IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 16, 2017, accepted July 24, 2017, date of publication July 28, 2017, date of current version August 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2733161

Modeling the Perceptual Quality of Stereoscopic Images in the Primary Visual Cortex

FENG SHAO[1], (Member, IEEE), WANTING CHEN[1], GANGYI JIANG[1], (Member, IEEE), AND YO-SUNG HO[2], (Fellow, IEEE)

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[2]School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea

Corresponding author: Feng Shao (shaofeng@nbu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61622109 and Grant 61271021, and in part by K. C. Wong Magna Fund in Ningbo University.

ABSTRACT Quality assessment of 3-D images encounters more challenges in better understanding of human visual system. In this paper, we propose a perceptual quality assessment approach for stereoscopic images by modeling visual properties of the primary visual cortex. For this purpose, we obtain a new feature encoding approach for the visual information, and define a new similarity measure approach to match the feature encoding to give more reliable and accurate quality assessment. Experimental results on three symmetrically, asymmetrically, and multiple distorted stereoscopic image quality assessment databases demonstrate that our method has high consistency with subjective assessment.

INDEX TERMS Stereoscopic image quality assessment, primary visual cortex, receptive field, microstructure, macrostructure.

I. INTRODUCTION

Qualty assessment of stereoscopic images is challenging but has made great progress in recent years. However, due to the limitations in fully understanding the neural mechanism that the human perceives an image in stereoscopic vision (to distinguish with the previous works that utilizes neural network techniques to predict the perceptual quality), computational model that mimics the visual neurons and predicts image quality accurately, nevertheless, may offer new insights to explain visual perception to a wide range of visual stimuli.

A. RELATED WORK

The essence of image quality assessment is to understand how we see and perceive the images. The classic full-reference (FR) methods, including structural similarity index (SSIM) [1], feature similarity (FSIM) [2], visual information fidelity (VIF) [3], gradient magnitude similarity deviation (GMSD) [4], and so on, are devoted to dig image features and measure their similarity. However, due to the difficulty in establishing clear principles in human vision perception, especially in the promising stereoscopic vision, it is important to consider how the existing computational models can be used to simulate the neural mechanisms of visual cortex. Computational models of the visual cortex could provide a platform for achieving such a target, integrating results across different cells to provide an overall explanation for the quality prediction.

The human perceives images in the ventral pathway, beginning with simple and complex cells in the primary visual cortex (V1). Then, the streams pass to the V2, V3 and V4 area for perception and recognition. The V1 is the largest part of the human visions system (HVS) responsible for the primary vision, and it receives the signals from the Lateral Geniculate Nucleus (LGN) located in both hemispheres of the brain. From the perspective of physiological point, the V1 has two main types of cells namely simple cells and complex cells responsible for perception [5]. Following neural processing in the receptive fields (RFs), two retinal images are passed to the V1 in the brain. The retinal information is first received by the simple cells which work independently to the left and right eyes. Then, each pair of RFs is connected to a complex cell to generate the binocular signal. As an alternative way to connect the left and right signals to gain more reliable perception, different types of binocular combination behaviors (also known as binocular integration behaviors in some literatures) are simulated to integrate the quality scores from left and right channels [6]–[10]. However, this type of approaches is still limited to relatively simple models of human perception, and does not capture the intrinsic behavior of complex cells, so that these approaches are unsuitable for stereoscopic image quality assessment (SIQA).

On the other hand, the challenge for SIQA mainly originates from the understanding of binocular vision. Aiming at utilizing the properties of binocular vision, our previous work [11] first interpreted the phenomena of monocular, binocular fusion and binocular suppression based on different region classification. Lee *et al*. [12] further discriminated coarse and fine 3D perception. For coarse 3D perception, regional 3D information is classified into one of stereopsis, binocular rivalry, or suppression. For fine 3D perception, binocular foveation and fusion are reflected to each segment. Cao *et al*. [13] classified image regions into monocular and binocular regions, and utilized several visual characteristics to simulate the primary visual processing mechanism. Even a complex visual process involves both binocular fusion and binocular rivalry that co-exist at different spatial regions, it remains unclear how the binocular vision affect the region classification because the findings have never tested by psychology experiments. Hence, a large number of works attempt to model the binocular fusion and binocular rivalry behavior on the so-called 'cyclopean' view, a binocular single vision created from two eyes. In [14], cyclopean image was created from the left image and disparity-compensated right image, and the local energy was used for weights to explain the binocular rivalry. In [15], the local phase and amplitude features extracted from the cyclopean image were used to measure 2D artifacts. Similarly, different feature representation methods were conducted on the cyclopean images in many works [16]–[19]. However, these methods do not concern the core issue that how the cyclopean vision is formed in the brain, and are distinguished only by different information representation for the cyclopean image.

As known, studies of visual cortex typically measure responses to a set of stimuli designed to investigate a psychology phenomenon [20]. Thus, we can measure the electroencephalograph (EEG)/functional magnetic resonance imaging (fMRI) responses in early visual cortex and develop different models to predict and measure these responses. To model the responses instead of measuring by devices, a set of V1-like filters are applied to the image to get the response signals. For binocular vision, binocular energy model is established from the left and right filters (defined as RFs in the two eyes), which depends on the disparity of the input stimulus, and the disparity tuning of the unit [21]. Different binocular energy models have been proposed to respond for stimulus from phase or position shift [22]–[24]. For SIQA task, Bensalma *et al*. [25] utilized Complex Wavelet Transform (CWT) to calculate the responses of two pairs of stereoscopic images, and built Binocular Energy Quality Metric (BEQM) to measure the quality. Galkandage *et al*. [26] extended the binocular energy model, and built Extended Binocular Energy Quality Metric (EBEQM) for stereoscopic vision. Our previous work [27] utilized binocular response and binocular mask components to modulate the perceptual quality. Overall, the energy responses are more in line with the process that human is responsible for simple and complex cells. However, the success of binocular energy models depends on accurate connections among the simple and complex cells, and need to dig much deeper features from the responses.

B. OBSERVATIONS

As mentioned above, previous works point out important visual properties of binocular combination behavior, e.g., cyclopean vision and binocular energy models in modeling the perceptual quality of stereoscopic images. However, these models or methods only explain limited phenomena of the V1. Most methods distinguished from others are only in different types and/or amounts of information extracted from the stereoscopic images. For SIQA task, even most of these models currently work well for symmetrically distorted stimuli, the core principle of a model should have the capability to handle asymmetrically distorted stimuli or multiple distorted stimuli. Therefore, from the perspective of visual pathway, the expansibility of these methods may be limited, because:

- Observing the impact of disparity activity, we note that the binocular energy model cannot well address the quality variations, and such variations require disparity stimulus, leading to poor performance in addressing asymmetrically distorted stimuli.
- The cyclopean vision is suggested a constraint for modeling binocular combination behavior, but the influence of different disparity stimuli is not well characterized.
- We observe that the V1 is responsible for the basic neuronal activity in early visual cortex. Increasing or decreasing the contrast induced by image distortion will affect the V1 activity. This motive us to simulate the visual pathway in the V1 for neural processing, rather than simply modeling one particular behavior.

C. CONTRIBUTIONS

In this work, we use hierarchical monocular simple cells, monocular complex cells, binocular simple cells and binocular complex cells to represent the visual pathway in the V1. The main motivation of this work is to develop a technique that more effectively models neural processing for the visual information representation in the V1. Towards this end, we obtain a new feature encoding approach, and define a new similarity measure approach to match the feature encoding to give more reliable and accurate quality assessment. For feature encoding, we construct responses across different scales, orientations and disparity planes to represent contrast and phase features. For similarity measure, we define microstructure similarity from the features to reflect the local perception, and macrostructure similarity to reflect the basic perception. The main contributions are summarized as follows:

1) We simulate the visual pathway in the V1 based on the RFs from monocular simple, monocular complex, binocular simple and binocular complex cells. Defining such a pathway can comprehensively measure quality degradations and variations.

2) We measure microstructure similarity from contrast and phase features to detect the local perception, and measure macrostructure similarity to reflect the basic

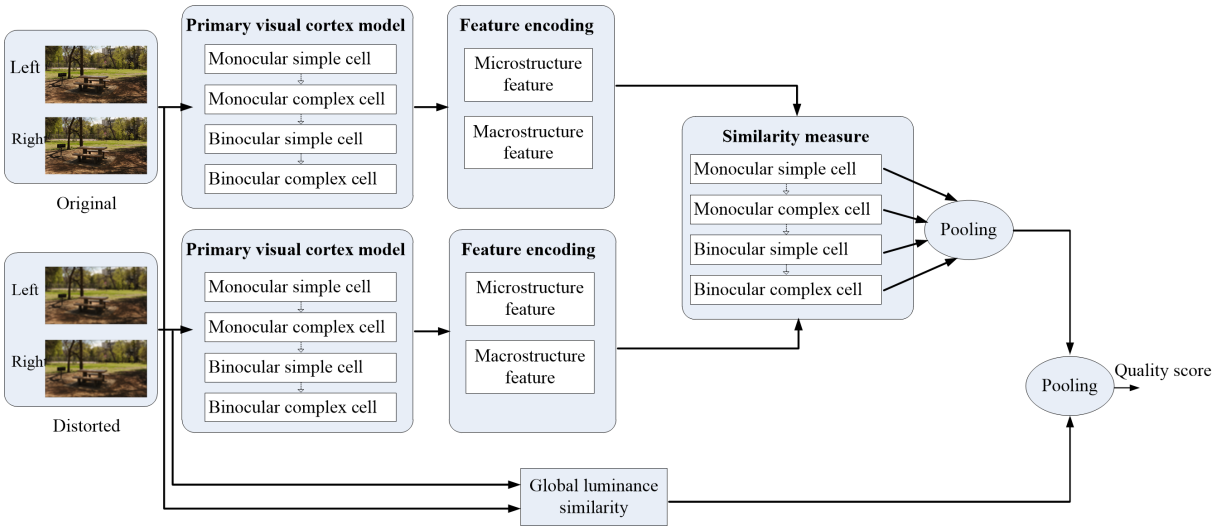


FIGURE 1. Procedure of the proposed method.

perception. We thus use different combination and pooling methods to combine the similarity indexes into a final quality.

3) Through comprehensive experiments and comparison, we analyze that our metric correlates well with subjective observations especially for asymmetrically and multiple distorted stereoscopic images, demonstrating that it can be used as a general quality evaluator for various stereoscopic images applications.

The remainder of the paper is organized as follows. Section 2 presents the proposed quality assessment approach. The experimental results are given and discussed in Section 3, and finally conclusions are drawn in Section 4.

II. THE PROPOSED APPROACH

Numerous studies have conducted visual quality analysis in the V1 [8]–[19], [28]. In this paper, we propose a perceptual quality assessment approach for stereoscopic images by modeling visual properties of V1. Our approach consists of four parts: visual information modeling, feature encoding, similarity measure and quality pooling, as shown in Fig. 1. The core of our approach is the modeling of monocular and binocular visual information to simulate the simple and complex cells. The visual information from each stage is exploited to extract contrast and phase features. Next, we measure the micro- and macro-structure similarities between the reference and distorted features, and finally employ different combination and pooling methods to combine the similarity indexes into a final quality.

A. V1 MODEL

Before processing by the V1 model, the input $I^{L/R}(\mathbf{x})$ is perceived by the lateral geniculate nucleus (LGN) neurons with a RF function $\mathbf{h}_\sigma(\mathbf{x})$. The response to the input is computed by convoluting with the circular symmetric on-center/off-surround RFs. Refer to [29], the output from this filter is subject to a multiplicative gain followed by a saturating nonlinearity, such that

$$X^{L/R}(\mathbf{x}) = k_{LGN} \times \left(\mathbf{h}_\sigma(\mathbf{x}) \otimes I^{L/R}(\mathbf{x})\right) \quad (1)$$

$$\mathbf{h}_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2}\frac{x^2+y^2-2\sigma^2}{\sigma^4}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2)$$

where L/R designates that the cell belongs to the left and right, "$\otimes$" is the convolution operation, σ is used to control the scales of LoG filter. In this paper, we set $\sigma = 2$ and $k_{LGN} = 2\pi$.

Based on the inputs coming from the LGN, the perceived V1 cells are classically divided into two types: simple cells and complex cells (also defined as neuron in some literatures). For stereoscopic vision, to characterize monocular and binocular V1-like RFs, these V1 cells can be divided into the following typical types: monocular simple and complex cells, binocular simple and complex cells. Similar to [30], we use Gabor-like filter to simulate the RF properties, defined as follows:

$$\mathbf{g}_{\theta,\omega}(\mathbf{x}) = \frac{1}{2\pi\gamma\sigma^2}\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\exp\left(j\omega x'\right) \quad (3)$$

where $x' = (x\cos\theta + y\sin\theta)$, $y' = (-x\sin\theta + y\cos\theta)$, $[x]^+ = \max(x, 0)$, ω represents the spatial frequency of the sinusoidal carrier, θ is the orientation (the angle of the normal to the sinusoid), γ is the spatial aspect ratio and σ is the sigma of the Gaussian envelope. Four orientations, 0°, 45°, 90° and 135°, are used with six different spatial frequencies, 1.74, 2.47, 3.49, 4.93, 6.98 and 9.87 (cycles/degree) (the scale is reflected by the spatial frequencies). Thus, the simple cell is computed with the response $S^{L/R,+}_{\theta,\omega}(\mathbf{x})$ of the Gabor filter to the input $X^{L/R}(\mathbf{x})$, responded to dark-light contrast polarity, which is defined as follows:

$$S^{L/R,+}_{\theta,\omega}(\mathbf{x}) = \mathbf{g}_{\theta,\omega}(\mathbf{x}) \otimes \left[X^{L/R}(\mathbf{x})\right]^+ \quad (4)$$

Next, the responses of simple cells are further processed by complex cells. To address directional selectivity of the RFs,

TABLE 1. The relationship between disparity *d* and allelotropic shift *s*.

Disparity (d)	V. near disparity	Near disparity	Zero disparity	Far disparity	V. Far disparity
Allelotropic shift (s)	-30	-15	0	15	30

all simple cells are decomposed only on the horizontal or vertical direction, the cells responded to horizontal or vertical boundary are obtained by

$$S_{\omega}^{H,L/R,+}(\mathbf{x}) = \sum_{\theta \in H} S_{\theta,\omega}^{L/R,+}(\mathbf{x}) \tag{5}$$

$$S_{\omega}^{V,L/R,+}(\mathbf{x}) = \sum_{\theta \in V} S_{\theta,\omega}^{L/R,+}(\mathbf{x}) \tag{6}$$

where H/V designates to the horizontal or vertical boundary. The cells in V1 layer 4 with light-dark contrast polarity are the inverse of those in (5) and (6):

$$S_{\omega}^{H/V,L/R,-}(\mathbf{x}) = -S_{\omega}^{H/V,L/R,+}(\mathbf{x}) \tag{7}$$

The cells in V1 layer 3B monocular are obtained by

$$B_{\omega}^{H/V,L/R,+/-}(\mathbf{x}) = 2\left[S_{\omega}^{H/V,L/R,+/-}(\mathbf{x})\right]^{+} \tag{8}$$

where the multiplicative factor of 2 is to compensate for the monocular simple cells received from only one eye. The response of a V1 monocular simple cell is calculated by

$$C_{\omega}^{H/V,L/R}(\mathbf{x}) = \left[B_{\omega}^{H/V,L/R,+}(\mathbf{x})\right]^{+} + \left[B_{\omega}^{H/V,L/R,-}(\mathbf{x})\right]^{+} \tag{9}$$

By integrating the simple cells within the horizontal and vertical directions, the response of a monocular complex cell is calculated by

$$C_{\omega}^{L/R}(\mathbf{x}) = \left(C_{\omega}^{H,L/R}(\mathbf{x}) + C_{\omega}^{V,L/R}(\mathbf{x})\right)^{2} \tag{10}$$

The response of a corresponding layer of binocular simple cells can be modeled as:

$$R_{\omega}^{B}(\mathbf{x}) = \left(C_{\omega}^{L}(\mathbf{x}) + C_{\omega}^{R}(\mathbf{x})\right)^{2} \tag{11}$$

Normally, by pooling the responses of simple cells with different monocular phase and/or position shift, we obtain binocular complex cells with a specific disparity tuning, as done in our previous work [31]. To focus on 3B and 2/3A layers in the V1 in this work, we derive binocular complex cell based on the laminar cortical model in [32]. The layer 3B inhibitory cells only respond to vertical boundary. The cell membrane potentials are determined by

$$\begin{aligned} &Q_{\omega,d}^{V,L,+/-}(\mathbf{x}) \\ &= \frac{1}{r_2}\left(\left[S_{\omega}^{V,L,+/-}(x+s,y)\right]^{+}\right. \\ &\quad -\beta\left(\left[Q_{\omega,d}^{V,R,+/-}(x,y)\right]^{+} + \left[Q_{\omega,d}^{V,R,-/+}(x,y)\right]^{+}\right. \\ &\quad \left.\left. + \left[Q_{\omega,d}^{V,L,-/+}(x,y)\right]^{+}\right)\right) \end{aligned} \tag{12}$$

$$\begin{aligned} &Q_{\omega,d}^{V,R,+/-}(\mathbf{x}) \\ &= \frac{1}{r_2}\left(\left[S_{\omega,d}^{V,R,+/-}(x+s,y)\right]^{+}\right. \\ &\quad -\beta\left(\left[Q_{\omega,d}^{V,L,+/-}(x,y)\right]^{+} + \left[Q_{\omega,d}^{V,L,-/+}(x,y)\right]^{+}\right. \\ &\quad \left.\left. + \left[Q_{\omega,d}^{V,R,-/+}(x,y)\right]^{+}\right)\right) \end{aligned} \tag{13}$$

where γ_2 and β represent the decay rate and the strength of the inhibition, respectively, d is the disparity plane to which the model neuron is tuned, and s is the allelotropic shift that depends on the disparity plane. The relationship between disparity plane d and allelotropic shift s is defined in Table 1. In the experiment, refer to [32], we set $\gamma_2 = 4.5$ and $\beta = 4$.

The layer 3B binocular cells receive input from the V1 layer 4 and layer 3B inhibitory cells corresponding to the same position and disparity, determined as follows:

$$\begin{aligned} &B_{\omega,d}^{V,B,+/-}(\mathbf{x}) \\ &= \frac{1}{r_1}\left(\left[S_{\omega,d}^{V,L,+/-}(x+s,y)\right]^{+} + \left[S_{\omega,d}^{V,R,+/-}(x-s,y)\right]^{+}\right. \\ &\quad -\alpha\left(\left[Q_{\omega,d}^{V,L,+/-}(x,y)\right]^{+} + \left[Q_{\omega,d}^{V,L,-/+}(x,y)\right]^{+}\right. \\ &\quad \left.\left. + \left[Q_{\omega,d}^{V,R,+/-}(x,y)\right]^{+} + \left[Q_{\omega,d}^{V,R,-/+}(x,y)\right]^{+}\right)\right) \end{aligned} \tag{14}$$

where γ_1 and α represent the decay rate and the strength of the inhibition, respectively. In the experiment, refer to [32], we set $\gamma_1 = 0.29$ and $\alpha = 6$.

The final V1 binocular complex cells are calculated by

$$C_{\omega,d}^{V,B}(\mathbf{x}) = \left[B_{\omega,d}^{V,B,+}(\mathbf{x})\right]^{+} + \left[B_{\omega,d}^{V,B,-}(\mathbf{x})\right]^{+} \tag{15}$$

After the above multi-stage processing, we obtain $C_{\omega}^{H/V,L/R}(\mathbf{x})$, $C_{\omega}^{L/R}(\mathbf{x})$, $R_{\omega}^{B}(\mathbf{x})$ and $C_{\omega,d}^{V,B}(\mathbf{x})$ to characterize monocular simple, monocular complex, binocular simple and binocular complex cells respectively. In the next stages, these different cells are fed to the perceptual quality model to obtain the perceptual quality. The same V1 model is also used in our previous work [33] for 3D image quality prediction, but more effective perceptual features are extracted from the perceptual information in this work.

B. PERCEPTUAL QUALITY MODEL

Even though the above responses can reflect low-level contrast features, they are not enough in detecting structure information to account for the HVS sensitivity. To extract robust structure information, phase features are further extracted from the response maps. Inspired by the great success of phase congruency (PC) in extracting invariant features, we obtain the local phase instead of the gradient as in our previous work [11]. Similar to [2], by modulating the image on different scales and orientations, a set of responses

at each point $\mathbf{x}$ ($[\eta_{s,\theta}(\mathbf{x}), \xi_{s,\theta}(\mathbf{x})]$) are obtained. Then, with the local amplitude on scale s and orientation ($A_{s,\theta}(\mathbf{x}) = \sqrt{\eta_{s,\theta}(\mathbf{x})^2 + \xi_{s,\theta}(\mathbf{x})^2}$), and the local energy along orientation θ ($E_\theta(\mathbf{x}) = \sqrt{F_\theta(\mathbf{x})^2 + H_\theta(\mathbf{x})^2}$) where $F_\theta(\mathbf{x}) = \sum_s \eta_{s,\theta}(\mathbf{x})$ and $H_\theta(\mathbf{x}) = \sum_s \zeta_{s,\theta}(\mathbf{x})$, the phase congruency is computed by

$$PC(\mathbf{x}) = \frac{\sum_\theta E_\theta(\mathbf{x})}{\varepsilon + \sum_s \sum_\theta A_{s,\theta}(\mathbf{x})} \tag{16}$$

where ε is a small positive constant. As an example in Fig. 2, the response map extracted from the WN distorted image still suffers from serious local quality degradation, while the extracted PC maps success in conducting local comparisons.

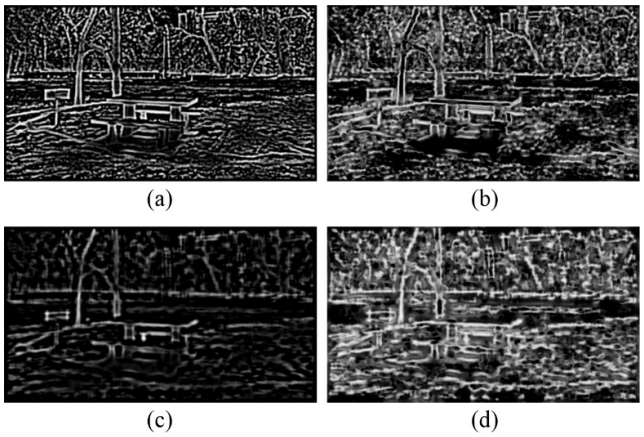


FIGURE 2. Example of original and distorted response and PC maps. (a) original response map; (b) the PC map of (a); (c) white noise distorted response map of (a); (d) the PC map of (c).

Focused on the responses extracted from the V1 model, we use the great successful similarity measure to detect the local difference between the original and its contaminated response maps. We take this similarity as microstructure similarity. Thus, let $\boldsymbol{C}_{org}(\mathbf{x})$ and $\boldsymbol{C}_{dis}(\mathbf{x})$ be the reference and distorted response maps, respectively, contrast microstructure similarity is defined as

$$\widehat{Q}_R(\mathbf{x}) = \frac{2\boldsymbol{C}_{org}(\mathbf{x}) \cdot \boldsymbol{C}_{dis}(\mathbf{x}) + T_1}{\left(\boldsymbol{C}_{org}(\mathbf{x})\right)^2 + \left(\boldsymbol{C}_{dis}(\mathbf{x})\right)^2 + T_1} \tag{17}$$

Similarly, let $\boldsymbol{E}_{org}(\mathbf{x})$ and $\boldsymbol{E}_{dis}(\mathbf{x})$ be the reference and distorted phase maps, respectively, phase microstructure similarity is defined as

$$\widehat{Q}_P(\mathbf{x}) = \frac{2\boldsymbol{E}_{org}(\mathbf{x}) \cdot \boldsymbol{E}_{dis}(\mathbf{x}) + T_1}{\left(\boldsymbol{E}_{org}(\mathbf{x})\right)^2 + \left(\boldsymbol{E}_{dis}(\mathbf{x})\right)^2 + T_1} \tag{18}$$

From another aspect, distortion will also affect the presentation of visual saliency. For example, human's fixation may change to those serious distorted regions. Therefore, the macrostructure difference between the original and distorted saliency maps is also crucial in determining the visual quality. To detect the saliency information, the input left and right images are first down-sampled by a factor of 2, and obtain the response maps and PC maps from the down-sampled inputs by the above similar process. Then, these response maps and PC maps are smoothed using a Gaussian kernel function to act as the saliency maps, denoted as $\breve{\boldsymbol{C}}_{org}(\mathbf{x})$ and $\breve{\boldsymbol{C}}_{dis}(\mathbf{x})$, $\breve{\boldsymbol{E}}_{org}(\mathbf{x})$ and $\breve{\boldsymbol{E}}_{dis}(\mathbf{x})$, corresponding to the above $\boldsymbol{C}_{org}(\mathbf{x})$ and $\boldsymbol{C}_{dis}(\mathbf{x})$, $\boldsymbol{E}_{org}(\mathbf{x})$ and $\boldsymbol{E}_{dis}(\mathbf{x})$, respectively. Fig. 3 shows the smoothed maps corresponding to Figs. 1(a)–(d).

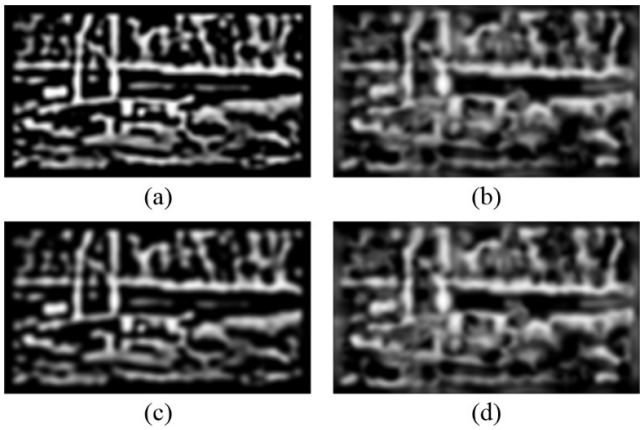


FIGURE 3. Illustration of smoothed response and PC maps corresponding to Figs. 2(a)–(d).

Then, contrast macrostructure similarity is defined as

$$\breve{Q}_R(\mathbf{x}) = \frac{\min\left(\breve{\boldsymbol{C}}_{org}(\mathbf{x}), \breve{\boldsymbol{C}}_{dis}(\mathbf{x})\right) + T_2}{\max\left(\breve{\boldsymbol{C}}_{org}(\mathbf{x}), \breve{\boldsymbol{C}}_{dis}(\mathbf{x})\right) + T_2} \tag{19}$$

Similarly, phase macrostructure similarity is defined as

$$\breve{Q}_P(\mathbf{x}) = \frac{\min\left(\breve{\boldsymbol{E}}_{org}(\mathbf{x}), \breve{\boldsymbol{E}}_{dis}(\mathbf{x})\right) + T_2}{\max\left(\breve{\boldsymbol{E}}_{org}(\mathbf{x}), \breve{\boldsymbol{E}}_{dis}(\mathbf{x})\right) + T_2} \tag{20}$$

Then, $\widehat{Q}_R(\mathbf{x})$, $\widehat{Q}_P(\mathbf{x})$, $\breve{Q}_R(\mathbf{x})$ and $\breve{Q}_P(\mathbf{x})$ are combined to get the global similarity. We define the final quality score Q_δ as follows:

$$Q_\delta = \left(\frac{1}{N}\sum_{\mathbf{x}\in\Omega} \widehat{Q}_R(\mathbf{x}) \cdot \widehat{Q}_P(\mathbf{x})\right)^\alpha \cdot \left(\frac{1}{N'}\sum_{\mathbf{x}\in\Omega'} \breve{Q}_R(\mathbf{x}) \cdot \breve{Q}_P(\mathbf{x})\right)^\beta \tag{21}$$

where α and β are two parameters to adjust the relative importance of micro- and macro-structure features, Ω means the whole spatial domain, Ω' means the sampled spatial domain, and N and N' denotes the number of the pixels in the two spatial domains.

C. 3D QUALITY POOLING

Recall that we use monocular simple cells, monocular complex cells, binocular simple cells and binocular complex cells to characterize the V1, and the similarity score in each stage is computed respectively. Taking monocular complex cells as

an example, quality scores for the left and right images are obtained first, and are combined into a binocular quality by binocular combination model

$$Q_{MS,\omega} = \mathrm{COM}(Q^L_{MS,\omega}, Q^R_{MS,\omega}) \tag{22}$$

where COM denotes the binocular combination operation. Here, we use the classical Gain-Control model [34] to address the binocular combination behavior. The weights in the Gain-Control model are computed across different scales as follows:

$$\rho^L = \frac{1+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^L_{dis}(\mathbf{x};\omega)}{1+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^L_{dis}(\mathbf{x};\omega)+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^R_{dis}(\mathbf{x};\omega)} \tag{23}$$

$$\rho^R = \frac{1+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^R_{dis}(\mathbf{x};\omega)}{1+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^L_{dis}(\mathbf{x};\omega)+\sum_{\omega\in\Omega_\omega}\sum_{\mathbf{x}\in\Omega} \boldsymbol{C}^R_{dis}(\mathbf{x};\omega)} \tag{24}$$

The Gain-Control model can accurately describe an early stage of binocular combination. In the subsequent stages, the fused binocular signals are directly perceived by the means of responses. Then, by applying average pooling on the quality scores across all scales (and across all disparity planes for the binocular complex cells), the final quality scores for the monocular simple cells, monocular complex cells, binocular simple cells and binocular complex cells are computed as

$$Q_{MS} = \underset{\omega\in\Omega_\omega}{\mathrm{AVG}}(Q^H_{MS,\omega} + Q^V_{MS,\omega}) \tag{25}$$

$$Q_{MC} = \underset{\omega\in\Omega_\omega}{\mathrm{AVG}}(Q_{MC,\omega}) \tag{26}$$

$$Q_{BS} = \underset{\omega\in\Omega_\omega}{\mathrm{AVG}}(Q_{BS,\omega}) \tag{27}$$

$$Q_{BC} = \underset{\omega\in\Omega_\omega, d\in\Omega_d}{\mathrm{AVG}}(Q_{BC,\omega,d}) \tag{28}$$

where AVG denotes the average pooling operation.

Once the quality scores for the four stages are obtained, we should combine them into a global 3D quality score. Theoretically, the influences of monocular/binocular simple/complex cells for different distortion types are not the same. As demonstrated by the comparative results in the following Subsection 3.3, the effectiveness of the quality scores on different types of distortion will be relatively different. To achieve a more natural perception (that is, they are not simply averaged), a pooling operation is performed. All the quality scores are fused to generate a 3D quality score (defined as PVC index) as:

$$Q_{PVC} = \mathrm{OP}\left(Q_{MS}, Q_{MC}, Q_{BS}, Q_{BC}\right) \tag{29}$$

where OP denotes pooling operation. Here, different weighting/multiplication/learning pooling strategies can be used. For simplicity, we use linear weighting to combine the four quality scores.

In addition to the above contrast/phase changes, luminance changes can also cause the visible distortion although they are also related to the contrast/phase changes. In the experiment, we found that only adopting PVC index cannot obtain satisfactory results for some distortions (the evaluation results will be analyzed). Therefore, besides measuring the above contrast/phase changes, we adopt additional global metric directly on the original and distorted left/right images to measure the global luminance similarity (GLS) by

$$Q^L_{GLS} = \mathrm{SSIM}(\mathbf{I}^L_o, \mathbf{I}^L_d) \tag{30}$$

Note that we have tested several popular FR IQA metrics, and find that the SSIM metric is comparatively well in addressing the luminance and contrast changes to compensate the PVC index. Similar to Eq. (22), the binocular combined GLS index is computed by

$$Q_{GLS} = \mathrm{COM}(Q^L_{GLS}, Q^R_{GLS}) \tag{31}$$

The final index is calculated by combining Q_{PVC} and Q_{GLS} into a quality score by

$$Q = \gamma \cdot Q_{PVC} + (1-\gamma)\cdot Q_{GLS} \tag{32}$$

where $0 < \gamma < 1$ is a parameter for adjusting the relative importance of the two components.

III. EXPERIMENTAL RESULTS AND ANALYSES

A. DATABASE AND EXPERIMENT DESCRIPTION

To validate the effectiveness of the proposed method, we conduct experiments on the LIVE 3D IQA Databases (Phase-I and Phase-II) [35], [36]. The LIVE 3D IQA Database Phase-I contains 20 reference stereopairs and 365 symmetrically distorted stereopairs corresponding to five distortion types: JPEG, JP2K, GBLUR, WN and Fast fading (FF). The LIVE 3D IQA Database Phase-II contains 8 reference stereopairs, and 120 symmetrically and 240 asymmetrically distorted stereopairs corresponding to the same distortion types with LIVE 3D IQA Database Phase-I.

In addition, to further verify the proposed method, the experiments are also conducted on our recent multiply-distorted stereoscopic image database (NBU-MDSID) [37], which is composed of 270 multiply-distorted stereoscopic images corrupted simultaneously by blurring, JPEG compression and noise injection symmetrically imposed on the left and right images. The database has been made available online at http://www.escience.cn/people/fshao/database.html.

Four commonly-used performance indicators, Pearson's Linear Correlation Coefficient (PLCC), Spearman's Rank Correlation Coefficient (SRCC), Kendall's Rank Correlation Coefficient (KRCC), and Root Mean-Squared Error (RMSE), are used to evaluate the IQA metrics. For the nonlinear regression, a four-parameter logistic mapping function is used:

$$DMOS_p = \frac{\beta_1 - \beta_2}{e^{(Q-\beta_3)/|\beta_4|}+1} + \beta_2 \tag{33}$$

where β_1, β_2, β_3 and β_4 are determined by using the subjective scores and the objective scores.

We compare our method with eight state-of-the-art SIQA approaches, including: Lin's scheme [8] (FI-PSNR and FI-SSIM), Benoit's scheme [39] (the $d1$ metric is used),

TABLE 2. Overall performance on LIVE 3D IQA Database Phase-I and Phase-II.

Criteria	LIVE-I				LIVE-II			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
FI-PSNR	0.8345	0.8341	0.6297	9.0360	0.6584	0.6375	0.5252	8.4956
FI-SSIM	0.8721	0.8765	0.6791	8.0236	0.6844	0.6795	0.5042	8.2295
Benoit [38]	0.8899	0.8901	0.6947	7.4786	0.7642	0.7475	0.5553	7.2806
Bensalma [25]	0.8874	0.8747	0.6756	7.5585	0.7699	0.7513	0.5761	7.2035
Chen [14]	0.9167	0.9157	0.7368	6.5503	0.9065	0.9013	0.7307	4.7663
Shao [11]	0.9228	0.9199	0.7409	6.3190	0.7585	0.7451	0.5646	7.3554
Shao [39]	0.9350	**0.9251**	0.7291	5.8155	0.8628	0.8494	0.6145	5.7058
Lin [15]	0.9242	0.9203	0.7415	6.2629	0.9113	0.8935	0.7206	4.6477
Proposed	**0.9389**	0.9308	**0.7648**	**5.6459**	**0.9263**	**0.9282**	**0.7627**	**4.1996**

TABLE 3. Performance evaluation for each individual distortion type on LIVE 3D IQA Database Phase-I.

Distortion	Criteria	FI-PSNR	FI-SSIM	Benoit[38]	Bensalma[25]	Chen[14]	Shao[11]	Shao[39]	Proposed
JPEG	PLCC	0.2187	0.4773	0.5766	0.3803	0.6344	**0.6563**	0.6533	**0.6654**
	SRCC	0.1212	0.4361	0.4983	0.3283	0.5582	0.6148	**0.6167**	**0.6339**
	KRCC	0.0735	0.2837	0.3301	0.2141	0.3718	**0.4270**	0.4243	**0.4395**
	RMSE	6.3810	6.5392	5.3429	6.0477	5.0548	**4.9341**	**5.0089**	6.4344
JP2K	PLCC	0.7851	0.8650	0.8859	0.8389	0.9164	0.9238	**0.9269**	**0.9360**
	SRCC	0.7993	0.8584	0.8730	0.8171	0.8956	0.8752	0.8789	**0.9000**
	KRCC	0.5918	0.6563	0.6791	0.6000	0.7139	0.6797	0.7000	**0.7259**
	RMSE	8.0215	6.4979	6.0073	7.0493	5.1852	4.9580	4.6080	**4.5571**
GB	PLCC	0.9095	0.9176	0.9217	0.9369	0.9417	0.9513	**0.9534**	**0.9542**
	SRCC	0.9020	0.8793	0.8802	0.9157	0.9261	**0.9375**	**0.9403**	0.9242
	KRCC	0.7333	0.7192	0.7212	0.7495	0.7717	**0.8000**	0.7576	**0.7697**
	RMSE	6.0169	5.7522	5.6146	5.0595	4.8698	**4.4606**	4.5208	**4.3308**
WN	PLCC	0.9351	0.9374	0.9354	0.9147	**0.9436**	0.9410	0.9368	**0.9441**
	SRCC	0.9316	0.9379	0.9369	0.9055	**0.9481**	**0.9431**	0.9366	0.9430
	KRCC	0.7665	0.7835	0.7816	0.7291	**0.8025**	**0.7949**	0.7329	0.7886
	RMSE	5.8963	5.7923	5.8831	6.7247	**5.5073**	5.6315	6.7650	**5.4836**
FF	PLCC	0.7026	0.7203	0.7477	0.7339	0.7580	**0.8403**	**0.8532**	0.8304
	SRCC	0.5875	0.5861	0.6242	0.6500	0.6879	**0.7814**	**0.8056**	0.7807
	KRCC	0.4168	0.4327	0.4491	0.4681	0.5118	0.5859	**0.5833**	**0.6029**
	RMSE	8.8425	8.6198	8.2509	8.4399	8.1048	**6.7363**	**6.7766**	6.9238

Bensalma's scheme [25], Chen's scheme [14], Lin's scheme [15] (Proposed-IS is used) and our two previous schemes (Shao's scheme [11] and Shao's scheme [39]). Note that FI-PSNR and FI-SSIM metrics use binocular combination behavior, Bensalma's scheme [25] uses binocular energy model, and Chen's scheme [14] and Lin's scheme [15] use cyclopean vision to measure the perceived quality.

TABLE 4. Performance evaluation for each individual distortion type on LIVE 3D IQA Database Phase-II.

	Criteria	FI-PSNR	FI-SSIM	Benoit[38]	Bensalma[25]	Chen[14]	Shao[11]	Shao[39]	Proposed
JPEG	PLCC	0.6124	0.6417	0.5328	**0.8577**	0.8422	0.7504	0.7857	**0.8506**
	SRCC	0.5234	0.6789	0.5078	**0.8461**	0.8396	0.7195	0.7367	**0.8340**
	KRCC	0.3544	0.4837	0.3592	**0.6411**	0.6364	0.5180	0.4835	**0.6466**
	RMSE	6.7968	7.3302	6.2033	**3.3302**	3.9525	4.8452	4.1394	**3.8537**
JP2K	PLCC	0.6696	0.7224	0.6467	0.6667	**0.8426**	0.8377	0.8124	**0.8768**
	SRCC	0.6557	0.7027	0.6325	0.8038	0.8334	**0.8477**	0.8234	**0.8747**
	KRCC	0.4640	0.5070	0.4499	0.6182	0.6416	**0.6635**	0.5674	**0.7011**
	RMSE	7.2909	6.7882	7.4878	9.8165	**5.2870**	5.3607	5.7167	**4.7574**
GB	PLCC	0.9100	0.8417	0.8814	0.9077	**0.9650**	0.8270	0.8645	**0.9445**
	SRCC	0.8650	0.8358	0.8545	0.8838	**0.9096**	0.8005	0.8490	**0.9241**
	KRCC	0.6667	0.6401	0.6635	0.6995	0.7441	0.6111	0.6515	**0.7739**
	RMSE	5.7723	7.5182	6.5777	5.8419	3.6532	7.8289	6.3012	4.5729
WN	PLCC	0.8547	0.9271	0.8610	**0.9436**	**0.9602**	0.8496	0.8646	0.9339
	SRCC	0.7098	0.9206	0.8569	**0.9386**	**0.9554**	0.8455	0.8589	0.9325
	KRCC	0.5454	0.7551	0.6823	**0.7809**	**0.8153**	0.6690	0.7001	0.7786
	RMSE	5.5610	4.0156	5.4488	**3.5457**	**2.9920**	5.6511	4.5764	3.8315
FF	PLCC	0.6550	0.8561	0.8472	**0.9097**	0.9097	0.8808	0.9035	**0.9330**
	SRCC	0.7660	0.8348	0.8319	0.8743	0.8890	0.8509	**0.8906**	**0.9409**
	KRCC	0.5814	0.6596	0.6455	0.7144	0.7402	0.6854	**0.7526**	**0.7958**
	RMSE	8.6955	5.9466	6.1126	**4.7779**	4.7794	5.4488	5.8804	**3.8973**

B. OVERALL PERFORMANCE ON LIVE 3D IMAGE DATABASE

To examine the performance of the proposed method in evaluating the symmetrically and asymmetrically distorted images, we reported the overall performance on LIVE 3D IQA Database Phase-I and Phase-II in Table 2, in which the case with the best performance is highlighted in bold. From Table 2, observe that the proposed method outperforms all the other comparative methods in terms of most of the performance criteria on two databases. Some other interesting phenomena can be observed from the table: 1) for the symmetrically distorted stereoscopic images, some SIQA algorithms perform quite well, and there seems to be less performance difference between the 3D methods. Due to the space limitation, we do not list the performance of 2D IQA algorithms on the LIVE 3D IQA Database Phase-I, but some well-developed algorithms also can achieve impressive predictive performance when directly applied on the two stereoscopic views and then averaged; 2) for the more challenging asymmetrically distorted stereoscopic images, Chen's scheme, Lin's scheme and our method are significantly superior to other methods. Chen's and Lin's schemes use cyclopean images to evaluate the quality, which have proved to be effective in predicting the quality of asymmetrically distorted cases. Our method is particularly effective for asymmetrically distorted stereoscopic images because the comprehensive monocular and binocular vision is considered.

To more comprehensively evaluate the prediction performance of the proposed scheme, we compare the reported results of nine methods for each distortion type. The PLCC, SRCC, KRCC and RMSE results are listed in Tables 3 and 4, where the top three metrics have been highlighted in boldface. One can see that the proposed scheme is among the top 3 metrics 8 times in terms of PLCC, 6 times in terms of SRCC, 8 times in terms of KRCC, and 6 times in terms of RMSE. Impressively, for the LIVE Phase-II database including both symmetrically and asymmetrically distorted stereoscopic pairs, our method performs the best on the JP2K, blurring and FF distortions. Therefore, performance gains on

TABLE 5. PLCC Performance of each quality component.

Database	Algorithm	JPEG	JP2K	GB	WN	FF	All
LIVE I	Q_{MC}	0.6816	0.9400	0.9512	0.9222	0.8188	0.9137
	Q_{BS}	0.6388	0.9353	0.9524	0.9256	0.8227	0.9238
	Q_{BC}	0.6164	0.9058	0.9448	0.9058	0.7169	0.8825
	Q_{PVC}	0.6657	0.9352	0.9533	0.9263	0.8293	0.9221
	Q_{GLS}	0.2741	0.8210	0.9080	0.9250	0.7297	0.8699
	Q	**0.6654**	**0.9360**	**0.9542**	**0.9441**	**0.8304**	**0.9389**
LIVE II	Q_{MC}	0.8006	0.8279	0.9697	0.9065	0.8697	0.8831
	Q_{BS}	0.8662	0.8953	0.9189	0.9447	0.9219	0.8125
	Q_{BC}	0.8700	0.7692	0.5439	0.9323	0.7311	0.6289
	Q_{PVC}	0.8291	0.8594	**0.9795**	0.9119	0.9205	0.9181
	Q_{GLS}	0.5486	0.7191	0.7250	0.9139	0.7342	0.6584
	Q	**0.8506**	**0.8768**	0.9445	**0.9339**	**0.9330**	**0.9263**

TABLE 6. SRCC Performance of each quality component.

Database	Algorithm	JPEG	JP2K	GB	WN	FF	All
LIVE I	Q_{MC}	0.6507	0.9038	0.9170	0.9264	0.7519	0.9142
	Q_{BS}	0.6109	0.8973	0.9245	0.9259	0.7699	0.9218
	Q_{BC}	0.6541	0.8858	0.9072	0.8796	0.5835	0.8890
	Q_{PVC}	0.6371	0.8992	0.9192	0.9292	0.7765	0.9256
	Q_{GLS}	0.2407	0.8222	0.8788	0.9282	0.6866	0.8559
	Q	**0.6339**	**0.9000**	**0.9242**	**0.9430**	**0.7807**	**0.9308**
LIVE II	Q_{MC}	0.7824	0.8111	0.8737	0.9056	0.8380	0.8868
	Q_{BS}	0.8563	0.8870	0.9261	0.9434	0.8902	0.8004
	Q_{BC}	0.8060	0.7660	0.5859	0.9324	0.7043	0.6541
	Q_{PVC}	0.8023	0.8505	**0.9330**	0.9121	0.9371	0.9228
	Q_{GLS}	0.5644	0.7003	0.7387	0.9091	0.7350	0.6375
	Q	**0.8340**	**0.8747**	0.9241	**0.9325**	**0.9409**	**0.9282**

the overall or individual distortions by our proposed method is indicative of its power in mimicking the visual aspects in the V1.

C. CONTRIBUTION OF EACH COMPONENTS

In fact, our method is composed of Q_{MS}, Q_{MC}, Q_{BS} and Q_{BC} to derive Q_{PVC}, and use Q_{PVC} and Q_{GLS} to derive the final score. Therefore, it is meaningful to investigate the impact of each quality component in the proposed method. To facilitate the analysis, we design the following schemes (e.g., denoting by Q_{MC}, Q_{BS}, Q_{BC}, Q_{PVC}, and Q_{GLS}) for comparison. The results of PLCC and SRCC are presented in Table 5 and Table 6. From the tables, we can see that, for different types of distortion, the influence of Q_{MC}, Q_{BS} and Q_{BC} will have large deviations in evaluating the perceptual quality. For example, for the symmetrically distorted LIVE Phase-I database, Q_{MC} and Q_{BS} are effective for most of the distortions (Q_{BS} and Q_{PVC} have the similar overall performance), while for the asymmetrically distorted cases, integrating Q_{BC} into the Q_{PVC} will largely promote the

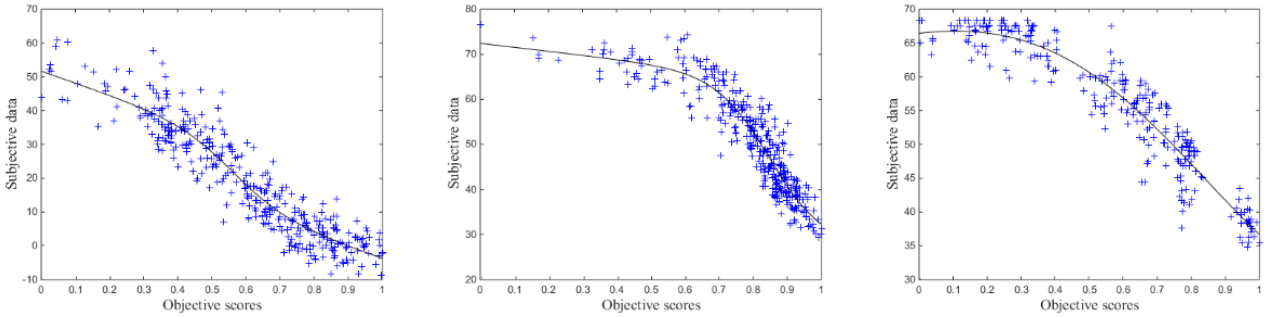


FIGURE 4. Scatter plots of objective scores versus subjective data (DMOS). (a) LIVE Database-Phase I. (b) LIVE Database-Phase II. (c) NBU-MDSID.

overall performance, even for FF distortion, Q_{BC} will have weakness in addressing the actual perception on both LIVE Phase-I and Phase-II databases. On the other hand, only using Q_{PVC} to predict the final quality is not a best choice except for blurring distortion in the LIVE Phase-II database. Therefore, by training a proper parameter λ ($\lambda = 0.285$ in this work), the performance on the overall or individual distortions can be further promoted.

TABLE 7. Overall performance on NBU-MDSID.

Criteria	PLCC	SRCC	KRCC	RMSE
FI-PSNR	0.8477	0.8461	0.6165	5.0966
FI-SSIM	0.8967	0.9227	0.7580	4.2374
Benoit [38]	0.8667	0.8313	0.7174	4.4209
Bensalma[25]	0.8564	0.8336	0.6408	4.9426
Chen[14]	0.8849	0.8772	0.6577	4.3851
Shao[11]	0.8342	0.7580	0.6585	5.3200
Shao [39]	0.9185	0.9048	0.7174	3.6867
Shao [37]	0.8781	0.8815	0.6496	4.5702
Proposed	0.9523	0.9359	0.7721	2.9209

D. PERFORMANCE ON MULTI-DISTORTED DATABASE

In addition to the above single-distorted cases, we also report the performances on our recent constructed multiply-distorted database (NBU-MDSID). Our no-reference metric designed for the database is also compared. We were surprised to find from Table 7 that, the performance of our method is significantly superior to other methods. The PLCC is promoted from 0.9185 (the highest performance of the comparative methods) to 0.9523, and SRCC is promoted from 0.9048 to 0.9359. The results further confirm that the proposed model have the capability to handle symmetrically and asymmetrically distorted stimuli, as well as multiple distorted stimuli. The scatter plots on the three databases shown in Fig. 4 demonstrate that our model has superior convergence and monotonicity.

IV. CONCLUSIONS

In this paper, we have presented a new stereoscopic image quality assessment approach by modeling visual properties of the primary visual cortex (V1). The main motivation of this work is to develop a technique that more effectively models neural processing for the visual information representation in the V1. Towards this end, we propose new feature encoding and similarity measure approaches to give more reliable and accurate quality assessment. The results demonstrated that the proposed approach can yield much better results than the other competing methods. As future work, we plan to consider deeper visual features and explore the connections them to better simulate the perception process. We also plan to extend this framework to blind reference applications.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[5] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge, MA, USA: MIT Press, 1981.

[6] J. Zhou, M. A. Georgeson, and R. F. Hess, "Linear binocular combination of responses to contrast modulation: Contrast-weighted summation in first- and second-order vision," *J. Vis.*, vol. 14, no. 13, p. 24, 2014.

[7] C. B. Huang, J. Zhou, Y. Zhou, and Z. L. Lu, "Contrast and phase combination in binocular vision," *PLoS One*, vol. 5, no. 12, 2010, Art. no. e15075.

[8] 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.

[9] 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.

[10] 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.

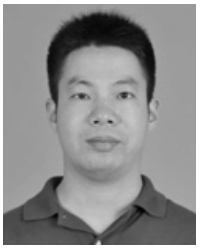
[11] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[12] K. Lee and S. Lee, "3D perception based quality pooling: Stereopsis, binocular rivalry, and binocular suppression," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 3, pp. 533–545, Apr. 2015.

[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] 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.

[15] Y. Lin, J. Yang, W. Lu, Q. Meng, Z. Lv, and H. Song, "Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 89–101, Feb. 2017.
[16] J. Yang *et al.*, "Quality assessment metric of stereo images considering cyclopean integration and visual saliency," *Inform. Sci.*, vol. 373, pp. 251–268, Dec. 2016.
[17] L. Jin, A. Boev, K. Egiazarian, and A. P. Gotchev, "Quantifying the importance of cyclopean view and binocular rivalry-related features for objective quality assessment of mobile 3D video," *EURASIP J. Image Video Process.*, vol. 2014, no. 6, pp. 1–18, 2014.
[18] 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.
[19] A. Maalouf and M.-C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 1161–1164.
[20] A. Polonsky, R. Blake, J. Braun, and D. J. Heeger, "Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry," *Nature Neurosci.*, vol. 3, no. 11, pp. 1153–1159, 2000.
[21] P. B. Hibbard, "Binocular energy responses to natural images," *Vis. Res.*, vol. 48, no. 12, pp. 1427–1439, 2008.
[22] Y. Chen and N. Qian, "A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms," *Neural Comput.*, vol. 16, no. 8, pp. 1545–1577, 2004.
[23] P. B. Hibbard, R. Goutcher, and D. W. Hunter, "Encoding and estimation of first- and second-order binocular disparity in natural images," *Vis. Res.*, vol. 120, pp. 108–120, Mar. 2016.
[24] J. C. A. Read and B. G. Cumming, "Testing quantitative models of binocular disparity selectivity in primary visual cortex," *J. Neurophysiol.*, vol. 90, no. 5, pp. 2795–2817, 2003.
[25] 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.
[26] C. Galkandage, J. Calic, S. Dogan, and J. Y. Guillemaut, "Stereoscopic video quality assessment using binocular energy," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 102–112, Feb. 2017.
[27] F. Shao, G. Jiang, M. Yu, F. Li, Z. Peng, and R. Fu, "Binocular energy response based quality assessment of stereoscopic images," *Digit. Signal Process.*, vol. 29, pp. 45–53, Sep. 2014.
[28] F. Zhang, W. Jiang, F. Autrusseau, and W. Lin, "Exploring V1 by modeling the perceptual quality of images," *J. Vis.*, vol. 14, no. 1, p. 26, Jan. 2014.
[29] M. W. Spratling, "Image segmentation using a sparse coding model of cortical area V1," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1631–1643, Apr. 2013.
[30] G. Azzopardi and N. Petkov, "A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model," *Biological*, vol. 106, no. 3, pp. 177–189, Mar. 2012.
[31] 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, Feb. 2016.
[32] S. Grossberg and P. D. L. Howe, "A laminar cortical model of stereopsis and three-dimensional surface perception," *Vis. Res.*, vol. 43, pp. 801–829, Apr. 2003.
[33] F. Shao, W. Chen, W. Lin, Q. Jiang, and G. I. Jiang, "Simulating receptive fields of human visual cortex for 3D image quality prediction," *Appl. Opt.*, vol. 5, no. 21, pp. 5488–5496, Jul. 2016.
[34] J. Ding and G. Sperling, "A gain-control theory of binocular combination," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 4, pp. 1141–1146, 2006.
[35] 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.
[36] 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. 379–391, Sep. 2013.
[37] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, "Learning sparse representation for blind quality assessment of multiply-distorted stereoscopic images," *IEEE Trans. Multimedia*, vol. 19, no. 9, pp. 1821–1836, Aug. 2017.
[38] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 389–392.
[39] 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.



FENG SHAO (M'16) received the B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in electronic science and technology. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. He was a Visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore, in 2012. He received Excellent Young Scholar Award by NSF of China in 2016. He has authored over 100 technical articles in refereed journals and proceedings in the areas of 3-D video coding, 3-D quality assessment, and image perception.

WANTING CHEN received the B.S. degree from Ningbo University, Ningbo, China, in 2015. She is currently pursuing the M.S. degree at Ningbo University. Her current research interests include image/video processing and quality assessment.

GANGYI JIANG received the M.S. degree from Hangzhou University in 1992, and the Ph.D. degree from Ajou University, South Korea, in 2000. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include digital video compression and multi-view video coding.

YO-SUNG HO (SM'06–F'16) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, South Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. He joined Electronics and Telecommunications Research Institute, (ETRI), Daejeon, South Korea, in 1983. From 1990 to 1993, he was with the Philips Laboratories, Briarcliff Manor, NY, USA, where he was involved in development of the advanced digital high-definition television system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, South Korea, where he is currently a Professor with the Information and Communications Department. His research interests include digital image and video coding, image analysis, and image restoration, advanced video coding techniques, digital video and audio broadcasting, 3-D video processing, and content-based signal representation and processing.