. Liu, Z. Gao, R. Chu, and Z. Song, "A perceptual stereoscopic image quality assessment model accounting for binocular combination behavior," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 138–145, Aug. 2015.

[14] Y. Liu, F. Kong, and Z. Zhen, "Toward a quality predictor for stereoscopic images via analysis of human binocular visual perception," *IEEE Access*, vol. 7, pp. 69283–69291, 2019.

[15] J. Yang, K. Sim, B. Jiang, and W. Lu, "No-reference stereoscopic image quality assessment based on hue summation–difference mapping image and binocular joint mutual filtering," *Appl. Opt.*, vol. 57, no. 14, pp. 3915–3926, May 2018.

[16] 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.

[17] F. Shao, Z. Zhang, Q. Jiang, W. Lin, and G. Jiang, "Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 573–585, Nov. 2018.

[18] 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.

[19] Y. Yao, L. Shen, and P. An, "Bivariate analysis of 3D structure for stereoscopic image quality assessment," *Signal Process., Image Commun.*, vol. 65, pp. 128–140, Jul. 2018.

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

[21] L. Chen and J. Zhao, "Perceptual quality assessment of stereoscopic images based on local and global visual characteristics," *Multimedia Tools Appl.*, vol. 78, no. 9, pp. 12139–12156, May 2019.

[22] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images," *ACM Trans. Appl. Percept.*, vol. 3, no. 2, pp. 95–109, Apr. 2006.

[23] Y. Lv, M. Yu, G. Jiang, F. Shao, Z. Peng, and F. Chen, "No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network," *Signal Process., Image Commun.*, vol. 47, pp. 346–357, Sep. 2016.

[24] J. Wang, S. Wang, K. Ma, and Z. Wang, "Perceptual depth quality in distorted stereoscopic images," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1202–1215, Mar. 2017.

[25] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder," *Pattern Recognit.*, vol. 76, pp. 242–255, Apr. 2018.

[26] 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.

[27] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

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

[29] Y. Zhan and R. Zhang, "No-reference JPEG image quality assessment based on blockiness and luminance change," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 760–764, Jun. 2017.

[30] W. Zhou, L. Yu, W. Qiu, Y. Zhou, and M. Wu, "Local gradient patterns (LGP): An effective local-statistical-feature extraction scheme for no-reference image quality assessment," *Inf. Sci.*, vols. 397–398, pp. 1–14, Aug. 2017.

[31] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.

[32] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 732–741.

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

[34] Q. Wang, J. Chu, L. Xu, and Q. Chen, "A new blind image quality framework based on natural color statistic," *Neurocomputing*, vol. 173, pp. 1798–1810, Jan. 2016.

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

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

[37] H. Shao, X. Cao, and G. Er, "Objective quality assessment of depth image based rendering in 3DTV system," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video*, May 2009, pp. 1–4.

[38] Y. Ding and Y. Zhao, "No-reference stereoscopic image quality assessment guided by visual hierarchical structure and binocular effects," *Appl. Opt.*, vol. 57, no. 10, p. 2610, Apr. 2018.

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

[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] 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.

[42] Y. Zhang, J. Wu, X. Xie, L. Li, and G. Shi, "Blind image quality assessment with improved natural scene statistics model," *Digit. Signal Process.*, vol. 57, pp. 56–65, Oct. 2016.

[43] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 721–734, Feb. 2018.

[44] K. J. Ciuffreda, "Binocular vision and stereopsis," *Optometry Vis. Sci.*, vol. 75, no. 5, p. 310, May 1998.

[45] H. R. Silva and S. H. Bartley, "Summation and subtraction of brightness in binocular perception," *Brit. J. Psychol. Gen. Sect.*, vol. 20, no. 3, pp. 241–250, Jan. 1930.

[46] Z. Li and J. J. Atick, "Efficient stereo coding in the multiscale representation," *Netw., Comput. Neural Syst.*, vol. 5, pp. 157–174, Oct. 1993.

[47] F. A. A. Kingdom, "Binocular vision: The eyes add and subtract," *Current Biol.*, vol. 22, no. 1, pp. R22–R24, Jan. 2012.

[48] K. A. May, L. Zhaoping, and P. B. Hibbard, "Perceived direction of motion determined by adaptation to static binocular images," *Current Biol.*, vol. 22, no. 1, pp. 28–32, Jan. 2012.

[49] K. A. May and L. Zhaoping, "Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns," *Current Biol.*, vol. 26, no. 12, pp. 1571–1576, Jun. 2016.

[50] L. Zhaoping, *Understanding Vision: Theory, Models, and Data*. New York, NY, USA: Oxford Univ. Press, 2014.

[51] 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.

[52] G. R. Engel, "The visual processes underlying binocular brightness summation," *Vis. Res.*, vol. 7, nos. 9–10, pp. 753–767, Sep. 1967.

[53] D. W. Curtis and S. J. Rule, "Binocular processing of brightness information: A vector-sum model," *J. Exp. Psychol., Hum. Perception Perform.*, vol. 4, no. 1, pp. 132–143, Jun. 2006.

[54] A. I. Cogan, "Human binocular interaction: Towards a neural model," *Vis. Res.*, vol. 27, no. 12, pp. 2125–2139, Jan. 1987.

[55] J. Ding and G. Sperling, "A gain-control theory of binocular combination," *J. Vis.*, vol. 2, no. 7, p. 327, Mar. 2010.

[56] S. Grossberg and F. Kelly, "Neural dynamics of binocular brightness perception," *Vis. Res.*, vol. 39, no. 22, pp. 3796–3816, Nov. 1999.

[57] J. Yang, B. Jiang, Y. Wang, W. Lu, and Q. Meng, "Sparse representation based stereoscopic image quality assessment accounting for perceptual cognitive process," *Inf. Sci.*, vols. 430–431, pp. 1–16, Mar. 2018.

[58] B. G. Cumming and A. J. Parker, "Responses of primary visual cortical neurons to binocular disparity without depth perception," *Nature*, vol. 389, no. 6648, pp. 280–283, Sep. 1997.

[59] B. G. Cumming and A. J. Parker, "Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque," *J. Neurosci.*, vol. 20, no. 12, pp. 4758–4767, Jun. 2000.

[60] J. Sponring, "The entropy of scale-space," in *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 1, 1996, pp. 900–904.

[61] Y. Li, F. Yang, W. Wan, J. Wang, M. Gao, J. Zhang, and J. Sun, "No-reference stereoscopic image quality assessment based on visual attention and perception," *IEEE Access*, vol. 7, pp. 46706–46716, 2019.

[62] 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.

[63] 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.

[64] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2281–2284.

[65] L. M. Hurvich, "An opponent-process theory of color vision," *Psychol. Rev.*, vol. 64, no. 6, pp. 384–404, 1957.

[66] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 15, no. 8, p. 2036, Aug. 1998.

[67] N. Qian and S. Mikaelian, "Relationship between phase and energy methods for disparity computation," *Neural Comput.*, vol. 12, no. 2, pp. 279–292, 2000.

[68] P. B. Hibbard and S. Bouzit, "Binocular energy responses to natural images," *J. Vis.*, vol. 6, no. 6, p. 833, Mar. 2010.

[69] 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.

[70] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[71] R. Song, H. Ko, and C. C. J. Kuo, "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source," *J. Inf. Sci. Eng.*, vol. 31, no. 5, pp. 1593–1611, 2014.

[72] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.

[73] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, "Learning sparse representation for no-reference quality assessment of multiply distorted stereoscopic images," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1821–1836, Aug. 2017.

[74] L. Liu, B. Yang, and H. Huang, "No-reference stereopair quality assessment based on singular value decomposition," *Neurocomputing*, vol. 275, pp. 1823–1835, Jan. 2018.

[75] 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.

YUN LIU received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, China.

She was a Visiting Ph.D. Student with the School of Optometry, University of California at Berkeley, USA. She is currently a Lecturer with the College of Information, Liaoning University. Her research interests include stereo vision research, pattern recognition, and image quality evaluation.

ZHI ZHENG is currently pursuing the Ph.D. degree with the Department of Electronic and Information Engineering, Beijing Jiaotong University.

His research interests focus on image processing and pattern recognition.

BAOQING HUANG is currently pursuing the M.S. degree with the College of Information, Liaoning University.

His research interests focus on stereo vision research, image processing, and pattern recognition.

WEIQING YAN received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, China.

She was a Visiting Ph.D. Student with the School of Optometry, University of California at Berkeley, USA. She is currently a Lecturer with the School of Computer and Control Engineering, Yantai University. Her research interests include image quality evaluation, stereo vision research, and machine learning.

HONGWEI YU is currently pursuing the M.S. degree with the College of Information, Liaoning University.

His research interests include image quality evaluation and pattern recognition.

• • •