IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 25, 2018, accepted September 1, 2018, date of publication September 13, 2018, date of current version October 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2869525

Stereoscopic Image Quality Assessment by Considering Binocular Visual Mechanisms

GUANGMING SUN[1], YONG DING[1], RUIZHE DENG[1], YANG ZHAO[1], XIAODONG CHEN[2], AND ANDREY S. KRYLOV[3]

[1]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[2]The 14th Research Institute, China Electronics Technology Group Corporation, Nanjing 210039, China
[3]Laboratory of Mathematical Methods of Image Processing, Lomonosov Moscow State University, 119991 Moscow, Russia

Corresponding author: Yong Ding (dingy@vlsi.zju.edu.cn)

This work was supported in part by the National Science and Technology Major Project under Grant 2016ZX01012101-003 and in part by the Fundamental Research Funds for the Central Universities.

ABSTRACT With the booming of 3-D image processing in the entertainment industry and 3-D multimedia applications todays, the technology for assessing the quality of stereoscopic image faces more challenging tasks than its 2-D counterparts, such as binocular combination, stereo matching, and binocular rivalry. In this paper, a novel stereoscopic image quality assessment method is proposed by jointly exploring binocular fusion and rivalry models. In the quality-aware feature extraction stage, multi-scale binocular combination and binocular rivalry energy responses are first generated from reference and distorted stereopairs. In addition, considering that the changes of luminance affect the quality of stereoscopic image greatly, multi-scale visual features relating to image quality are obtained from its luminance maps as another binocular combination and rivalry features. Then, the dissimilarity of quality-aware features between the reference stereopair and its distorted version is quantified. Finally, such dissimilarities are mapped into an objective score to represent the perceptual quality of stereoscopic image through the pooling strategy of support vector regression. Experiments conducted on LIVE 3-D databases demonstrate that the proposed method achieves 96.61% and 96.03% in terms of Pearson's linear correlation coefficient on Database Phase I and Phase II, respectively, outperforming most of the state-of-the-art methods.

INDEX TERMS Stereoscopic image quality assessment, stereo vision, binocular fusion, binocular rivalry.

I. INTRODUCTION

Imaging is one of the main ways for human beings to obtain visual information todays. Due to compression, retargeting, coding or decoding operations during image processing, the quality of images degrades more or less through image transmission channels. So it's necessary to recognize and evaluate the quality degradation of images [1], [2]. Over the past few years, many classic methods have been created to deal with the problem about two-dimensional Image Quality Assessment (2D IQA), such as Structural SIMilarity (SSIM) [3], Visual Information Fidelity (VIF) [4] and so on. With the development of three-dimensional (3D) imaging technologies, the researches on evaluating degradation of stereoscopic images are needed to be explored urgently [5]–[8]. Yet, these existing methods proposed to evaluate the plant image quality are not involved in the field of stereoscopic image. Compared to its 2D counterparts, there have many limitations on how to explain the processing mechanism of binocular vision in exploring the quality degradation of stereopairs. Recently, although large amounts of experimental studies on Stereoscopic Image Quality Assessment (SIQA) have been performed, it's still a puzzle in fully understanding the neural mechanism of visual cortex about how the human brain perceives and deals with stereoscopic natural image [5].

Early methods [6] tried to extend some well-known 2D IQA methods to 3D tasks by directly employing these approaches (e.g., SSIM and VIF) to each of views in stereopair independently and combining the two obtained objective scores into a final quality score with different weights. However, these methods achieved poor performance due to ignoring the internal relationship between the left and right views. To improve the performance, binocular perceptual properties (e.g., binocular disparity and binocular combination) had been taken into account in some approaches later. You *et al.* [7] first pointed out that the degradation of quality in disparity maps also influences the overall objective quality

score, and they extended the previous experiments by analyzing the quality of disparity maps extremely and improved the performance to a certain extent.

Motived by the effects of binocular masking and binocular rivalry when viewing asymmetrical stereopairs, Chen *et al.* [8] took the local energy response of left and disparity-compensated right views as corresponding weighting coefficients to explain the binocular rivalry behaviors of human perception, and then integrated two views into a monocular image. Bensalma and Larabi [9] developed a Binocular Energy Quality Metric (BEQM) by estimating the associated binocular energy of simple and complex cells, and utilizing complex wavelet transform to obtain the final responses of stereoscopic images. In [10], the energy responses of the reference and distorted stereopairs, constructed by log-Gabor filter, were modulated by binocular response and binocular mask components. However, due to the lack of fully understanding on the binocular perceptual vision, these methods do not capture the intrinsic properties of binocular vision (i.e., how the cyclopean vision is created and recognized from two eyes in the brain).

It has been claimed that there still exist other binocular properties (e.g., binocular fusion or suppression) for stereo vision in human brain [11]–[13]. By jointly utilizing the effects of different binocular properties on stereo vision, Shao *et al.* [11] divided stereoscopic images into monocular regions, binocular fusion regions and binocular suppression regions. And then each region was evaluated respectively and the measured results were integrated into an overall score. This framework was further improved in another study of Shao *et al.* [12] by building monocular and binocular visual perception models. Considering monocular and binocular cells in primary visual processing mechanism, Cao *et al.* [13] classified image regions into monocular and binocular visual regions, extracted several visual characteristics from those regions to calculate monocular and binocular local quality, and generate the overall quality. Yet, it remains quite unclear how to classify the binocular fusion and suppression spatial regions by complex human visual cortex, and the visual processing cannot be simulated by these proposed methods correctly.

Hence, more and more researches focus on the core issue that how to generate the so-called 'cyclopean' image from two eyes by simulating the binocular behaviors of simple and complex cells in human brain. Lin *et al.* [14] focused on the simulations of binocular properties and applications of low-level features, and then created the cyclopean phase and amplitude maps to measure perceptual quality. In [15], monocular energy response, binocular energy response and binocular rivalry response were established from original and distorted stereopairs to simulate the Receptive Fields of simple and complex cells in the primary visual cortex (V1) area, respectively. Shao *et al.* [16] extracted energy responses and phase maps from stereopair by modeling visual properties of human visual cortex, and then transformed them into microstructure and macrostructure features by applying a new feature encoding and similarity measure approach.

As mentioned above [8]–[16], the facts are need to be pointed that there are two visual pathways carrying visual information through neural processing in human brain, in which the ventral pathway represents for the perception and recognition of natural scene. The retinal information, captured from the RFs of the ganglion cells, are firstly transferred to the Lateral Geniculate Nucleus (LGN), and then sent further to the ventral stream for visual information representations and integrations. The ventral stream begins from simple and complex cells in V1 area, and then goes through the V2, V3 and V4 areas [17], [18]. As the largest area of the visual cortex, the area V1, which mainly consists of two types of cells named simple cells and complex cells, is critical to the generation of early vision and responsible for the perception of human visual. The simple cells in V1 deal with the retinal information of left and right views receiving from corresponding LGN respectively, and the complex cells are used to connect the left and right retinal signals to binocular signals, responsible for binocular combination in stereo vision.

Inspired by the mechanism of stereo vision in human brains, an effective algorithm for assessing quality of stereoscopic images is proposed in the paper. By simulating the neural processing of simple and complex cells, features related to both binocular fusion and binocular rivalry are first extracted from reference and distorted stereopairs, respectively. Then after the stages of similarity measurement and features encoding, several similarities of binocular features between pristine stereopair and its distorted version are quantified. Finally, all the similarities are synthesized and mapped into an overall objective quality score by Support Vector Regression (SVR). Experimental results on the publicly available 3D databases demonstrate that the proposed algorithm achieves better performance in predicting the quality of stereoscopic distorted images than the existing state-of-the-art methods in both symmetric and asymmetric distorted stereopairs. The main contributions in the paper are summarized as follows:

1) Through the deep analysis of binocular visual mechanisms, several quality-aware features are captured to characterize stereoscopic image structure and information by jointly simulating binocular fusion and binocular rivalry properties in human brain.

2) We spread a novel binocular visual integration algorithm based on an existing stereoscopic image integration model (called DSKL model) by using gain-control and gain-enhancement models. The novel integration algorithm can explain the intrinsic properties of binocular fusion well.

The rest of this paper is organized as follows. In Section II, details of the proposed method jointly considering binocular fusion and rivalry properties of stereo vision are presented. Section III shows the experimental results and discussions. And conclusions are drawn in Section IV.

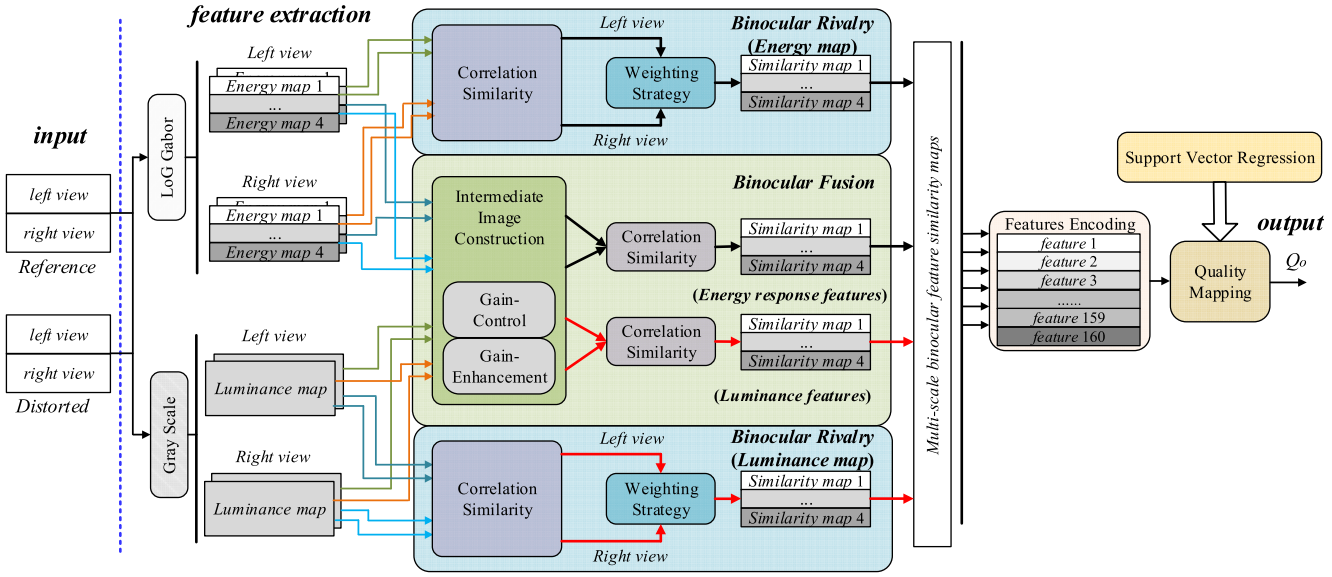


FIGURE 1. Framework of the proposed SIQA method.

II. PROPOSED METHOD

Based on the detailed analysis of neural processing and the review of some existing objective SIQA methods, a novel algorithm is proposed by jointly considering binocular fusion and rivalry properties of Human Visual System (HVS), in which the framework is illustrated as Figure 1. For given reference and distorted stereopairs, multi-scale energy responses are extracted respectively by log-Gabor filter since it can better model the receptive field profile of simple cells in V1. Moreover, luminance maps are generated from stereopairs considering that the changes of luminance reflect the quality degradation of stereo images. Further, various binocular visual features are extracted from multi-scale energy responses and luminance maps by applying binocular fusion and binocular rivalry models to simulate complex cells in V1 of binocular vision. Then, a set of feature vectors are obtained through the features encoding strategy using Local Binary Patterns (LBPs), which are relative to subjective visual system. Finally, these similarities of all binocular visual features between the reference and distorted stereopairs are quantified by dissimilarity measurement and mapped into an objective quality score by the SVR.

A. BINOCULAR FUSION MODEL

The retinal information received from the left and right RFs of the ganglion cells is combined into the single visual signals by complex cells of V1 in the visual nervous system, called binocular fusion. There have developed numerous biological models to simulate the binocular combination behaviors of complex cells, in which the main goal is to generate a single ''cyclopean'' perceptual image from visual signals captured by two eyes. Ding *et al.* [19] provide a contrast binocular combination method, in which a single combined image can be created from two eye's signals by jointly employing gain-control and gain-enhancement models, called the Ding, Sperling, Klein and Levi (DSKL) model.

One problem directly employing the DSKL model to the proposed algorithm is that the DSKL model takes two sine waves as input, but the inputs of our algorithm are the left and right images, which only contain contrast information and lack of phase components. Inspired by the DSKL model, we spread the model by employing the position shift mechanism instead of the phase shift mechanism of DSKL model to generate the disparity map [20], and then integrate the left and disparity-compensated right views to the cyclopean image using gain-control and gain-enhancement models of the DSKL model. To represent the position shift mechanism, a robust dense stereo reconstruction algorithm proposed in [21] is adopted to generate a binocular disparity map owing to its better performance in dealing with the issues of occlusion and depth discontinuities. In the features extraction of binocular fusion, log-Gabor responses on different spatial frequencies and orientations, and luminance maps of left and disparity-compensated right views are firstly extracted as the quality-aware features [18]. And then, applying the improved DSKL model to represent binocular fusion behavior of complex cells, multi-scale binocular fusion energy responses and multi-scale binocular fusion luminance maps are generated.

1) THE GENERATION OF MULTI-SCALE LOG-GABOR RESPONSES

Previous researches have shown that each eye's RFs properties in the V1 can be modeled by Gabor-like filters, in which the log-Gabor filter is selected in this paper due to its better performance in modeling the RFs properties than other Gabor-like filters [18]. The filter can be described as follows:

$$LG(f,\theta) = \exp\{-\frac{[\log(f/f_0)]^2}{2[\log(\sigma_f/f_0)]^2} - \frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\} \quad (1)$$

where f_0 and θ_0 are the central frequency and orientation angle, and 4 scales and 4 orientations (i.e., 0°, 45°, 90° and 135°) are selected in this paper. σ_θ and σ_f define the filter's angular bandwidth and scale bandwidth.

2) GAIN-CONTROL AND GAIN-ENHANCEMENT MODEL

The gain-control model is one of the most effective binocular combination models because it can explain the neural processing of binocular vision well, including Fencher's paradox and the cyclopean perception [22]. In addition, the phenomenon of gain-enhancement has been found in center-surround interactions that the gain enhancement from one eye to the other eye also receives the suppression from the other eye [23]. In the DSKL model, a cyclopean image can be generated by jointly considering the gain-control and gain-enhancement models as following:

$$E_v(x,y;f) = \frac{\sum_\theta |C_v(x,y;f,\theta)|}{g_c}$$
$$E_v^*(x,y;f) = \frac{\sum_\theta |C_v(x,y;f,\theta)|}{g_e} \quad (2)$$

where $C_v(x, y: f, \theta)$ denotes the energy response map of a single view in a specific scale and orientation, in which $v \in \{l, r\}$. $E_v(x, y: f)$ and $E_v^*(x, y: f)$ are the total energy response across different orientations in different scales for gain-control and gain-enhancement. g_c and g_e represent the gain-control threshold and the gain-enhancement threshold, respectively.

Considering the two gain effects of total contrast energy applied by the other eye, the monocular energy responses of left or right view can be re-expressed.

$$|\mathrm{C}_l(x,y;f)|_{ce} = \frac{1+\frac{E_r^*(x,y;f)}{1+\beta E_l(x,y;f)}}{1+\frac{E_r(x,y;f)}{1+\alpha E_l(x,y;f)}} \sum_\theta |\mathrm{C}_l(x,y;f,\theta)|$$
$$|\mathrm{C}_r(x,y;f)|_{ce} = \frac{1+\frac{E_l^*(x,y;f)}{1+\beta E_r(x,y;f)}}{1+\frac{E_l(x,y;f)}{1+\alpha E_r(x,y;f)}} \sum_\theta |\mathrm{C}_r(x,y;f,\theta)| \quad (3)$$

where α and β are different gain-control efficiencies to represent different effects in final monocular energy response $|C_v(x, y; f)|_{ce}$ of gain-control and gain-enhancement, in which $v \in \{l, r\}$.

3) BINOCULAR ENERGY RESPONSES

By combination of the left and disparity-compensated right views after interocular interaction, the binocular energy response in different scales $|C(x, y; f)|$ can be finally expressed by the following function (4), as shown at the top of the next page: where φ_r and φ_l denote the phase of energy response of the left and right views.

4) SIMILARITY MEASUREMENT

Through the similar binocular combination method, the binocular energy response of distorted stereopairs can be also created, denoted as $|C_D(x, y; f)|$. Due to the degree of corruption dependents on the image's distortion level, or quality degradation, dissimilarity quantification between the reference stereopair and its corresponding distorted version should be used to calculate the degree of corruption in distorted stereopair and assess the quality of stereoscopic distorted image as follows:

$$\mathrm{S}_{BF}(x,y;f) = \frac{2\,|\mathrm{C}_R(x,y;f)| \times |\mathrm{C}_D(x,y;f)| + T_1}{|\mathrm{C}_R(x,y;f)|^2 + |\mathrm{C}_D(x,y;f)|^2 + T_1} \quad (5)$$

where $S_{BF}(x, y: f)$ represents the similarity map of binocular energy response in different scales for binocular fusion property, and T_1 is a constant to avoid the denominator being equal to zero.

5) BINOCULAR LUMINANCE MAP

In addition, the changes of the luminance can also reflect the distortion level of stereopair [24]. Then luminance maps of left and right views can be extracted and transformed into binocular luminance response through simple and complex cells of area V1 by the similar contrast binocular combination way. Note that, different with energy response, the luminance map lacks of orientation angle information. Thus in the stage of binocular combination, the combination function can be simplified to

$$\begin{aligned}&|\mathrm{I}(x,y;f)| \\ &= \sqrt{I_l(x,y;f)_{\mathrm{ce}}^2 + I_r(x,y;f)_{ce}^2 + 2I_l(x,y;f)_{ce} \times I_r(x,y;f)}_{ce} \\ &= \mathrm{I}_l(x,y;f)_{\mathrm{ce}} + \mathrm{I}_r(x,y;f)_{ce} \end{aligned} \quad (6)$$

where likely to the binocular generation of energy responses, $I_v(x, y; f)_{ce}$ denotes the final luminance map of the view v for binocular fusion property by considering the effect of two gain models.

B. BINOCULAR RIVALRY MODEL

In addition to binocular fusion in V1 area of nervous system, another very important physiological phenomenon in human vision is binocular rivalry [17]. Binocular rivalry responses in V1 area play vital roles in simulating the effects of binocular interaction in human brain. Specifically, when different monocular stimuli signals of complex cells are transmitted to corresponding retinal locations of the two eyes, the cyclopean visual signals can be dominated by the visual signals from one eye with the higher energy response, instead of the average results of the two monocular stimuli signals. Based on this observation, binocular rivalry, as the result of the monocular stimuli dominating stereo perception, can be recognized as a plausible explanation. Thus, it is wise to consider the property of binocular rivalry into the paper to improve the performance of the proposed algorithm.

Through the simple cells of V1 area, the visual signals of left and right eyes are transformed into corresponding energy responses, respectively. In order to measure the quality degradation of corresponding distorted image in two views,

$$|\mathrm{C}(x,y;f)| = \left\| |\mathrm{C}_l(x,y;f)|_{ce} + |\mathrm{C}_r(x,y;f)|_{ce} \right\| = \sqrt{|\mathrm{C}_l(x,y;f)|^2_{ce} + |\mathrm{C}_r(x,y;f)|^2_{ce} + 2\,|\mathrm{C}_l(x,y;f)|_{ce}\,|\mathrm{C}_r(x,y;f)|_{ce}\cos(\phi_r - \phi_l)} \quad (4)$$

the similarity maps are generated by dissimilarity quantification between original and distorted images in each of views independently, as follows:

$$\mathrm{S_v}(x,y;f) = \frac{2\left|\mathrm{C}_{R,v}(x,y;f)\right| \times \left|\mathrm{C}_{D,v}(x,y;f)\right| + T_2}{\left|\mathrm{C}_{R,v}(x,y;f)\right|^2 + \left|\mathrm{C_{D,v}}(x,y;f)\right|^2 + T_2} \quad (7)$$

where $C_{R,v}(x, y: f)$ and $C_{D,v}(x, y: f)$ denote the energy response maps of reference and distortion stereopairs in different scales, in which $v \in \{l, r\}$. $S_v(x, y: f)$ is the similarity map between the reference stereopair and its distorted version in the left or right view. T_2 is a constant and not equal to zero.

According to the researches on binocular rivalry model, each of two eyes exerts gain control on the signals received by the other eye in corresponding spatial locations, and the ability of gain control to the other eye is in proportion to the energy response of its own visual signal. Usually, a linear model is used to simulate the binocular interaction mechanism of HVS, as follows:

$$\mathrm{S_{BR}}(x,y;f) = w_r(x,y;f) \times \mathrm{S}_l(x,y;f) + w_r(x,y;f) \times \mathrm{S}_r(x,y;f) \quad (8)$$

where $S_{BR}(x, y: f)$ represents the similarity maps of binocular energy response in different scales for binocular rivalry property, $w_l(x, y: f)$ and $w_r(x, y: f)$ are the binocular rivalry weighting coefficients to indicate the relative receiving contribution of both eyes, denoted as

$$w_l(x,y;f) = \frac{1 + \mathrm{M}_l(x,y;f)}{1 + \mathrm{M}_l(x,y;f) + \mathrm{M}_r(x,y;f)}$$
$$w_r(x,y;f) = \frac{1 + \mathrm{M}_r(x,y;f)}{1 + \mathrm{M}_l(x,y;f) + \mathrm{M}_r(x,y;f)} \quad (9)$$

where $M_l(x, y: f)$ and $M_r(x, y: f)$ are the energy responses of the left and right views in different scales, respectively.

C. FEATURES ENCODING

The various binocular feature maps (i.e., multi-scale binocular fusion and rivalry energy responses, multi-scale binocular fusion and rivalry luminance maps) have been obtained from phase A and B. However, there has a high dimensionality and easy to overfit if directly using those binocular feature maps as inputs of regression learning. Instead of directly mapping those binocular features into an objective score, dimension reduction strategy can be considered to solve the problems about high dimensionality and complicated calculation.

LBPs are recognized as an effective structural and textural information operator and usually used in dimension reduction [25]. By comparing eight local neighbor pixels with a central pixel and concatenating the results, it can form a non-directional binary local structural pattern. However, the statistical histogram of LBP maps can output 256-dimensional features, which also means the high dimensionality. As an improved version of standard LBPs, the rotation-invariant uniform LBPs [25], only having 10-dimensional outputs for eight neighboring pixels, are adopted in the paper.

After features encoding strategy by the rotation-invariant uniform LBPs, the 160-dimensional feature vectors are generated from 16 different binocular feature maps. In order to estimate whether the features extracted by the features encoding are quality-aware for stereo images, the joint normalized histograms of feature vectors in the different distortion levels of JPEG2000 compression (JP2K) and Gaussian Blur (GB) created from two distorted stereopairs on LIVE 3D IQA Database Phase I are drawn in Figure 2.

From Figure 2, we can find that: (i) the distorted types of stereopairs can be easily distinguished through the shapes of the joint normalized histograms in low distortion level, which indicates these feature vectors are content-independent and type-dependent. (ii) the joint normalized histograms of feature vectors in different distorted levels are also different, and thus these extracted binocular features can reflect the degree of quality degradation of distorted stereopairs, which is called quality-aware features. In other words, these joint normalized histograms can be recognized as stable binocular quality-aware features to estimate stereo image quality.

D. POOLING STRATEGY

As what mentioned before, 160-dimensional features can be generated from 16 binocular feature maps. However, it is still difficult to understand the quality of 3D images from 160-dimensional features, intuitively. For the purpose of mapping the feature vectors into a final objective quality score, a pooling strategy should be established. In the early days, the pooling strategy generally adopt a linear function approach, by assigning different weights to each of the features, and allowing them to be linearly superposed to map the final quality score. Recently, the technology of machine learning has been successfully applied in the regression analysis. In the paper, SVR is employed to construct a regression function to give out an objective score (Q_o) due to its effectiveness and moderate computational complexity [26].

$$Q_\mathrm{o} = SVR(\mathrm{Index}_1, \mathrm{Index}_2, \cdots, \mathrm{Index}_{160}) \quad (10)$$

where $Index_1$ to $Index_{160}$ represent 160-dimensional quality-aware features, and through the pooling strategy of SVR, Q_o can be finally generated.

SVR consists of the training and testing processes. In the training process, we randomly divide the dataset into five subsets with no overlap by cross-validation scheme. Four of them (i.e., 80%) are used for training and the rest one (i.e., 20%) is used for testing. Since the performance of the

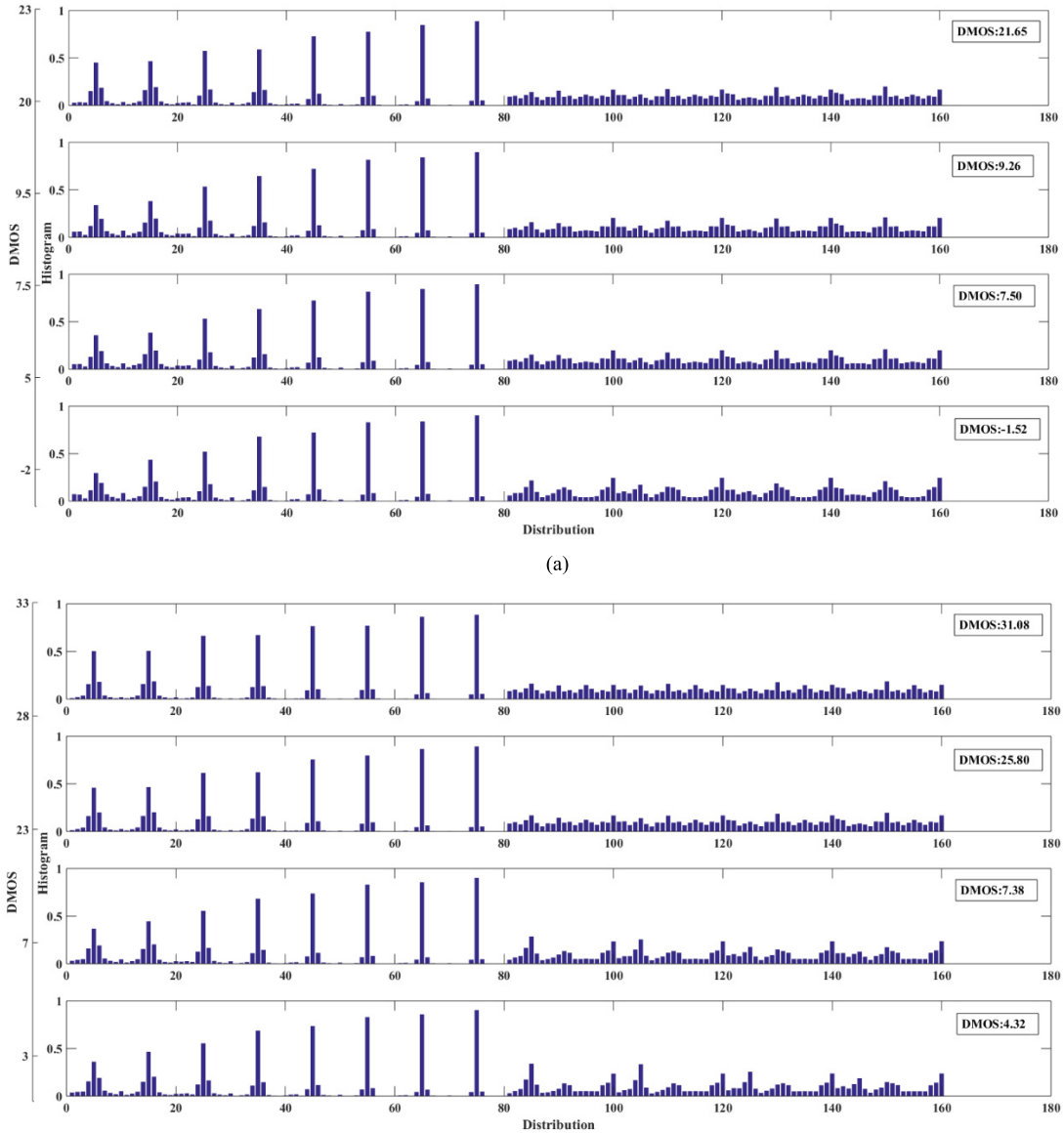


FIGURE 2. Joint normalized histograms of the quality-aware features in different distortion types and levels. (a) JP2K. (b) GB.

train-test procedure is dependent on the selection of training and testing content, we repeat it 1000 times and obtain the median results of 1000 iterations of cross-validation.

III. EXPERIMENTAL RESULTS

A. DATABASES

Experiments are conducted on two large-scale databases: the LIVE 3D IQA Database Phase I [8] and Phase II [27]. Phase I includes 365 symmetrically distorted stereopairs while Phase II has 120 symmetrically distorted stereopairs. In addition, phase II also contains 240 asymmetrically distorted stereopairs, where the degrees of stereoscopic image distortion in the left and right views are different. The distorted stereopairs in phase I and phase II are contaminated by five types of distortions (i.e., JP2K, JPEG compression, White Noise (WN), GB, and Fast-Fading (FF)). In order to further validate the performance of the proposed algorithm, the Waterloo IVC Database Phase II [28], [29] is also tested in the paper. The Waterloo IVC Database Phase II is created from 10 reference stereopairs, containing 130 symmetrically and 330 asymmetrically distorted stereopairs with three types of distortions including WN, GB and JPEG in four distorted levels. It's worthy emphasis that Waterloo IVC Database Phase II contains the Hybrid distortion in asymmetrical distortion, where the left and right views of asymmetrically distorted stereopairs have different distortion types.

To better illustrate the experimental results of the proposed algorithm, some classical Full-Reference (FR)

TABLE 1. Overall performance in LIVE 3D databases phase I and II.

	Phase I			Phase II		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
You'10	0.9303	0.9247	6.0161	0.7970	0.7840	6.7720
Chen'13	0.9167	0.9157	6.5503	0.9010	0.8930	4.9870
Cao'16	0.9380	0.9340	5.7330	0.9040	0.9020	4.8020
Shao'17	0.9389	0.9308	5.6459	0.9263	0.9282	4.1996
Ma'17	0.9469	0.9340	5.2111	0.9300	0.9218	4.1232
Xu'16	0.9381	0.9310	5.6789	0.8979	0.8879	4.9680
Lin'17	0.9242	0.9203	6.2629	0.9113	0.8935	4.6477
Geng'17	0.9430	0.9320	5.5140	0.9210	0.9190	5.4000
Jiang'18	0.9460	0.9378	5.3160	0.9261	0.9257	4.2679
Our	**0.9661**	**0.9567**	**4.1927**	**0.9603**	**0.9533**	**3.1210**

methods including You'10 [7] and Chen'13 [8], and several state-of-the-art methods such as Cao'16 [13], Shao'17 [16], Ma'17 [15], Xu'16 [30], Lin'17 [14], Geng'17 [31] and Jiang'18 [32] are selected. In order to evaluate the performance of SIQA methods quantitively, three metrics including Pearson's Linear Correlation Coefficient (PLCC), Spearman's Rank-order Correlation Coefficient (SRCC) and Root Mean Squared Error (RMSE) are adopted. PLCC and RMSE are used to evaluate prediction accuracy and consistency with human subjective evaluation, while SRCC is an indicator of prediction monotonicity. Larger PLCC and SRCC and lower RMSE indicate better performance.

B. OVERALL PERFORMANCE

The overall experimental performance comparison on LIVE databases between the proposed algorithm and other state-of-art methods is shown in Table 1, where the best results are highlighted in boldface.

From these results of the overall assessment performance, we can find that the overall performance of our method is far better than the compared methods both on LIVE Phase I and on Phase II, especially on Phase II. Owing to that Phase II contains both symmetrically and asymmetrically distorted images while Phase I only contains symmetrically distorted images, the performance of all the listed methods on Phase II is worse than that on Phase I more or less. In real applications, a stereoscopic pair may be distorted symmetrically or asymmetrically. Therefore, the comparison on Phase II is more constructive. The results indicate that the proposed algorithm can achieve better performance on the prediction of asymmetric distortion.

In addition, some important information is released from the table. First, the early methods, like You's scheme [7], achieve poor performance in all distortion types on both two LIVE 3D Databases because it is a simple extension from 2D image quality assessment methods, indicating that directly applying 2D methods into SIQA is unreasonable, especially in asymmetrically distorted stereopairs. Second, Chen's method [8] achieves better performance on 3D LIVE Databases Phase II than the early methods because the method considers binocular properties, like binocular combination and binocular rivalry. However, considering binocular visual system as a simplistic combination model, the performance of Chen's method cannot be greatly improved. With the researches of perceptual images on visual cortex system go deeper, the framework of 3D IQA constructed by Shao, Ma and other researchers become more and more reasonable, and the performance is getting better and better. Finally, the performance of those methods that consider binocular vision is more impressive than early methods, which indicates the facts that these quality-aware features extracted by binocular visual properties can reflect the objective quality of stereopair well.

C. ROBUSTNESS ACROSS DIFFERENT DISTORTION TYPES

In order to more comprehensively evaluate the prediction performance of the proposed algorithm, the experiments comparing with these state-of-art algorithms are also operated on the individual distortion types. Table 2 and Table 3 show the individual performance comparisons of different distortion types on LIVE Phase I and Phase II and the top metrics are highlighted in boldface. Due to part of the results are not provided in the original articles, "-" is used to indicate for unavailable results.

From Table 2 and 3, it's clear to find that the performance of the proposed algorithm achieves the top results on different distortion types ten times in terms of PLCC and SRCC, followed by Ma's method (six times), Jiang's method (two times), Xu's method and Geng's method (one time). In addition, the proposed algorithm also achieves perfect results on the individual distortion types in terms of RMSE. Although the performance on the types of asymmetrical JPEG compression and symmetrical Gaussian Blur distortion types is slightly worse than some of the best performance, there is only a narrow margin comparing to the leading methods. In addition, the proposed algorithm achieves better results than other FR algorithms for other distortion types across all databases. Overall, the performance of the proposed algorithm is more stable and satisfying across all different symmetrical and asymmetrical distortion types than other algorithms, and well closest to human subjective evaluation.

D. ROBUSTNESS ACROSS DIFFERENT DATABASES

For further validating the effectiveness and robustness of the proposed algorithm, experiments on the Waterloo IVC Database Phase II are conducted and the overall performance is illustrated in Table 4. Note that, since the Waterloo IVC Database Phase II is only created in recent years, there are not much research on the database, so we can only list the experimental results of several available FR and No-Reference (NR) SIQA algorithms (i.e., NR Chen's [33], SINQ [34], ADD-SSIM [35], FR Chen's [8] and Wang's [28]) on the database for comparison. In addition, the performance in symmetrical and asymmetrical distortion of the proposed algorithm are also listed in Table 5. As we can see from Table 4, the performance of the proposed algorithm still outperforms that of all other compared methods on the Waterloo IVC Database Phase II. These compared results again confirm the efficacy

TABLE 2. The performance comparison of individual distortion types on phase I.

Distortion	Criteria	You'10	Chen'13	Cao'16	Shao'17	Ma'17	Xu'16	Lin'17	Geng'17	Jiang'18	Our
JP2K	PLCC	0.9050	0.9164	-	0.9360	**0.9610**	0.9373	0.9520	0.9420	0.9408	0.9559
	SRCC	0.9051	0.8956	0.9010	0.9000	**0.9143**	0.8920	0.9127	0.9050	0.9027	0.9029
	RMSE	6.2066	4.7842	-	4.5571	-	-	-	-	-	**3.6770**
JPEG	PLCC	0.6333	0.6344	-	0.6654	0.7746	0.6547	0.7546	0.7190	0.6975	**0.8257**
	SRCC	0.6008	0.5582	0.5970	0.6339	0.6659	0.6368	0.7164	0.6530	0.6628	**0.7676**
	RMSE	5.7097	5.0681	-	6.4344	-	-	-	-	-	**3.5509**
WN	PLCC	0.9351	0.9436	-	0.9441	0.9412	0.9373	0.9266	0.9630	0.9516	**0.9684**
	SRCC	0.9403	0.9481	0.9350	0.9430	0.9077	0.9393	0.9289	**0.9560**	0.9529	0.9412
	RMSE	5.6216	4.8168	-	4.3308	-	-	-	-	-	**4.0003**
GB	PLCC	0.9545	0.9417	-	0.9542	**0.9711**	0.9633	0.9583	0.9620	0.9578	0.9494
	SRCC	0.9300	0.9261	0.9380	0.9247	0.9030	**0.9445**	0.9332	0.9310	0.9361	0.8833
	RMSE	5.6798	4.8063	-	5.4836	-	-	-	-	-	**4.2515**
FF	PLCC	0.8589	0.7580		0.8304	0.8941	0.8756	0.8620	0.8670	0.8554	**0.8874**
	SRCC	0.8030	0.6879	0.7740	0.7807	0.8312	0.8324	0.8286	0.8160	0.8079	**0.8353**
	RMSE	8.4923	8.0323	-	6.9238	-	-	-	-	-	**5.4412**

TABLE 3. The performance comparison of individual distortion types on phase II.

Distortion	Criteria	You'10	Chen'13	Cao'16	Shao'17	Ma'17	Xu'16	Geng'17	Jiang'18	Our
JP2K	PLCC	0.7320	0.8426	-	0.8768	**0.9670**	0.9218	0.8610	0.8463	0.9633
	SRCC	0.7309	0.8334	0.8580	0.8747	0.9328	0.9027	0.8360	0.8497	**0.9357**
	RMSE	4.1860	5.5620	-	4.7574	-	-	-	-	**2.5375**
JPEG	PLCC	0.6741	0.8422	-	0.8506	**0.9350**	0.9038	0.8350	0.8771	0.8881
	SRCC	0.8334	0.8396	0.8690	0.8340	**0.8968**	0.8893	0.8410	0.8547	0.8374
	RMSE	4.0860	3.8650	-	3.8537	-	-	-	-	**3.1960**
WN	PLCC	0.5464	0.9602	-	0.9339	0.9341	0.9045	0.9490	0.9553	**0.9715**
	SRCC	0.4820	0.9554	0.9290	0.9325	0.8893	0.8861	0.9390	**0.9563**	0.9464
	RMSE	4.3960	3.3680	-	4.5729	-	-	-	-	**2.4513**
GB	PLCC	0.9763	0.9650	-	0.9445	0.9384	0.8974	0.9790	0.9845	**0.9910**
	SRCC	0.9227	0.9096	0.9110	0.9241	0.8992	0.8914	0.9200	0.9383	**0.9536**
	RMSE	8.6490	3.7470	-	3.8315	-	-	-	-	**1.7858**
FF	PLCC	0.8561	0.9097	-	0.9330	0.9489	0.9183	0.9480	0.9601	**0.9618**
	SRCC	0.8392	0.8890	0.9220	0.9409	0.9167	0.8996	0.9160	**0.9555**	0.9385
	RMSE	4.6490	4.9660	-	3.8973	-	-	-	-	**2.8901**

TABLE 4. Overall performance in Waterloo IVC Database Phase II.

	Type	PLCC	SRCC
Chen	NR	0.882	0.884
SINQ	NR	0.922	0.911
ADD-SSIM	FR	0.501	0.378
Chen	FR	0.569	0.444
Wang	FR	0.891	0.868
Our	FR	**0.938**	**0.934**

TABLE 5. The performance of the proposed algorithm in Waterloo IVC Database Phase II.

	Symmetric	Asymmetric	ALL
PLCC	0.9660	0.9153	0.9378
SRCC	0.9574	0.9067	0.9344
RMSE	4.2348	7.7935	6.7391

of the proposed 3D IQA algorithm. From Table 5, it's easy to see that the proposed algorithm achieves impressive performance in prediction of symmetrically distorted images, while for asymmetric distortion, the performance is particularly poor. That's because the Hybrid distortion in Waterloo IVC Database Phase II can't be predicted well by the proposed algorithm.

E. INTUITIVE REPRESENTATION OF ALGORITHM PERFORMANCE

Furthermore, in order to intuitively demonstrate the performance of the proposed algorithm, the scatter plots of the overall predicted objective quality scores against subjective quality scores on LIVE Phase I and Phase II are drawn in Figure 3. In the scatter plots, the horizontal axis means predicted objective quality scores after objective score mapping and the vertical axis shows subjective image quality scores of distorted stereopairs on LIVE 3D Database Phase I and Phase II. The distribution of scatter plot represents the prediction performance of the proposed metric, in which we can get the conclusion that the performance of the proposed metric achieves high consistency with the subjective quality scores and the performance on asymmetrical distorted stereopairs is not worse than that on symmetrical distorted stereopairs.

We also draw the scatter plots of DMOS versus the predicted quality scores of different distortion types on LIVE Phase I and Phase II, and different distortion types are represented by scatters with different colors, shown in Figure 4.

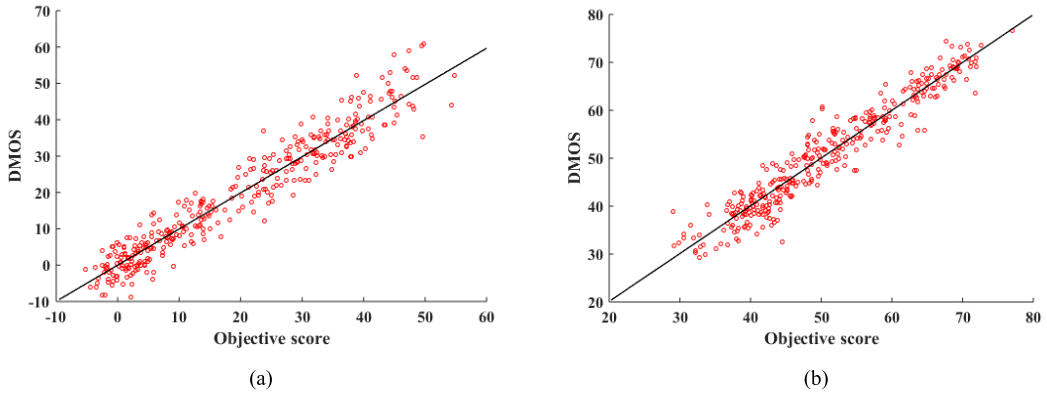


FIGURE 3. Scatter plots of the overall predicted quality scores against the subjective scores (DMOS) of the proposed metric on LIVE 3D IQA Database Phase I and II. (a) LIVE 3D Phase I. (b) LIVE 3D Phase II.

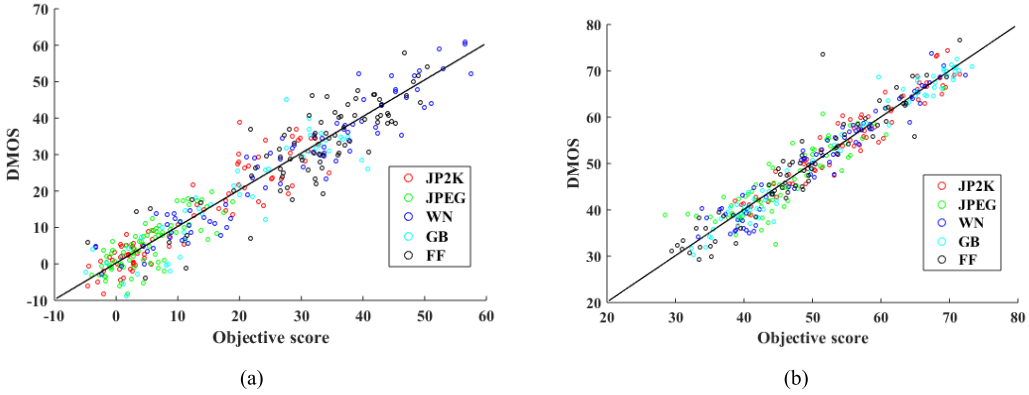


FIGURE 4. Scatter plots of predicted quality scores against the subjective scores (DMOS) of the proposed metric on five types of distortion on LIVE 3D IQA Database Phase I and II. (a) LIVE 3D Phase I. (b) LIVE 3D Phase II, respectively.

According to the scatter plots, it is clear that predicted objective quality scores of the proposed algorithm are well correlative with DMOSs in most distortion types (i.e., JP2K, WN and FF), as the scatter plots of JP2K, WN and FF distortion types are more distributed closely beside the fitting curve than that of other distortion types. The same conclusions can also be seen from the indices of PLCC, SRCC and RMSE.

Besides, in order to obtain content-independent performance results, 5000 results of performance metrics have been conducted across 1000 iterations of cross-validation in the quality mapping stage. For a good SIQA algorithm, 5000 performance results should be consistent and similar, called ''stabilization''. To demonstrate the stabilization of the proposed algorithm, the box-plots of PLCC metric on three different databases are drawn in Figure 5. From Figure 5, it is easy to find that the proposed algorithm is concentrated and only has small whiskers on the three different databases, which means the performance of our algorithm is stable and impressive. In addition, we can also find that the stabilization of performance on LIVE 3D Database Phase I is significantly better than the other two databases, in which that on Waterloo IVC Database Phase II is the worst. The results concluded from box-plots are consistent with the previous obtained conclusions.

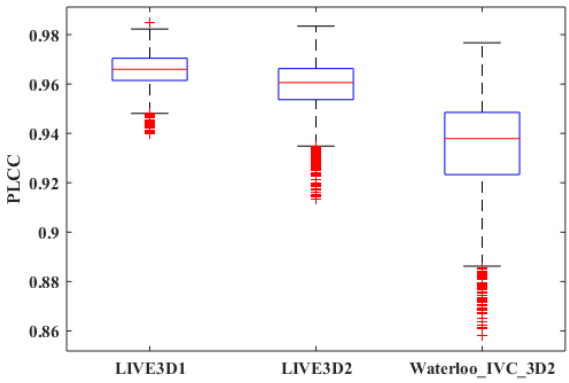


FIGURE 5. Box-plots of PLCC metric of the proposed algorithm across 1000 iterations of cross-validation on LIVE 3D IQA Database Phase I and II, and Waterloo IVC Database Phase II, respectively.

To sum up, the proposed algorithm outperforms most existing state-of-the-art algorithms in terms of the high consistency with subjective evaluation for most of all distortion types in both databases, as well as the robustness across symmetrical or asymmetrical distortions as a result of taking the properties of binocular fusion and binocular rivalry into account.

IV. CONCLUSION

A novel stereoscopic image quality assessment method considering primary visual cortex mechanism including binocular fusion and binocular rivalry is presented in the paper. With two binocular visual pathways for exploring the quality of stereopairs, binocular fusion and binocular rivalry features are captured and then the similarities between the distorted stereopair and its corresponding undistorted version are quantified. After features encoding by LBPs, several quality-aware features are generated and finally pooled into an objective quality score. Experiments prove that the proposed method is more accurate and robust in predicting stereoscopic image quality than most existing methods.

REFERENCES

[1] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "PMFS: A perceptual modulated feature similarity metric for stereoscopic image quality assessment," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 1003–1006, Aug. 2014.

[2] 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.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[4] 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.

[5] 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.

[6] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.

[7] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, Scottsdale, AZ, USA, Jan. 2010, pp. 61–66.

[8] 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.

[9] 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.

[10] F. Shao, G.-Y. Jiang, M. Yu, F. Li, Z. Peng, and R. Fu, "Binocular energy response based quality assessment of stereoscopic images," *Digit. Signal Process.*, vol. 29, no. 1, pp. 45–53, Mar. 2014.

[11] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[12] F. Shao, W. Lin, G. Jiang, and Q. Dai, "Models of monocular and binocular visual perception in quality assessment of stereoscopic images," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 123–135, Jun. 2016.

[13] Y. Cao, W. Hong, and L. Yu, "Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2016, pp. 1–6.

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

[15] 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, Oct. 2017.

[16] 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, Jul. 2017.

[17] N. Kruger *et al.*, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1847–1871, Aug. 2013.

[18] W. E. L. Grimson, "From images to surfaces: A computational study of the human early visual system," *Neuropsychologia*, vol. 20, no. 4, p. 512, 1982.

[19] J. Ding, S. A. Klein, and D. M. Levi, "Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model," *J. Vis.*, vol. 13, no. 2, pp. 1–37, Feb. 2013.

[20] K. Li, F. Shao, G. Jiang, and M. Yu, "Full-reference quality assessment of stereoscopic images by learning sparse monocular and binocular features," *Proc. SPIE*, pp. 927312-1–927312-10, Nov. 2014, doi: 10.1117/12.2073641.

[21] S. Lee, J. H. Lee, J. Lim, and I. H. Suh, "Robust stereo matching using adaptive random walk with restart algorithm," *Image Vis. Comput.*, vol. 37, pp. 1–11, May 2015.

[22] 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.

[23] J. Ding, S. Klein, and D. Levi, "Binocular combination in abnormal binocular vision," *J. Vis.*, vol. 13, no. 2, pp. 1–31, Feb. 2013.

[24] 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.

[25] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[26] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[27] 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, Dec. 2013.

[28] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

[29] J. Wang, K. Zeng, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic images from single views," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Chengdu, China, Jul. 2014, pp. 14–18.

[30] H. Xu, M. Yu, T. Luo, Y. Zhang, and G. Jiang, "Parts-based stereoscopic image assessment by learning binocular manifold color visual properties," *J. Electron. Imag.*, vol. 25, no. 6, p. 061611, Dec. 2016.

[31] 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, 2017.

[32] G. Jiang, M. He, M. Yu, F. Shao, and Z. Peng, "Perceptual stereoscopic image quality assessment method with tensor decomposition and manifold learning," *IET Image Process.*, vol. 12, no. 5, pp. 810–818, Feb. 2018.

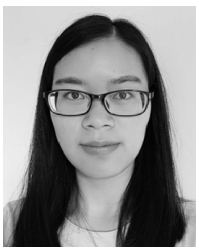
[33] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[34] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Process., Image Commun.*, vol. 58, pp. 287–299, Aug. 2017.

[35] 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.

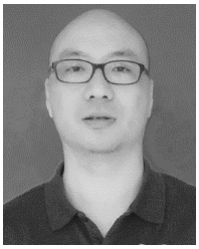
GUANGMING SUN received the B.S. degree from the Huazhong University of Science and Technology, China, in 2017. He is currently pursuing the M.S. degree with the College of Information Science and Electronic Engineering, Zhejiang University, China. His research interests include stereoscopic image quality assessment and computer vision.

YONG DING received the B.S. and M.S. degrees from the School of Electronic Science and Applied Physics, Hefei University of Technology, Hefei, China, in 1997 and 2000, respectively, and the Ph.D. degree from the College of Electronic Science and Engineering, Southeast University, Nanjing, China, in 2008. From 2000 to 2006, he was a Senior Engineer with the Research and Development Center, Hisense. From 2006 to 2008, he was a Senior Project Leader of the Architecture Design Department, OmniVision. In 2011, he joined the Laboratory of Mathematical Methods of Image Processing, Lomonosov Moscow State University, Russia, as a Visiting Scholar. Since 2009, he has been an Associate Professor with the Institute of VLSI Design, Zhejiang University. He holds over 23 Chinese patents. His research interests concentrate on image objective quality assessment, digital video processing, and associated system-on-chip architectures. He has authored over 50 papers in journals in these fields of research. He has made several plenary or invited talks on international conferences. He is currently in charge of several projects supported by the Chinese Government, including the National Science and Technology Support Program and the National High Technology Program (863 Program). These projects focus on image quality assessment and video processing.



RUIZHE DENG received the B.S. degree in electrical engineering and automation from Qinghai University, China, in 2015. She is currently pursuing the M.S. degree with the College of Information Science and Electronic Engineering, Zhejiang University, China. Her research interests include image quality assessment and computer vision.

YANG ZHAO received the B.S. degree from Shandong University, China, in 2015, and the M.S. degree from Zhejiang University, China, in 2018. His research interests include planar and stereoscopic image quality assessment, medical image processing, and computer-aided diagnosis.



XIAODONG CHEN received the B.S. degree in engineering from the School of Engineering, Ningbo University, Ningbo, China, in 2003, and the Ph.D. degree in mechatronic engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008.

ANDREY S. KRYLOV received the B.S. degree and the Ph.D. degree in applied mathematics from Lomonosov Moscow State University, Russia, in 1978 and 1983, respectively. Since 1984, he has been with the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, where he is currently a Full Professor and the Head of the Laboratory of Mathematical Methods of Image Processing. His research interests include mathematical methods of image processing, computer vision, ill-posed and inverse problems, and mathematical methods in physical chemistry of liquid metals.