IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 3, 2019, accepted May 20, 2019, date of publication May 27, 2019, date of current version June 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919155

Toward a Quality Predictor for Stereoscopic Images via Analysis of Human Binocular Visual Perception

YUN LIU [1], FANHUI KONG [2], AND ZHIZHUO ZHEN [3]

[1] College of Information, Liaoning University, Liaoning 110036, China
[2] School of Management, Tianjin University of Technology, Tianjin 300384, China
[3] School of Artificial Intelligent, Hebei University of Technology, Tianjin 300401, China

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

ABSTRACT Perceptual stereoscopic image quality assessment (SIQA) has become a challenge research problem due to the poor understanding of human binocular visual characteristics. For the task of SIQA, an intuitive idea is to develop effective models on the basis of the image content and depth perception. In this paper, we propose a full-reference objective quality evaluator for stereoscopic images by simulating binocular behaviors of the human visual system (HVS): Binocular interaction and depth perception. This model is based on a cyclopean image from a novel binocular combination model as image content quality description and a depth binocular combination model from a depth synthesized procedure as depth perception description. The final quality score of the distorted stereoscopic images is calculated by integrating the above two perception indicators. The experimental results on two stereoscopic image quality databases demonstrate that our proposed metric works efficiently for both symmetric and asymmetric distortions and achieves high consistent alignment with subjective observations.

INDEX TERMS Binocular combination, cyclopean image, depth perception, stereoscopic images.

I. INTRODUCTION

Image quality assessment (IQA) is an important and challenging research problem. During the past few decades, a raising number of IQA methods have been studied to predict the image quality via mathematical models [1]–[4]. Compared with two-dimensional (2D) image, three-dimensional (3D) image provides the cues of depth perception, which makes 3D IQA difficult. According to the availability of perfect image as the reference information, three categories of stereoscopic images quality assessment (SIQA) methods are classified: no reference (NR), reduced reference (RR), and full reference (FR) SIQA. Without considering any reference information, NR SIQA models are proposed [5]–[10]. Besides, based on part of reference information, RR SIQA methods [11]–[15] are built for stereoscopic image quality prediction. Different from RR and NR methods, FR methods, requiring the complete reference information, are widely developed during the last decade due to their efficient results [16], [17].

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

In this paper, we mainly focus on FR SIQA method which can be divided into three categories. Since stereoscopic images consist of two views, the first category is to apply the 2D IQA algorithm to each view directly to get the final evaluation score [18], [19]. However, since the depth information exists in stereoscopic images, the above methods are not validated. By combining the depth information, the second category of SIQA model is extensively studied in the literature. Benoit *et al.* [20] and Campisi *et al.* [21] applied a straightforward way to evaluate the image quality based on 2D quality index and depth map, which shows a good performance. Other similar works also have been conducted in [22], [23]. However, even with the depth information considered in the above SIQA models, they cannot obtain higher accuracy in prediction performance based on 2D metrics. The third category of SIQA metric is then proposed by considering binocular perception properties [24]–[26]. Binocular perception properties are the cyclopean mechanisms to percept the "cyclopean image" created from two eyes. Chen *et al.* [27] created a "cyclopean image" to model human binocular rivalry behavior and then proposed a FR stereoscopic images quality estimator. Bensalma and Larabi [28] proposed a binocular energy quality metric to obtain the perceived image quality by using the Complex Wavelet Transform (CWT). Lin and Wu [29] constructed an effective evaluation model by the frequency binocular integrated model. Focus on the characteristics of receptive fields, Shao *et al.* [30] applied the phase energy and complexity of the binocular information to calculate the quality score of the stereoscopic images. Later, they [31] proposed a novel perceptual quality assessment approach for stereoscopic images by modeling visual properties of the primary visual cortex. Besides, Yao *et al.* [32] proposed a bivariate-based model to capture image quality by extracting features from binocular and monocular perception regions, respectively. By considering the important of the color information in human visual binocular perception, Xu *et al.* [25] proposed a parts-based SIQA method by learning manifold color visual properties.

For SIQA, various factors should be considered, since symmetric or asymmetric distortions will involve binocular confusion, depth perception discomfort, and so on. Therefore, SIQA should at least account for binocular perception and depth perception. As mentioned above, previous works indicated that human binocular characteristics are the key component in stereoscopic images quality assessment. However, these models didn't consider the disparity issue and couldn't explain how the depth perception is formed in human brain for the binocular combination behavior. In this work, we both address the image quality perception and depth perception based on the cyclopean mechanism and depth binocular property to simulate 3D scene understanding to get a more accurate quality assessment metric. Firstly, based on human binocular characteristics, we apply the bank of Gabor filter to obtain the local energies of response for the cyclopean image, and build a combined depth perception index based on the contrast energy and signal strength from the two eyes. The binocular image and the depth perception model are obtained respectively based on the above two quality indexes, which can achieve high consistent alignment with subjective assessment and improve the prediction accuracy of the model. The main advantages of our paper are as follows: 1) The binocular combination model we applied in this paper is easy to implement and replaceable. 2) A combined depth perception index is proposed for stereoscopic images quality evaluation based on binocular contrast gain-control model, which presents a promising way to compute the depth perception quality with low computational complexity.

The remainder of the paper is organized as follows: in Section II, we briefly introduce the related background. Section III provides the proposed metric step by step. In section IV, we discuss the experimental results. Finally, Section V summarizes this paper.

II. BACKGROUND

Since human beings are the final receiver to judge image quality, human visual system (HVS) characteristics, such as binocular perceptual mechanism, have aroused lots of focus [33], [34]. To investigate how human eyes process the image information and get the 3D scene in human brain, it needs to clearly understand the properties and limitations of HVS. In the past years, various HVS property models have been discussed, but how human eyes send visual information from the two views of the stereoscopic images to human brain and how human get the depth perception are still confused in vision science, which has aroused lots of psychophysical and physiological researchers to investigate. In recent years, two popular types of human binocular visual perception model are described to analysis human visual process: single-channel and double-channel models. Many plausible models have been proposed to explain how human eyes combine two views information together to generate the depth perception in human brain. Here we briefly describe the existing findings about human brain vision models, as follows.

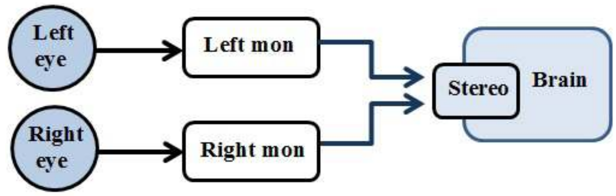


FIGURE 1. The flowchart of the single-channel model.

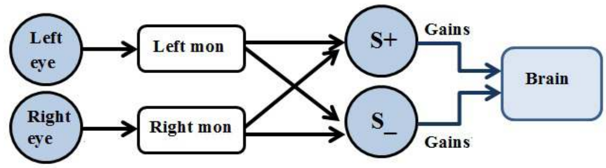


FIGURE 2. The flowchart of the double-channel model.

A. SINGLE-CHANNEL MODEL

HVS is self-adapted and sensitive to the stimulation of the external light signal [35]. Left (L) and right (R) eyes monocular (mon) neurons firstly detect the information of each view simultaneously and then combine them together to obtain the stereopsis [36], as illustrated in Fig.1. In general, human eyes collect the visual signals through the photoreceptors in the retina or each eye separately, and feed the response into the disparity sensitive neurons to generate the unique perspective in human brain.

B. DOUBLE-CHANNEL MODEL

An alternative view, however, suggests that there exist two separate pathways in binocular combination used for stereopsis. Silva and Bartley [37] investigated the summation and subtraction of brightness in binocular perception and demonstrated the existence of double channel mechanism. Besides, Li and Atick [38] explained how the double channel works (shown in Fig.2) with two view signals after gain control. They indicated that the visual gain information is needed to optimize the interocular correlation.

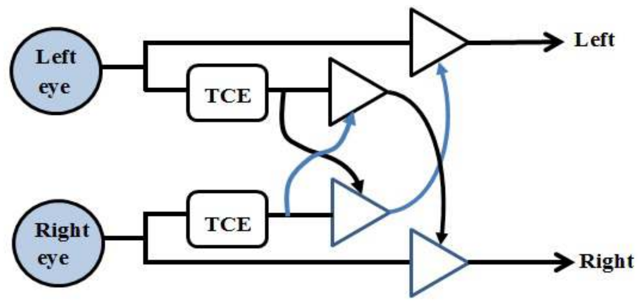


FIGURE 3. The flowchart of the binocular combination model. TCE means the total visually weighted contrast energy.

Compared with the single-channel model, the double channel model plays an important role in determining the reasonable human combination process and can well explain human binocular behaviors. There are many effective visual models that can successfully reflect how a cyclopean image fused based on the monocular image [39]–[41]. In [42], Ding and Sperling conducted a binocular vision experiment to study the appearance of the binocular image and developed a gain-control (GC) model by using sine-wave gratings with different phases and contrasts, which can well model human binocular phase perception. Meanwhile, a binocular combination model, shown in Fig.3, was developed to investigate human double channel mechanism [43]. Motivated by the interaction between visual properties and quality assessment, in this paper, we apply the double channel model to build the cyclopean perception in SIQA model to show the binocular interaction.

C. STEREO DEPTH PERCEPTION

Depth perception is the ability of human to perceive the object in 3D space and help our brain to judge its space distance. Stereo depth perception involves a complicated visual process and hard to interpret how human eye works. Many depth perception works have been conducted using the monocular vision cues [44], [45]. In work [46], a cross-correlation model was proposed to predict the depth perception qualitatively. This model gives an introduction of the interocular correlation and stereoacuity by using the signal strength in the cyclopean domain. Later, Cormack *et al.* [47] investigated the relationship among stereo paradox, contrast, and spatial frequency, and proposed a hybrid model to explain how the stereoscopic system works:

$$D \propto \sqrt{\frac{1}{w_L C_L} + \frac{1}{w_R C_R} - \frac{2\rho}{\sqrt{w_L C_L w_R C_R}}}. \tag{1}$$

where D is the depth threshold. C_L and C_R are contrasts of the left and right images, respectively. ρ is the correlation parameter in the two eyes. $w_L = C_L^P / C_L^P + KC_R^P$ and $w_R = C_R^P / C_R^P + KC_L^P$. p is the exponent of the power function.

Hou *et al.* [48] proposed a multipathway contrast gain-control model to account for stereo depth, which gives a more plausible way to explain how human eyes to generate the depth perception. By using the above combination perception model, we build the other component of the quality index in our proposed model.

III. THE PROPOSED METRIC

As discussed in previous sections, human binocular behaviors reflect how our brain perceives image information, whereas depth combination process enables us to perceive the disparity information. Therefore, it is valid to combine the cyclopean image process and depth model together to conduct the quality assessment. With this inspiration, an efficient cyclopean perception algorithm for stereoscopic images quality assessment is proposed, as shown in Fig.4.

A. CYCLOPEAN IMAGE MODEL

Our previous study [49] has shown that the binocular model can simulate the procedure of visual perception by using the above physiological discoveries in Section II.B. Since GC model proposed in [42] can well model human binocular phase perception, such as binocular fusion and rivalry, so in this paper we adopt the GC model to obtain the cyclopean image, as follows:

$$C = \frac{1+E_L}{1+E_L+E_R} I_L + \frac{1+E_R}{1+E_L+E_R} I_R. \tag{2}$$

where C is the cyclopean image. I_L and E_L are the left view and the left visually weighted energy response. I_R and E_R are the right view and the right visually weighted energy response.

It needs to mention that the above cyclopean image model is defined based on sine-wave gratings, in which the relative amplitudes play an important role in the procedure of binocular combination. Therefore, we focus on the amplitude information during proposing the cyclopean image index. Firstly, we extract the amplitude information for further processing by transforming the reference stereoscopic images and the distorted stereoscopic images to LAB color space, respectively. Then the cyclopean image can be obtained based on the GC model proposed in work [42].

B. COMBINED DEPTH PERCEPTION MODEL

In work [48], Hou *et al.* gave us a plausible way to measure the disparity threshold for depth perception. They found that the depth threshold shows the binocular contrast GC properties, and has the same front-end gain control procedure as the binocular contrast perception. The extended MCM model in [48] provided a good mathematic model to calculate the depth perception D based on the signal strength and the GC model:

$$D = \frac{1}{\left(L\frac{1}{1+\frac{E_R}{1+E_L}}\right)\left(R\frac{1}{1+\frac{E_L}{1+E_R}}\right)}. \tag{3}$$

where L and E_L are the contrast of the left image and the left visually weighted contrast energy, respectively. R and E_R are

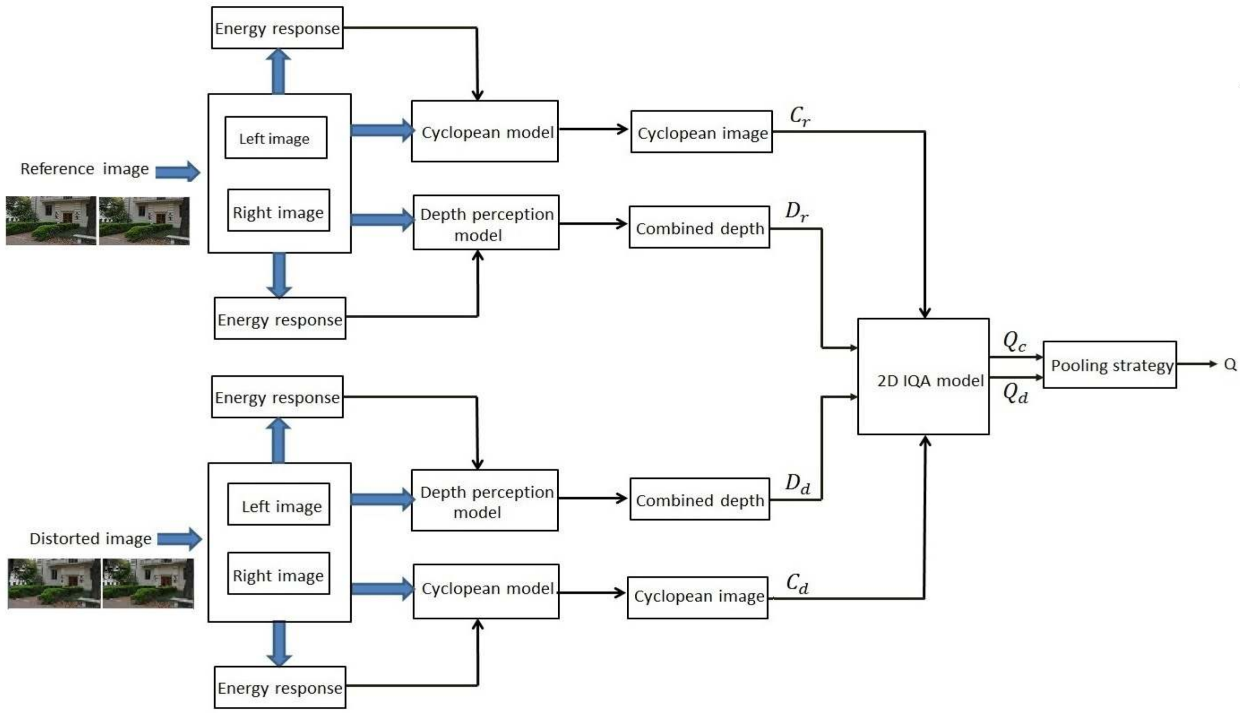


FIGURE 4. The framework of the proposed metric.

the contrast of the right image and the right visually weighted contrast energy, respectively.

In work [50], the depth perception mechanism was investigated based on monocular images with different contrast and phases. The experimental results indicated the existence of contrast GC model in direct and indirect interocular inhibition, and concluded a contrast GC cyclopean mechanism based on the binocular interaction. Therefore, we take the depth combination model as the factor of depth perception information instead of calculating depth map, which has the advantage of efficient computing time.

Hibbard [51] indicated that Gabor filter can well model human eye's receptive field, and log-Gabor can be constructed with arbitrary bandwidth [52]. Thus, we adopt the log-Gabor filter to extract monocular responses, and obtain the energy responses of two eyes like our previous paper [49]. More details can refer to [53]. Define the spatial scale as s, and the orientation index is o, then the log-Gabor filter with the radial frequency $\boldsymbol{\omega}$ and orientation angle $\boldsymbol{\theta}_0$ can be formulated as follows:

$$G_{s,o}(\boldsymbol{\omega},\boldsymbol{\theta})=exp[-\frac{(log(\boldsymbol{\omega}/\boldsymbol{\omega}_s))^2}{2\sigma_s^2}]\times exp[-\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2}{2\sigma_o^2}]. \quad (4)$$

where $\boldsymbol{\sigma}_s$ and $\boldsymbol{\sigma}_0$ determine the filter's strengths. $\boldsymbol{\omega}_s$ is the center frequency. It's out of the scope to study the impacts of these parameters on the Gabor filter in this paper, so we sets these parameters by strictly following reference [27].

Denote the responses of log-Gabor on scale s and along orientation o as $[\boldsymbol{\eta}_{s,o}, \boldsymbol{\zeta}_{s,o}]$, then the final local energy of log-Gabor filter responses at location X can be computed as follows:

$$E = \boldsymbol{sum}\sqrt{(\sum_s \eta_{s,o}(X))^2 + (\sum_s \zeta_{s,o}(X))^2}. \quad (5)$$

Hence, the log-Gabor filter response combined with the above cyclopean model can synthesize the cyclopean images, and that combined with Hou's [48] depth perception model can synthesize the corresponding combined depth perception model. The final quality of the stereoscopic image can be obtained by integrating the qualities of the cyclopean image and the combined depth perception model together with a pooling strategy.

C. POOLING STRATEGY

Based on the above steps, we denote the cyclopean image of the reference stereoscopic images as C_r, and that of the distorted stereoscopic images as C_d. We denote the combined depth perception model of reference stereoscopic images as D_r, and that of the distorted stereoscopic images as D_d. A pooling strategy is then applied to C_r and C_d to calculate the quality index of the cyclopean image. There are many pooling strategies, such as PSNR, SSIM [54], MS-SSIM [55], and ADD-GSIM [56]. In this section, we validate the MS-SSIM pooling strategy. Denote Q_C as the quality of the cyclopean image, it can be calculated based on the similarity between C_r and C_d, shown as follows:

$$\mathrm{Q}_c = \mathrm{MS-SSIM}\left(\mathrm{C}_r, \mathrm{C}_d\right). \quad (6)$$

TABLE 1. Performances of the proposed SIQA method and other eleven methods in terms of PLCC, SROCC, and RMSE on LIVE 3D Database Phase I and Phase II (cases in bold denote best performance).

	SSIM	MS-SSIM	ADD-GSIM	Benoit	You	Chen	Bensalma	Shao[30]	Lin	Shao[31]	Proposed
						LIVE 3D	Phase I				
PLCC	0.8725	0.9261	0.9388	0.8899	0.9303	0.9167	0.8874	0.9350	0.8721	0.9389	**0.9430**
SROCC	0.8767	0.9239	0.9335	0.8901	0.9247	0.9157	0.8747	0.9251	0.8765	0.9308	**0.9402**
RMSE	10.6841	8.2486	7.5337	7.4786	6.0161	8.7385	7.5585	5.8155	8.0236	5.6459	**5.4238**
						LIVE 3D	Phase II				
PLCC	0.7467	0.7824	0.7359	0.7642	0.7744	0.9010	0.7699	0.8628	0.6844	**0.9263**	0.8417
SROCC	0.7292	0.7774	0.7304	0.7475	0.7206	0.8930	0.7513	0.8494	0.6795	**0.9282**	0.8307
RMSE	8.0074	8.0217	7.6422	7.2806	7.1413	10.5800	7.2035	5.7058	8.2295	**4.1996**	6.0946

Likewise, the depth perception quality index Q_d can be measured based on the combined depth perception model D_r and D_d, shows as follows:

$$Q_d = MS - SSIM\,(D_r, D_d)\,. \tag{7}$$

The final quality index of the stereoscopic images Q can be derived by combing Q_C and Q_d these two quality scores together. A linear weighting model is applied to integrate the above two quality scores to obtain the final score, shown as follows:

$$Q = \omega_1 \times Q_c + \omega_2 \times Q_d. \tag{8}$$

where $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are the weights of each synthesized map, respectively, and $\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 = 1$.

IV. EXPERIMENTAL RESULTS

A. DATABASE DESCRIPTION

The comparison experiments are conducted on two databases from University of Texas at Austin [57]: LIVE 3D Database Phase I and LIVE 3D Database Phase II. The database Phase I consists of 365 reference images generated from 20 natural content 3D images, and 365 distorted images by introducing five types of distortions to the reference images symmetrically at various levels (80 for JP2K, JPEG, WN, and FF respectively; 45 for Blur). For the database Phase II, it consists of 360 reference images generated from 8 natural content 3D images, and 360 distorted images (240 asymmetrically distorted and 120 symmetrically distorted). The distortion types are the same as that in Phase I.

B. EXPERIMENTAL PROTOCOLS

For the non-liner regression, we apply a logistic function with five parameters ($\beta_1, \beta_2, \beta_3, \beta_4, and\,\beta_5$) to map The Difference Mean Opinion Score ($DMOS$) based on the recommendation from VQEG [58]:

$$f(x) = \beta_1 \times \left[\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2\left(x - \beta_3\right)\right)}\right] + \beta_4 x + \beta_5. \tag{9}$$

where x is the predicted score. $f(x)$ is the mapped score.

To estimate the performance of the proposed SIQA method, three correlation coefficients are adopted to estimate the linear correlation between the objective score and the subjective score: PLCC (Pearson linear correlation coefficient) in (10), SROCC (Spearman rank order correlation coefficient) in (11), and RMSE (root mean squared error) in (12). When PLCC and SROCC are close to 1, and RMSE is close to 0, it indicates better fit to the data, which means the better assessment method.

$$PLCC(X,Y) = \frac{\sum_{i=1}^{N}(o_i - \overline{o})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(o_i - \overline{o})^2 \sum_{i=1}^{N}(s_i - \overline{s})^2}}. \tag{10}$$

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N\left(N^2 - 1\right)}. \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(o_i - s_i)^2}{N}}. \tag{12}$$

where N is the total number of the images. $O_i\,(i = 1, 2, \ldots N)$ is the ith X value, and o is the mean value of X. S_i $(i = 1, 2, \ldots N)$ is the ith Y value, and s is the mean value of Y.

To get the values of the parameters in (8), a small set of database Phase I are selected as the train database to optimizing PLCC and RMSE. We find that when $\boldsymbol{\omega}_1 = 0.69$ and $\boldsymbol{\omega}_2 = 0.31$, the proposed model yields the best performance. So we take these two parameters to conduct the further experiment. It should be noted that the weight of the combined depth image is smaller than that of the cyclopean image, which means the combined depth information is critical for SIQA.

C. OVERALL ASSESSMENT PERFORMANCE

In this section, we compare the proposed model with ten existing models: three SQIA models based on 2D IQA model (SSIM, MS-SSIM, and ADD-GSIM), You's scheme [19], Benoit's method [20], Chen's scheme [27], Bensalma's scheme [28], Lin's method [29], Shao's method [30] in 2015, and Shao's method [31] in 2017. The experimental results are listed in Table 1, where the best metric has been highlighted in boldface. Table 1 indicates that the 2D IQA- based models

TABLE 2. Performance comparison of nine metrics on each distortion type in terms of PLCC.

		Benoit	*You*	Chen	Bensalma	Shao[30]	Lin	Shao[31]	Proposed
LIVE 3D Phase I	JPEG	0.5597	0.6333	0.6344	0.3803	0.5200	**0.6654**	**0.6654**	**0.7315**
	JP2K	0.8897	**0.9410**	0.9163	0.8389	0.9213	0.8381	0.9360	**0.9423**
	WN	0.9360	0.9351	0.9432	0.9147	**0.9448**	0.9280	0.9441	**0.9463**
	Blur	0.9256	**0.9545**	0.9416	0.9369	**0.9592**	0.8249	0.9542	0. 9530
	FF	0.7514	0.8589	0.7573	0.7339	**0.8594**	0.7086	0.8304	**0.8658**
LIVE 3D Phase II	JPEG	0.5328	0.6741	0.8422	**0.8577**	0.7472	0.6417	0.8506	**0.8758**
	JP2K	0.6467	0.7320	0.8426	0.6670	0.7823	0.7224	**0.8768**	**0.8701**
	WN	0.8610	0.5464	**0.9602**	0.9436	**0.9464**	0.9271	0.9339	0.9325
	Blur	0.8814	**0.9763**	**0.9650**	0.9077	0.9580	0.8417	0.9445	0.9430
	FF	0.8472	0.8561	0.9097	0.9097	0.9046	0.8561	**0.9330**	**0.9218**

TABLE 3. Performance comparison of nine metrics on each distortion type in terms of SROCC.

		Benoit	*You*	Chen	Bensalma	Shao[30]	Lin	Shao[31]	Proposed
LIVE 3D Phase I	JPEG	0.5189	0.5967	0.5582	0.3283	0.4951	0.1960	**0.6339**	**0.6952**
	JP2K	0.8701	0.8979	0.8956	0.8170	0.8945	0.8388	**0.9000**	**0.9040**
	WN	0.9347	0.9389	**0.9481**	0.9055	0.9405	0.9284	0.9430	**0.9468**
	Blur	0.8967	0.9278	0.9261	0.9157	**0.9403**	0.7910	0.9242	**0.9294**
	FF	0.6142	**0.8030**	0.6879	0.6500	0.7963	0.6581	0.7807	**0.8108**
LIVE 3D Phase II	JPEG	0.5078	0.5229	0.8396	**0.8461**	0.7330	0.6789	0.8340	**0.8640**
	JP2K	0.6325	0.7309	0.8334	0.8038	0.7845	0.7027	**0.8747**	**0.8642**
	WN	0.8569	0.4820	**0.9554**	0.9386	**0.9651**	0.9206	0.9325	0.9230
	Blur	0.8545	**0.9227**	0.9096	0.8838	0.9204	0.8358	**0.9241**	0.9114
	FF	0.8319	0.8392	0.8890	0.8743	0.8905	0.8348	**0.9409**	**0.8977**

are simply combined the quality indexes of two images without considering HVS characteristics, so they have limited performances on both two databases. Shao's scheme [31] achieves the best performance for database Phase II, since the model of monocular and binocular visual information can well simulate human primary visual cortex (V1) which correlates well with subjective perception for the asymmetric distortions. It should be noted that the binocular combination properties are also taken into consideration in Chen's scheme, but the combined depth perception is missing. The proposed framework, considering both binocular interaction and depth perception, obtains the best performance on database Phase I and a competitive performance on database Phase II based on these three criteria, which demonstrates the necessity and rationality of the combination of two components: cyclopean image and combined depth perception. Overall, our proposed model is an effective predictor to evaluate the stereoscopic images quality on these two databases.

D. DISTORTION SENSITIVE

To evaluate the predictive performance of our proposed SIQA model in predicting different types of distortion, we compare it with the following seven SIQA models: You's scheme, Benoit's scheme, Chen's scheme, Bensalma's scheme [28], Lin's method [29], Shao's method [30] in 2015, and Shao's method [31] in 2017. Table 2 and Table 3 are present the results of PLCC and SROCC on different types of distortions, where the top two metrics have been highlighted in boldface.

Based on the results shown in Table 2 and Table 3, the proposed model is between the top two metrics 15 times compared with You's scheme (5 times), Chen's scheme (4 times), Bensalma's scheme [28] (2 times), Shao's scheme [30] (6 times), Lin's scheme (1 time), and Shao's scheme in [31] (8 times), which explains that the proposed SIQA method can well reflect human perception. Besides, for the LIVE 3D Phase I, the proposed SIQA algorithm yields perfect performance on JPEG, JP2K, WN, and FF distortions, respectively, which indicates the distortion sensitivity of our model. Although the overall performance of our model is not very prominent for the LIVE 3D Phase II, the values of PLCC and SROCC of the proposed model are close to the highest values, which indicates that the proposed model can achieve great consistent performance for these two databases.

Fig.5 shows the scatter plots of DMOS versus the objective scores based on our proposed model and three SIQA metrics (Benoit's, Chen's, and Lin's schemes) on Phase I, in which the proposed model achieves a better convergence than other models. In general, we can conclude that the SIQA metric proposed in this paper can be effectively applied to evaluate the quality of stereoscopic images contaminated by different types of distortion.

E. COMPONENT EVALUATION

Since the proposed model is composed by two quality indexes: the cyclopean image quality (CIQ) and the combined depth perception quality (CDQ), it is necessary to testify the necessity of the combination of two components. To this end, we test the performances of the methods separately using CIQ and CDQ on Phase I and Phase II. The experimental results are listed in Table 4. It can be concluded

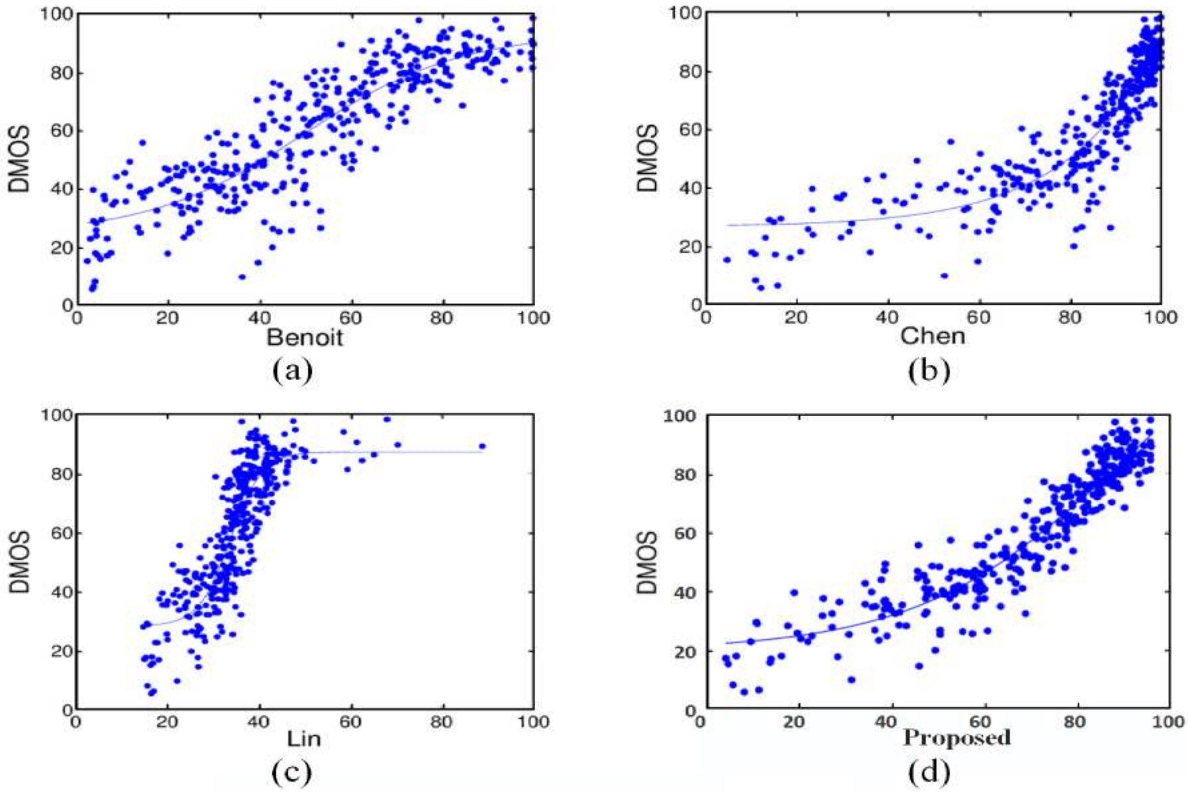


FIGURE 5. Scatter plots of predicted objective scores versus DMOS for the SIOA schemes. (a) Scheme in [20]. (b) Scheme in [27]. (c) Scheme in [29]. (d) Proposed scheme.

TABLE 4. Comparison of PLCC, SROCC, and RMSE for different schemes (the case in bold: the best performance).

	LIVE 3D Phase I			LIVE 3D Phase I		
Metric	CIQ	CDQ	Proposed	CIQ	CDQ	Proposed
PLCC	0.9205	0.9184	**0.9437**	0.8185	0.8000	**0.8417**
SROCC	0.9173	0.9148	**0.9402**	0.8005	0.7910	**0.8307**
RMSE	6.4064	6.4862	**5.4238**	6.4849	6.7722	**6.0943**

that the metric integrating both components achieves best performance on both symmetric distortion and asymmetric distortion. This indicates the necessity and rationality of the combination of two components, which reflect different aspects of human visual perception. Moreover, the method only using CIQ performs better than that only using CDQ, which indicates that CIQ make more contribution to SQIA than CDQ. This also explain why ω_1 is bigger than ω_2 in (8). Although the weight of the combined depth quality CDQ is weaker than that of the cyclopean image quality CIQ, it reveals that the combined depth perception is very critical for stereoscopic image quality assessment.

Fig.6 lists the experimental results of PLCC and SROCC of CIQ and CDQ models on each type of distortion on Phase I, which shows that the proposed model yields the best performance across the five types of distortion. In summary, the cyclopean perception and combined depth perception both play an important role in quality assessment. The most effective way to conduct the image quality is combing them together as the proposed model, which has proved to be an effective and accuracy tool in SIQA across both symmetric and asymmetric distortions.

F. INFLUENCE OF THE CYCLOPEAN MODEL

The proposed model in this paper adopts the Gain-Control (GC) model to obtain the cyclopean image. In order to study the influence of the cyclopean model on the overall performance, we replace the GC model in our proposed model with other three cyclopean models: Eye-Weighting (EW) model [39], Vector Summation (VS) model [40], and Neural Network (NN) model [41]. Denote the corresponding SIQA model as Q-EW, Q-VS and Q-NN, respectively, the overall performances of each SIQA model are listed in Table 5. The results indicate that the performance is very close to the proposed model, which proves that the above three cyclopean models can also effectively applied to conduct the quality evaluation, and the proposed model is not sensitive to the form of the cyclopean model.

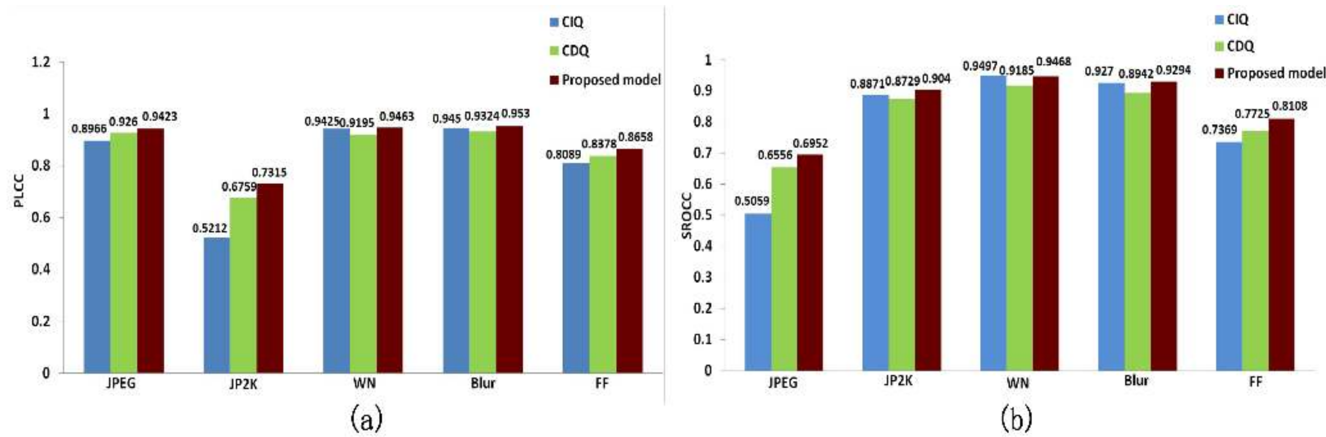


FIGURE 6. Performances of CIQ, CDQ, and the proposed model in terms of PLCC and SROCC on LIVE 3D Phase I. (a) Performance comparison in terms of PLCC. (b) Performance comparison in terms of SROCC.

TABLE 5. Comparison of different cyclopean models(the case in bold: the best performance).

	Metric	PLCC	SROCC	RMSE
LIVE	Q-EW	0.9420	0.9390	5.5029
3D	Q-VS	0.9430	0.9399	5.4550
Phase I	Q-NN	0.9312	0.9312	5.9777
	Proposed	**0.9437**	**0.9402**	**5.4238**
LIVE	Q-EW	0.8441	0.8334	6.0524
3D	Q-VS	**0.8468**	**0.8365**	**6.0033**
Phase I	Q-NN	0.7510	0.7458	7.4532
	Proposed	0.8417	0.8307	6.0946

V. CONCLUSION

In this paper, we have introduced a full-reference SIQA evaluator by considering human binocular interaction and combined depth perception. The main contributions of this work are: 1) we derive a quality prediction model based on image content quality and combined depth perception to account for human binocular characteristics; 2) a combined depth image is extracted by using the signal strengths after gain control to quantify the depth perception; 3) as the final quality index only contains binocular combination operations, our metric holds a low computational complexity. Experimental results indicate that the proposed model has high consistency with subjective assessment across the symmetric and asymmetric databases. In the future, other binocular visual mechanism should be explored and considered in SIQA method.

ACKNOWLEDGMENT

The author would like to thank Prof. Alan C. Bovik for providing the LIVE 3D IQA Database.

REFERENCES

[1] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, Sep. 2017.

[2] K. Gu, J. Zhou, J. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

[3] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.

[4] B. Jiang, J. Yang, Q. Meng, B. Li, and W. Lu, "A deep evaluator for image retargeting quality by geometrical and contextual interaction," *IEEE Trans. Cybern.*, to be published.

[5] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 341–353, Jun. 2018.

[6] L. Liu, B. Yang, and H. Huang, "No-reference stereopair quality assessment based on singular value decomposition," *Neurocomputing*, vol. 275, no. 31, pp. 1823–1835, Jan. 2018.

[7] 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.

[8] F. Shao, Q. Yuan, W. Lin, and G. Jiang, "No-reference view synthesis quality prediction for 3D videos based on color–depth interactions," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 659–674, Mar. 2018.

[9] 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.

[10] 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.

[11] F. Qi, D. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2338–2344, Dec. 2015.

[12] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1077–1084, Jun. 2016.

[13] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh, "No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models," *Signal Process., Image Commun.*, vol. 54, pp. 107–117, May 2017.

[14] F. Li, F. Shao, Q. Jiang, R. Fu, G. Jiang, and M. Yu, "Local and global sparse representation for no-reference quality assessment of stereoscopic images," *Inf. Sci.*, vol. 422, pp. 110–121, Jan. 2018.

[15] K. Gu, V. Jakhetiya, J.-F. Qiao, X. Li, W. Lin, and D. Thalmann, "Model-based referenceless quality metric of 3D synthesized images using local image description," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 394–405, Jan. 2018.

[16] G. Jiang, H. Xu, M. Yu, T. Luo, and Y. Zhang, "Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 269–279, Jul. 2017.

[17] J. Ma, P. An, L. Shen, and K. Li, "Full-reference quality assessment of stereoscopic images by learning binocular visual properties," *Appl. Opt.*, vol. 56, no. 29, pp. 8291–8302, 2017.

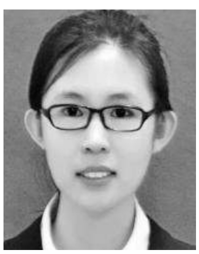
[18] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, Feb. 2008, Art. no. 680305.

[19] 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. 5th Int. Workshop Video Process. Qual. Metrics Consum. Electron. (VPQM)*, vol. 9, Jan. 2010, pp. 1–6.

[20] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, Dec. 2008, Art. no. 659024.
[21] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. 15th Eur. Signal Process. Conf.*, vol. 32, Sep. 2007, pp. 2110–2114.
[22] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "A perceptual quality metric for stereoscopic crosstalk perception," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 4033–4036.
[23] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. B. Akar, "Towards compound stereo-video quality metric: A specific encoder-based framework," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Denver, CO, USA, Mar. 2006, pp. 218–222.
[24] Y. Zhang and D. M. Chandler, "3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3810–3825, Nov. 2015.
[25] H. Xu, M. Yu, T. Luo, Y. Zhang, and G. Y. Jiang, "Parts-based stereoscopic image assessment by learning binocular manifold color visual properties," *Proc. SPIE*, vol. 25, no. 6, Oct. 2016, Art. no. 061611.
[26] X. Geng, L. Shen, K. Li, and P. An, "A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property," *Signal Process., Image Commun.*, vol. 52, pp. 54–63, Mar. 2017.
[27] 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.
[28] 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.
[29] 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.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] J. Yang, K. Sim, W. Lu, and B. Jiang, "Predicting stereoscopic image quality via stacked auto-encoders based on stereopsis formation," *IEEE Trans. Multimedia*, to be published.
[35] A. Boev, M. Poikela, A. Gotchev, and A. Aksay, "Modelling of the stereoscopic HVS," MOBILE 3DTV, Eur. 7th Framework Programme Res. Technol. Develop., Project No. 216503, 2009. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D5.3_Mobile3DTV_v2.0.pdf
[36] B. J. Rogers, *Binocular Vision and Stereopsis*. New York, NY, USA: Oxford Univ. Press, 1995.
[37] H. R. De Silva and S. H. Bartley, "Summation and subtraction of brightness in binocular perception," *Brit. J. Psychol.*, vol. 20, no. 3, pp. 241–250, Jan. 1930.
[38] Z. Li and J. J. Atick, "Efficient stereo coding in the multiscale representation," *Netw., Comput. Neural Syst.*, vol. 5, no. 2, pp. 157–174, Mar. 1994.
[39] G. R. Engel, "The visual processes underlying binocular brightness summation," *Vis. Res.*, vol. 7, nos. 9–10, pp. 753–767, Sep. 1967.
[40] D. W. Curtis and S. J. Rule, "Binocular processing of brightness information: A vector-sum model," *J. Exp. Psychol. Hum.*, vol. 4, no. 1, pp. 132–143, Feb. 1978.
[41] A. I. Cogan, "Human binocular interaction: Towards a neural model," *Vis. Res.*, vol. 27, no. 12, pp. 2125–2139, Feb. 1987.
[42] 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.
[43] C.-B. Huang, J. Zhou, Y. Zhou, and Z.-L. Lu, "Contrast and phase combination in binocular vision," *PLoS ONE*, vol. 5, no. 12, Dec. 2010, Art. no. e15075.
[44] G. E. Legge and Y. C. Gu, "Stereopsis and contrast," *Vis. Res.*, vol. 9, no. 8, pp. 989–1004, Feb. 1989.
[45] L. L. Kontsevich and C. W. Tyler, "Analysis of stereothresholds for stimuli below 2.5 c/deg," *Vis. Res.*, vol. 34, no. 17, pp. 2317–2329, Sep. 1994.
[46] L. K. Cormack, S. B. Stevenson, and C. M. Schor, "Interocular correlation, luminance contrast and cyclopean processing," *Vis. Res.*, vol. 31, no. 12, pp. 2195–2207, Feb. 1991.
[47] L. K. Cormack, S. B. Stevenson, and D. D. Landers, "Interactions of spatial frequency and unequal monocular contrasts in stereopsis," *Perception*, vol. 26, no. 9, pp. 1121–1136, Feb. 1997.
[48] F. Hou, C.-B. Huang, J. Liang, Y. Zhou, and Z.-L. Lu, "Contrast gain-control in stereo depth and cyclopean contrast perception," *J. Vis.*, vol. 13, no. 8, pp. 1–19, Jul. 2013.
[49] Y. Liu, J. Yang, Q. Meng, Z. Lv, Z. Song, and Z. Gao, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Process.*, vol. 125, pp. 237–248, Aug. 2016.
[50] C. B. Huang, J. Zhou, Z. L. Lu, and Y. Zhou, "Deficient binocular combination reveals mechanisms of anisometropic amblyopia: Signal attenuation and interocular inhibition," *J. Vis.*, vol. 11, no. 6, p. 4, May 2011.
[51] P. B. Hibbard, "Binocular energy responses to natural images," *Vis. Res.*, vol. 48, no. 12, pp. 1427–1439, Jul. 2008.
[52] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: Energy models, position shifts and phase shifts," *Vis. Res.*, vol. 36, no. 12, pp. 1839–1857, Jun. 1996.
[53] 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.
[54] 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.
[55] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Dec. 2003, pp. 1398–1402.
[56] 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.
[57] 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.
[58] *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*, VQEG, 2008. [Online]. Available: http://www.vqeg.org/

YUN LIU received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, China, in 2016.

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.



FANHUI KONG received the M.S. degree in supply chain engineering from the Tianjin University of Technology, Tianjin, China, in 2016, where she is currently pursuing the Ph.D. degree with the School of Management. She is also a Visiting Scholar with Jacksonville University, Jacksonville, FL, USA. Her research interests include information management, structure information processing, traffic flow information management, supply chain strategy, circular economy, big data analysis, and the Internet-of-Things applications.

ZHIZHUO ZHEN is currently pursuing the B.S. degree with the School of Artificial Intelligent, Hebei University of Technology, Tianjin, China.

His research interests include artificial intelligence, structure information processing, machine learning, image processing, big data analysis, and the Internet of Things.