A full-reference stereoscopic image quality assessment index based on stable aggregation of monocular and binocular visual features

Jianwei Si1 | Huan Yang1,2 | Baoxiang Huang1 | Zhenkuan Pan1 | Honglei Su3

1 College of Computer Science and Technology, Qingdao University, Qingdao, China
2 School of Automation, Qingdao University, Qingdao, China
3 College of Electronic Information, Qingdao University, Qingdao, China

Abstract

In stereoscopic image quality assessment, human visual system has been universally taken into account to detect perceptual characteristics. A novel full-reference stereoscopic image assessment metric by considering both monocular and binocular visual features of human visual system is proposed. In particular, a new region segmentation algorithm is firstly proposed to divide 3D images into occluded and non-occluded regions. The just noticeable difference model is employed on the occluded regions to formulate the monocular vision, while the binocular just noticeable difference model is applied to the non-occluded regions to reveal the binocular vision of the human visual system. In the proposed region segmentation, disparity information and Euclidean distance between stereo pairs are both adopted to solve the unstable segmentation problem of traditional methods. A new pooling strategy based on global edge features is then presented to aggregate the just noticeable difference and binocular just noticeable difference evaluation maps. In addition, some local image features as supplementary of just noticeable difference to describe visual characteristics of the human visual system are also extracted. Finally, an overall quality score is calculated based on the above-mentioned features to measure the visual quality of distorted stereo pairs. Experimental results show that the proposed metric achieves high consistency with the human visual system, and outperforms state-of-the-art algorithms on stereoscopic image quality assessment.

1 INTRODUCTION

With development of society and technology, two-dimensional (2D) images cannot satisfy human's growing demand for perceiving the real world. Three-dimensional (3D) images become increasingly prevalent [1]. A new challenge is how to assess the quality of 3D images for various application scenarios [2–7]. Human visual system (HVS) has been seen as a vital factor to produce objective metrics for image quality assessment (IQA) [8–11]. In order to reveal perceptual characteristics of HVS, various just noticeable difference (JND) and binocular just noticeable difference (BJND) models have been proposed and employed for 2D/3D IQA [12–14].

For 2D images, JND models are proposed based on monocular vision to indicate the minimum perceptible difference threshold. Humans cannot perceive the difference if image intensity variation is less than the threshold. In 2D IQA, JND threshold map of a distorted image can be obtained to predict a quality score of the image. However, for 3D images, monocular vision and binocular vision occur simultaneously. JND models are not suitable to express both of them. BJND ones are proposed based on binocular vision, considering perceptible difference of both left and right views. BJND indicates that HVS cannot find the asymmetric distortions that are less than BJND threshold. In stereoscopic IQA (SIQA), BJND values of both left and right images are used to get the final quality score.

In order to get better performance, some researchers make efforts to integrate both monocular and binocular perceptual factors into SIQA models [14–17]. The most representative one is the model in [14] proposed by Fezza et al. They proposed a SIQA model based on both JND and BJND. Two areas named occluded (OC) region and non-occluded (NOC) region are...
divided in stereo images. For a stereo pair, these pixels that have the same information between left and right images are classified into OC region. NOC region includes other pixels. In the OC region monocular vision occurs, while in the NOC region binocular vision appears. In the OC region, JND is employed to get a predicted score and BJND is used to obtain another one in the NOC region. The final prediction result is aggregated by these two scores. Considering both monocular and binocular perception, the model in [14] achieved better performance. Apart from monocular and binocular sight, some other features are also considered in [18, 19]. For example, visual attention and natural scene statistics are included to assess image quality. Other models have also taken some features such as depth perception, disparity and binocular fusion into account [15–17].

For HVS, global and local features both play significant roles. The comprehensive consideration for HVS features can boost the performance of SIQA models. However, the characteristics of the mentioned works above are all based on the global images and have ignored some local information of stereo images. Due to the insufficient consideration, these SIQAs cannot achieve expected assessment results.

Up to now, few people have paid close attention to the segmentation of OC and NOC regions, which is also a key procedure in SIQA. An effective dividing algorithm can not only coincide with HVS better but also improve the accuracy of existing SIQAs that are based on region partition. The most commonly used region segmentation method to divide OC and NOC regions can be found in [20], which also has been used in [14, 21]. As introduced in [20], two disparity maps can be obtained for a stereo image pair. By computing the difference of these two disparity maps, a Left Right Consistency (LRC) map can be obtained by acquiring the absolute value of the difference. And each pixel in the LRC map is seen as an index of region segmentation. Finally, an appropriate threshold is assigned to determine OC and NOC regions. For a pixel in the left or right image, it can be classified into OC region if the value in the same location of its LRC map is greater than the threshold, so as other pixels. However, in the method [20], the calculation of disparity maps can be easily affected by noises, especially for images with severe distortion. For distorted stereo pairs of the same reference images, their disparity maps may vary greatly. The consequence of this is that the resulted LRC maps are unstable for region segmentation.

In this paper, we propose a full-reference (FR) SIQA model based on a novel region segmentation method and a global feature-based pooling strategy. The frame of the proposed model is illustrated in Figure 1. On the one hand, a better performed Region Segmentation method based on Disparity maps and Euclidean distance (named RSDE) is developed to divide stereo pairs into OC and NOC regions. In order to simulate monocular and binocular vision, JND [12] and BJND [13] are employed in different regions. Moreover, a Pooling Strategy based on Global Edge features (PSGE) is proposed to get a predicted score from both JND and BJND maps. On the other hand, we extract local phase and amplitude in spatial domain to get local information that is not acquired by JND and BJND. Then, we can get similarity of them and produce another predicted quality score. The final evaluation score can be obtained by aggregating the two predicted quality scores. The contributions can be summarized as follows:

1) Considering the importance of region division, we develop a novel region segmentation method based on disparity map and Euclidean distance (ED; i.e. RSDE). Compared with conventional methods based on LRC maps, segmentation results of the proposed RSDE is more stable, reducing the effect of distortions on points. Therefore, we can get a better balance between monocular and binocular vision.

2) In SIQA, monocular and binocular vision have different proportions. The proposed PSGE polling strategy provides an effective way to evaluate the visual contributions of NOC and OC regions.

3) Considering the lack of local information, some local HVS features including local amplitude and phase are extracted to get more information for better quality evaluation, which can be seen as the complementary of JND and BJND characteristics.

The remaining part of the paper is established as follows. In Section 2, we briefly review related works. Some basic foundations of the proposed model can be seen in Section 3. The proposed SIQA model is introduced in Section 4. The experiments and analysis of the proposed method are demonstrated in Section 5. Finally, the conclusion is depicted in Section 6.
In the field of SIQA, similar to 2D IQA, research works can be categorized into subjective and objective methods. Subjective ones depend on human, while objective ones provide metrics that are consistent to subjective results. Many subjective experiments have been conducted including a growing number of factors that affect prediction scores. However, due to the expensive cost consumption of human resources and time, objective SIQA methods consistent with HVS are more practical in applications.

According to the known reference information, objective SIQA can be classified into three types: FR, reduced-reference (RR) and no-reference (NR) approaches. FR SIQA models [15, 22–30] use both distorted and reference stereoscopic images to obtain visual scores. RR SIQA methods [31–33] only need partial information of the reference image to estimate SIQA scores. But for NR SIQA approaches [34–42], they can assess visual quality without any information of reference images.

Some SIQA methods adopt some well-known 2D IQA models such as Structural Similarity (SSIM) [43] and Universal Quality Index (UQI) [44] into SIQA models. Campisi et al. [22] used four 2D models SSIM, UQI, C4 [45] and RR IQA (RRIQA) [25] to assess the performance of SIQA models. In [23], SSIM and peak signal-to-noise ratio (PSNR) are used directly by Yasakethu et al. to assess stereoscopic images. The quality score is the weighted average of two scores. In [24], Kuo et al. obtained a reference cyclopean image and a distorted one by combining left and right images. Then, information content weighting SSIM (IW-SSIM) [46] is used to calculate the similarity between the two images. In [13], apart from cyclopean as [24], depth perception is considered by Liu et al. Two depth maps and cyclopean images can be obtained from reference and distorted stereo pairs. Then, MS-SSIM [47] is used to get the similarity scores between depth maps and cyclopean images, respectively. The combination of the scores is seen as the final quality score. In [48], SSIM and C4 are applied for disparity maps and stereo views. Fan et al. [49] used disparity maps to produce cyclopean images. And, UQI is used to calculate the similarity between cyclopean images and disparity maps, respectively. The quality score is the pooling of UQI results and JND values for cyclopean images. Many 2D IQA models are adopted to construct SIQA models, and a comprehensive performance comparison can be found in [50]. In addition, Zhou et al. [34] proposed an NR-SIQA model based on binocular combination and an learning machine. Two binocular combinations of stimuli are generated using different combination strategies. Then some quality-aware features of the combinations are extracted and used to produce the final quality result. In [35], Kim et al. proposed a patch-based Convolutional Neural Network (CNN) to extract features. A series of stereo images are divided into a desirable number of patches to produce training data. Then they are fed into the network and get the quality score of each patch. Finally, the patch quality results of the image are pooled into the overall quality score of it. Ding et al. [36] extracted features from cyclopean images and disparity maps using a two-column dense CNN model, and fuses these features together to get the final SIQA score of stereo images. Yue et al. [40] proposed an NR-SIQA method via obtaining the quality-sensitive features from cyclopean phase map and applying an Support Vector Regression (SVR) model to combine the features. In [41], Liu et al. proposed an NR-SIQA method based on hierarchical learning. The stereoscopic images are divided into different groups according to distortion firstly. And the classified image group is fed into image quality predictor including five different perceptual channels. Finally, an Support Vector Machine (SVM) is applied to combine the outputs of five channels and obtain quality assessment results. Oh et al. [51] proposed a deep NR IQA model in terms of local and global feature aggregation. A CNN-based model is adopted to extract features of left and right patches and get the final quality score. These CNN-based methods can extract features automatically, but may ignore the relationship between patches of stereo pairs. And attributed to enclosed feature extraction, these CNN-based models are lack of interpretability, which can lead to difficult optimizations of various networks. In addition, the training of CNNs need a high time complexity. In contrary, traditional metrics have been studied widely due to the advantages such as higher interpretability, lower time complexity and so on. Hence, here, we mainly focus on investigating some effective artificial features on SIQA, and want to explore more aggregation manners to reveal monocular and binocular mechanism of the HVS.

Some SIQA models also take some other HVS characteristics into account. JND-based models have attracted great interest recently. For conventional 2D images, JND models have been proposed to indicate maximum image distortions that are undetectable to HVS [52–59]. They can be classified into two categories: frequency domain and spatial domain. These models in frequency domain often transfer an image into sub-band representation, for example, Discrete Cosine Transform (DCT) and wavelet decomposition [52–54]. Models in spatial domain seem to be more convenient because these models are calculated directly on pixels in an image [57–61]. Therefore, we focus on spatial (pixel-wise) models. For pixel-wise JND models, some HVS features such as Luminance Adaption (LA) and Contrast Masking (CM) effects are often considered. Based on both of them, early JND models are proposed by Chou et al. [57, 58]. Further, Zhang et al. [59] used Contrast Sensitivity Function (CSF) [62] as well as LA and CM to obtain JND thresholds. Liu et al. [60] developed a more accurate JND model on the basis of [59] in the advantages of edge masking and texture masking. Wang et al. [61] proposed a JND model for screen content images based on edge profile, which is decomposed into three features including luminance, contrast and structure information.

But JND models are not suitable for stereoscopic images. In order to keep compatible for stereo pairs, some stereo characteristics such as binocular vision and depth features must be considered. For 3D images, several studies have concentrated on finding a visibility threshold for 3D images/videos. Zhao et al. [13] proposed a BJND model considering LA and CM. Silva et al. [63, 64] studied JND in depth perception for 3D video displays. And the model can be utilized to get depth perception when the depth information is lost or lossy compressed.
Some other features are considered in [18, 19, 65–68]. Based on free energy principle and HVS, Zhu et al. [65] imported 2D IQA methods into SIQA ones by using dual-weight model. In [66, 67], Liu et al. and Kang et al. used depth component, structure component and gradient component to establish their SIQA models. Khan et al. [68] used the sparse representations of luminance images and depth maps to obtain SIQA scores.

3 | FOUNDATIONS OF ALGORITHM

3.1 | Just noticeable difference (JND)

In recent works, JND is used widely in image processing to keep consistent with HVS. In the OC region, single sight occurs. In order to maintain high consistency with HVS, JND is used in our model to simulate monocular vision [12]. It is acquired by LA and CM, which can be calculated as follows:

\[
JND(u, v) = LA(u, v) + CM(u, v) - 0.3 \times \min \{LA(u, v), CM(u, v)\},
\]

where \((u, v)\) represents a pixel location in an image, LA which represents LA of HVS, can be described as follows [69]:

\[
LA(u, v) = \begin{cases} 
17 \times (1 - \sqrt{(I(u, v) - 127)/128}) + 3, & \text{if } I(u, v) \leq 127 \\
3 \times (I(u, v) - 127)/128 + 3, & \text{otherwise},
\end{cases}
\]

where \(I(u, v)\) represents the pixel value at location \((u, v)\).

CM, as another phenomenon of HVS, indicates the sensitivity of different feature changes (i.e. spatial frequency, direction and location) [70]. It is represented as follows:

\[
CM(u, v) = EM(u, v) + TM(u, v),
\]

where \(EM\) and \(TM\) are edge masking and texture masking, respectively. In image decomposition processing, an image \(I\) is seen as the sum of the structural image \(I_s\) and the textural image \(I_t\) (i.e. \(I = I_s + I_t\)) [71]. Let \(C_s\) and \(C_t\) denote the maximum luminance contrast of \(I_s\) and \(I_t\). \(EM\) and \(TM\) can be obtained as follows:

\[
\begin{align*}
EM(u, v) &= C_s(u, v) \times \beta \times \rho_s, \\
TM(u, v) &= C_t(u, v) \times \beta \times \rho_t,
\end{align*}
\]

where \(\rho_s\) and \(\rho_t\) are signs of \(EM\) and \(TM\), which are set as 1 and 3, the constant \(\beta\) is assigned to 0.117.

3.2 | Binocular just noticeable difference (BJND)

In NOC regions, binocular vision occurs. BJND is the best choice to model visual factors mentioned in the last section. Similar to JND, CM and LA are also considered in BJND models. The overall BJND can be defined as follows:

\[
BJND(u, v) = T_{I_{1/1632}}(u, v) \times \left[ 1 - \left( \frac{N_a}{L_{I_{1/1632}}(u, v)} \right)^{1/\lambda} \right],
\]

where the parameters \(N_a\) and \(\lambda\) are equal to 0.3 and 3.76. \(T_{I_{1/1632}}(u, v)\) is the maximum JND threshold in location \((u, v)\) of the left or right image, and it can be calculated as follows:

\[
T_{I_{1/1632}}(k) = \begin{cases} 
\frac{c_1 L_{20\theta}(u, v) + c_2 L_{30\theta}(u, v) + c_3 d(u, v) + c_4}{c_5 L_{20\theta}(u, v) + c_6 L_{30\theta}(u, v) + c_7 d(u, v) + c_8}, & \text{if } 0 \leq L_{20\theta} < t_b \\
c_5 L_{20\theta}(u, v) + c_6 L_{30\theta}(u, v) + c_7 d(u, v) + c_8, & \text{otherwise},
\end{cases}
\]

where \(L_{20\theta}(u, v)\) is the luminance of background in location \((u, v)\) of stereo pairs and it is filtered by a filter \(B\) in Equation (7). \(d\) represents the disparity between left and right images. The different constants \(c_1 \sim c_8\) are set to \(4.299 \times 10^{-4}, -0.42, 14.792 \times 10^{-4}, 9.66, 0.986 \times 10^{-4}, 57.842, 2.333 \times 10^{-4}, 2.585\), respectively. The parameter \(t_b\) is assigned to 33.

4 | PROPOSED SIQA MODEL BASED ON RSDE AND HVS CHARACTERISTICS

As mentioned in Section 1, existing SIQA works based on JND and BJND have strong dependence on regional division. And the conventional region segmentation method based on LRC maps (RSLM) has shown its weaknesses. We propose a novel FR SIQA model in which an effective region segmentation method is proposed to divide OC and NOC regions. And, a PSGE is presented to fuse JND and BJND scores on OC and NOC regions. Moreover, as supplement of JND and BJND evaluation, some local features (i.e. local phase and amplitude) are also adopted to predict the visual quality of distorted images. Finally, evaluation results of these two parts are combined together to indicate the real quality scores of distorted images.

4.1 | Proposed RSDE algorithm for region segmentation

As mentioned in Section 1, traditional methods (e.g. RSLM) to divide stereo images into OC and NOC regions have shown disadvantages, for example, they cannot obtain stable LRC map and get the expected result of region segmentation. ED, which represents the distance between two objects, has been widely adopted to reflect the similarity between two locations of a
Therefore, we develop a novel region segmentation algorithm (i.e., RSDE) based on disparity maps and ED to distinguish OC and NOC districts and to get better results of region segmentation. For a distorted stereo pair in FR SIQA, its reference pair can be obtained. Based on the reference left and right images, we can get two disparity maps ($d_l$ and $d_r$) as [20]. As shown in Figure 2, besides the disparity information, we also consider the impact of image distortions on the region segmentation. The similarity of distorted left and right images can be evaluated by ED. In order to solve the problem that the results of ED calculation may be effected by different distortions on points or some points whose disparity are wrongly calculated, we calculate Local Euclidean Distance (LED) with a $5 \times 5$ window. The LED of one pixel in distorted left view and its responding pixel in the right one can be obtained as Equation (8). The right LED map related to the distorted right image can also be computed as Equation (9).

$$E_l(u, v) = \sum_{i=-2}^{2} \sum_{j=-2}^{2} (P(u + i, v + j) - P'(u + i, v + d_l(u, v) + j))^2,$$

where $E_l(u, v)$ and $E_r(u, v)$ are two LED maps on the basis of distorted left and right images, respectively. $P$ represents the matrix of $(u, v)$ with $5 \times 5$ neighbourhoods in distorted left(right) view and $P'$ represents the matrix of $(u, v \pm d_l(u, v))$ with $5 \times 5$ neighbourhoods in responding right(left) view. $d_l$ and $d_r$ are disparity maps of the reference pair.

As to the two LED maps above, less ED values indicate higher similarity [72]. For stereoscopic pairs, the number of pixels belonging to the NOC region is much more than that attributed to the OC region. In other words, for a random pixel in one view, the probability that it belongs to NOC region is greater. Based on the above, to the greatest extent we can accept the truth that pixels are divided into the NOC region by mistake. Therefore, we can get an LED map that is used to divide regions by a minimum strategy as follows:

$$LED(u, v) = \frac{\min\{E_l(u, v), E_r(u, v)\}}{255}.$$
Finally, an appropriate threshold, which is viewed as the bound of different areas, is used to get the ultimate result of RSDE as follows:

\[
\text{RSDE}_{(u,v)} = \begin{cases} 
\text{OC region,} & \text{if } LED(u,v) \geq \theta_t \\
\text{NOC region, otherwise,} &
\end{cases}
\] (11)

where \( \theta_t \) is the threshold to divide areas.

### 4.2 PSGE pooling strategy

Based on the region segmentation map \( RSDE (u,v) \), JND model is then carried out to evaluate the visual distortion of the OC region while BJND is employed on the NOC region. And a novel proposed pooling strategy PSGE is used to pooling different parts and a predicted quality score reflecting monocular and binocular characteristics can be obtained after the congregation.

In the proposed PSGE pooling strategy, global edge features are extracted as weights to evaluate the visual contributions of NOC and OC regions. We use Global Contrast (GC) and Global Width (GW) to reflect edge properties.

For a distorted image and its reference image, let \( \otimes \), \( a \), \( w \), \( b \) denote the convolution operation, the contrast parameter, the width parameter, and the basis parameter, respectively, a practical model \( s(u; a, b, \tau, n_0) \) used to extract GC map and GW map by convoluting the ideal edge model \( e(\xi; b, c, x_0) \) with the Gaussian kernel \( g(\xi; w) \) [73] can be seen in Equation (12). And the ideal edge model can be described as follows:

\[
e(u; a, b, \tau, n_0) = a \cdot U(u - n_0) + b,
\] (13)

where \( U(\cdot) \) denotes the unit-step function. And \( \text{erf}(\cdot) \) is the error function as follows:

\[
\text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-\eta^2} d\eta.
\] (14)

And the parameters \( a \) and \( \tau \) can be defined as follows:

\[
a = d_1 \cdot \sqrt{2\pi / \ln(l_1)} \cdot l_2^{1/4},
\] (15)

\[
\tau = \sqrt{1/\ln(l_1) - \sigma_d^2},
\] (16)

where \( l_1 = d_0^2 / d_3 d_3 \) and \( l_2 = d_3 / d_1 \). And they can be defined with the derivative of the Gaussian filter \( g(u, \sigma_d) \) as follows:

\[
d(u; a, \tau, \sigma_d, n_0) = \frac{a}{\sqrt{2\pi (\tau^2 + \sigma_d^2)}} \cdot \exp \left[ -\frac{(u - n_0)^2}{2(\tau^2 + \sigma_d^2)} \right].
\] (17)

Here we set \( d_1 = d(0; a, \tau, \sigma_d, n_0) \), \( d_2 = d(1; a, \tau, \sigma_d, n_0) \) and \( d_3 = d(-1; a, \tau, \sigma_d, n_0) \), respectively.

After calculating feature maps \( \text{GC} \) and \( \text{GW} \) as Equation (12), the similarity between left (right) distorted and reference images can be calculated in the same way as follows:

\[
S_{\text{GC}} = \frac{2 \cdot S_{\text{GC}} + \varepsilon_m}{(S_{\text{GC}}^2 + S_{\text{GW}}^2)^{1/2} + \varepsilon_m}, m \in \{ \text{GC, GW} \},
\] (18)

where \( S_{\text{GC}}^2 \) and \( S_{\text{GW}}^2 \) represent the GC and GW map of distorted and reference left (right) image. And the weights of left (right) distorted image can be obtained as follows:

\[
PSGE_{\text{LC}} = \left( S_{\text{GC}}^2 \right)^{\theta_1} \cdot \left( S_{\text{GW}}^2 \right)^{\theta_2},
\] (19)

where \( \theta_1 \) and \( \theta_2 \) are both set to 1.

And a result of quality score in OC region can be defined as follows:

\[
Q_{\text{OC}} = \frac{\sum J_{\text{NC}} \cdot PSGE_i}{\sum J_{\text{NC}}},
\]

In a similar way, we can get a quality score in the NOC region as follows:

\[
Q_{\text{NOC}} = \frac{\sum J_{\text{NC}} \cdot PSGE_i}{\sum J_{\text{NC}}},
\] (21)

### 4.3 Local feature description for quality evaluation

Image local features are also of great importance in SIQA. Phase, as an important characteristic of images, has been applied in image processing due to the advantages of high stability and adaptability [74]. However, different regions are evaluated independently in traditional methods, in this way global phase from direct Fourier transform is not an effective measure to express local features [21]. Moreover, existing works have shown that log-Gabor is an excellent choice to imitate simple cells in binocular vision [75–77]. Thus, we calculate PC for distorted and reference images by maximizing the Fourier components in phase with log-Gabor filter.

Let \( [\eta_{\text{LC}}, \xi_{\text{LC}}] \) denote log-Gabor response on spatial scale \( s \) and orientation \( \theta \) for distorted and reference images, we can...
get different responses on four scales and four orientations by applying the log-Gabor filter \(G_{\omega, \theta}(r, \omega)\) in the Fourier frequency domain:

\[
G_{\omega, \theta}(r, \omega) = \exp \left[ -\frac{(\log(\omega \setminus \omega_0))^2}{2\sigma_\omega^2} \right] \cdot \exp \left[ -\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2} \right],
\]

where the parameters \(\omega\) and \(\theta\) are the normalized radial frequency and the orientation angle of the filter, and \(\omega_0\) and \(\theta_0\) are the corresponding centre frequency and orientation of the filter, respectively. The parameters \(\sigma_\omega\) and \(\sigma_\theta\) determine the strength of the filter. Then the Local aMplitude (LM) at location \(x\) on scale \(s\) and along orientation \(\omega\) can be calculated as follows:

\[
LM_{\text{local}}(x) = \sqrt{\sum_{i} \eta_{i,s}(x) + \sum_{i} \xi_{i,s}(x)},
\]

and the local energy along orientation \(\omega\) can be computed as follows:

\[
LE_{\omega}(x) = \sqrt{\left(\sum_{i} \eta_{i,\omega}(x)\right)^2 + \left(\sum_{i} \xi_{i,\omega}(x)\right)^2}.
\]

The PC along orientation \(\omega\) can be computed as follows:

\[
PC_{\omega}(x) = \frac{LE_{\omega}(x)}{\varepsilon_1 + \sum_s LM_{\text{local},s}(x)},
\]

where \(\varepsilon_1\) is a small constant that is greater than 0.

However, we use Local Phase (LP) and LM to describe local features rather than PC directly. Let \(RF_{\omega}(x) = \sum_i \eta_{i,\omega}(x)\) and \(RH_{\omega}(x) = \sum_i \xi_{i,\omega}(x)\). The LP is defined as the angle between \(RF_{\omega}(x)\) and \(RH_{\omega}(x)\) along the orientation with the maximum PC value:

\[
LP_{\text{local}}(x) = \arctan(RF_{\omega}(x), RH_{\omega}(x)),
\]

where \(\omega_m\) denotes the orientation which corresponds to the maximum PC value. Then the local amplitude is defined as the sum of local amplitudes of all the scales along the orientation \(\omega_m\):

\[
LM_{\text{local}}(x) = \sum_s LM_{\text{local},s}(x)
\]

In our experiment, \(\omega_m = 1/6, \theta_m = 0, \sigma_r = 0.3, \sigma_s = 0.4\), and the numbers of scale and orientation of the filter are set to 4.

### 4.4 Overall quality estimation

Based on score maps of OC and NOC regions as Equation (20) and Equation (21), we can get a predicted score of monocular and binocular characteristics. Moreover, In Section 4.3, we have obtained LM and LP maps of distorted and reference images. In this section, we get similarities of LM and LP maps and pool them into another predicted score. Finally, a total quality assessment result can be obtained by combining both of them.

The predicted score of monocular and binocular characteristics can be obtained as follows:

\[
Q_{bf} = W_{\omega_{\text{oc}}} \cdot Q_{\omega_{\text{oc}}} + W_{\omega_{\text{nc}}} \cdot Q_{\omega_{\text{nc}}},
\]

where \(W_{\omega_{\text{oc}}}\) and \(W_{\omega_{\text{nc}}}\) are weight maps of OC and NOC regions. Let \(M, N\) denote the number of pixels of OC and NOC region, respectively, and they can be obtained as follows:

\[
W_{\omega_{\text{oc}}} = \frac{1}{M} \cdot e^{Q_{\omega_{\text{oc}}} - \frac{1}{N} \cdot e^{Q_{\omega_{\text{nc}}}}},
\]

\[
W_{\omega_{\text{nc}}} = \frac{1}{N} \cdot e^{Q_{\omega_{\text{nc}}} - \frac{1}{M} \cdot e^{Q_{\omega_{\text{oc}}}}},
\]

For LA and LM maps, we can get similarity maps by using Equation (18) as follows:

\[
S_{\omega_{\text{LM}}}^{i,j} = \frac{2 \cdot LM_{\text{local},i,j}^{\text{ref}} \cdot LM_{\text{local},i,j}^{\text{dis}} + \varepsilon_2}{\left(LM_{\text{local},i,j}^{\text{ref}}\right)^2 + \left(LM_{\text{local},i,j}^{\text{dis}}\right)^2 + \varepsilon_2},
\]

\[
S_{\omega_{\text{LP}}}^{i,j} = \frac{2 \cdot LP_{\text{local},i,j}^{\text{ref}} \cdot LP_{\text{local},i,j}^{\text{dis}} + \varepsilon_2}{\left(LP_{\text{local},i,j}^{\text{ref}}\right)^2 + \left(LP_{\text{local},i,j}^{\text{dis}}\right)^2 + \varepsilon_2},
\]

where \(\text{ref}\) and \(\text{dis}\) denote reference and distorted stereo images. \(\varepsilon_2\) is a constant and it is set to 0.01. The similarity maps of left (right) views can be obtained as follows:

\[
S_{\omega_{\text{LM}}}^{i,j} = S_{\omega_{\text{LM}}}^{i,j} \cdot W_{\omega_{\text{LM}}}^{i,j} + S_{\omega_{\text{LP}}}^{i,j} \cdot W_{\omega_{\text{LP}}}^{i,j},
\]

where \(W_{\omega_{\text{LM}}}^{i,j}\) and \(W_{\omega_{\text{LP}}}^{i,j}\) are weights of \(S_{\omega_{\text{LM}}}^{i,j}\) and \(S_{\omega_{\text{LP}}}^{i,j}\) by log-Gabor filter, respectively. Then the prediction score of local features can be obtained as follows:

\[
Q_{ij} = \frac{\sum S_{ij} + \sum S_{ij}}{2 \cdot \Omega},
\]

where \(\Omega\) denotes the number of pixels for a local feature map and it can be obtained as follows:

\[
\Omega = M + N.
\]

Finally, we can get the predicted score of our proposed model as follows:

\[
Q = f_1 \cdot e^{Q_{ij}} + f_2 \cdot e^{Q_{ij}^{-1}},
\]
where \( f_1 \) and \( f_2 \) are constants and \( f_1 + f_2 = 1 \). Here, we set \( f_1 = f_2 = 0.5 \).

5 | EXPERIMENTS

5.1 | Description of SIQA database

As mentioned in Section 2, more and more subjective experiments have been conducted in the case of taking a growing number of factors that affect prediction scores into account. In order to achieve the same evaluation criteria of objective SIQA models, many databases have been created. Some prevalent databases used widely in SIQA are LIVE 3D Phase I [78], LIVE 3D Phase II [79], IEEE-SA DB [80] and IVC [81]. We use LIVE 3D Phase I and Phase II to assess our proposed model.

Similar to [82], five same symmetric distorted types such as JPEG, JPEG2000 (JP2K), Gaussian White Noise (WN), Gaussian blur and Raleigh Fast Fading (FF) have been used to establish LIVE 3D Phase I. For all reference images, they are all distorted by five different distortions at different extents. But for left and right images of an image pair, both of them are all distorted at the same degree. Here, the number of 20 reference images and 365 distorted images, which includes 80 pairs each for JPEG, JP2K, WN, FF and 45 for Blur, have been the components of Phase I.

Similar to Phase I, Phase II consists of 8 reference images and 360 distorted images with co-registered human scores in the form of DMOS. For each reference stereo pair, they are all processed to three symmetric distorted stereo pairs and six asymmetric distorted stereo pairs.

5.2 | Performance index

Three widely used indicators including Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC) and Root Mean Squared Error (RMSE) between subjective and objective scores are applied to assess performance of our proposed model after non-linear regression. And we use a four-parameter logistic mapping function for non-linear regression [83] as follows:

\[
DMOS_p = \frac{b_1 - b_2}{\frac{e^{x|a|}}{a} + 1} + b_2, \tag{37}
\]

where \( b_1, b_2, b_3 \) and \( b_4 \) are constant parameters.

For three indexes above, the thresholds of PLCC and SROCC are both [0 1] while the value of RMSE is greater than 0. A higher value of PLCC and SROCC indicates a better correlation between subjective and objective prediction results. But for RMSE, a lower value reflects the same result.

In order to display the performance of our proposed model, 15 existing models SSIM [43], MS-SSIM [47], eight FR models including Fezza [14], Shao [21], Lin [26], Chen [27], Ma [28], Liu [15], Benoit [81] and Chen [29], and five NR models including Kim [35], Chen [37], Yue [40], Liu [41] and Oh [51] are taken into account as comparison objects. Some details can be seen in Section 5.5.

5.3 | Performance of region segmentation

We propose a novel region segmentation method (i.e. RSDE) based on disparity and ED. Thinking of the effects of the points wrongly calculated, we calculate the ED between the points with 5×5 neighbourhoods. In order to prove the effectiveness of the different sizes of neighbourhoods, we provide the results of various sizes of neighbourhoods including 3×3, 5×5, 7×7 and 9×9 in our model, which can be seen in Table 1. In the table, apart from the difference of sizes, other conditions are all the same to our model in Section 4. From the table, we can draw a conclusion that the size of 5×5 neighbourhoods have the best result of our model.

As for the threshold in Equation (11), we test different thresholds from 0.02 to 0.20 for our model. In Table 2, some well-performed results are shown here. All the tests are based on the condition that the only variable is threshold and other conditions are all the same as our model in Section 4. From the table, we can clearly find that the threshold 0.12 has the best performance.

In Figure 3, we provide results of the conventional method RSLM [20] and our RSDE. Figures 3(a) and 3(b) are four stereoscopic pairs of different scenes selected randomly in LIVE 3D Phase I. Figure 3(c) contains the results of RSLM to divide regions while Figure 3(d) presents results of RSDE in Section 4.1. For the results above, white area represents NOC region while black area denotes OC region. To prove the effectiveness of RSDE, we do a performance comparison of SIQAs by using RSLM and RSDE for region segmentation in Table 3. From the table, we can learn that the values of PLCC and SROCC by RSDE are more than that of conventional method, which means that a better performance of the SIQA model is shown by using our proposed RSDE.

In order to prove the stability of RSDE, we also use the third stereo pair of Figure 3 as reference images. Then for each of

| TABLE 1 | Performance of different size of neighbourhoods |
|----------|-----------------------------------------------|
| Size of neighbourhoods | 3×3 | 5×5 | 7×7 | 9×9 |
| PLCC     | 0.9371 | 0.9441 | 0.9435 | 0.9420 |
| SROCC    | 0.9346 | 0.9426 | 0.9430 | 0.9400 |
| RMSE     | 5.8269 | 5.3128 | 5.2725 | 5.3506 |

| TABLE 2 | Performance of different thresholds of RSDE |
|----------|------------------------------------------|
| Threshold | 0.10 | 0.11 | 0.12 | 0.13 | 0.14 |
| PLCC     | 0.9371 | 0.9410 | 0.9441 | 0.9370 | 0.9300 |
| SROCC    | 0.9346 | 0.9390 | 0.9426 | 0.9336 | 0.9286 |
| RMSE     | 5.8269 | 5.3523 | 5.3128 | 6.1516 | 7.2315 |
the distortions including JP2K, JPEG, WN, FF and Blur, we randomly select one stereo pair and get five stereo pairs. In Figures 4 and 5, we provide region segmentation results of different methods including RSLM and RSDE for the pairs. And all of the results are all mapped to [0 255]. By using RSLM, Figure 4(a) shows the result of reference images and Figures 4(b)–4(f) show the results of different distortions. Figure 5 shows the same things of RSDE. From the figures, we can learn about that compared with the similarity between (a) and any other one such as (b), (c), (e) and (f) in Figure 5, the similarity in Figure 4 is lower, which means the effectiveness of LED maps with 5x5 window and the stability of proposed RSDE. But for WN distortion, the similarity between Figures 4(a) and 4(c) is higher than that between Figures 5(a) and 5(c), which means that the proposed RSDE is not highly effective for White Noise.

5.4 Performance of local feature extraction

Feature extraction is one of the most important processes in SIQA. An excellent result of feature extraction is not only benefit to the improvement of accuracy for SIQA models but also instrumental in coinciding with subjective results. Here, the image in Figure 3(a) is selected as the original image to show the results of local feature extraction. We provide the maps of LP and amplitude mapped to [0 255] in Figures 6 and 7. Figures 6(a)–6(f) show the LP maps of original image and its five distorted images including the distortions of JP2K(b), JPEG(c), WN(d), FF(e) and Blur(f). Similar to Figure 6, Figure 7 shows the amplitude maps of the same original image. In order to demonstrate the effects of different distortions, distorted images with the similar values of DMOS are selected here.

As can be seen from Figure 6, it is obvious that the similarity between (a) and (d) is lower than that between (a) and any other one such as (b), (c), (d) and (f). Except (e) with little important structural visibility, other maps (b, c, e, f) all have a high similarity with (a), which means that LP is a fantastic description of features in SIQA due to the preservation of essential structure information. At the same time, we can find the same situation in Figure 7. The LM maps in (b), (c), (e), (f) are all more similar with (a) than (d). These two alike phenomenons indicate that the method to extract features using PC may be not highly effective for White Noise, which may influence the results of the proposed model.

In order to prove the effectiveness of feature extractions, four different schemes are used in Phase I. JND and BJND are all included in all schemes. For Scheme I, only JND and BJND are applied to produce the final prediction scores as Equation (28). For Scheme II, besides JND and BJND, only LP is used to calculate the local scores. For Scheme III, only local amplitude is used and other operations are the same as Scheme I. But for JP2K and WN distortion, the performance of the proposed model is less than Scheme II and Scheme III,

| Criteria  | With RSLM | With RSDE |
|-----------|-----------|-----------|
| PLCC      | 0.9350    | 0.9441    |
| SROCC     | 0.9245    | 0.9426    |
| RMSE      | 5.3551    | 5.3128    |
which can be attributed to the fact as follows. Both the distortions are produced by energy calculation, therefore a single energy-based way for feature extraction such as local amplitude and phase have a better performance than the proposed model consisted of both energy-based features and other features. For Scheme IV, it is made up of JND, BJND and PC in Equation (25). The performance results of them and our proposed model can be seen in Tables 4 and 5. From the tables we can learn that the result of Scheme I has lower PLCC and SROCC than others, regardless of some structure features. For Scheme II, Scheme III and Scheme IV, we can see that Scheme II and Scheme III have higher results than Scheme IV because the use of global PC is not a best choice to extract features as mentioned in Section 4.3. And compared with Scheme II and Scheme III, we can get a conclusion that in the process of feature extractions, LP has more contribution than local amplitude. The proposed model has the best result than others by considering both HVS characteristics and local features.

5.5 | Performance of proposed model

All experiments are done in LIVE 3D Phase I and Phase II mentioned in Section 5.1. Besides, PLCC, SROCC and RMSE mentioned in 5.2 are regarded as indexes to measure the effectiveness of different methods.
**TABLE 4** PLCC performance of different schemes for different distortions

|       | Scheme-I | Scheme-II | Scheme-III | Scheme-IV | Proposed |
|-------|----------|-----------|------------|-----------|----------|
| JP2K  | 0.8935   | 0.9043    | 0.8873     | 0.8832    | 0.9450   |
| JPEG  | 0.5996   | 0.5948    | 0.5885     | 0.4950    | 0.7456   |
| WN    | 0.9179   | 0.9293    | 0.9428     | 0.9414    | 0.9330   |
| Blur  | 0.9173   | 0.9315    | 0.9041     | 0.8609    | 0.9609   |
| FF    | 0.6561   | 0.6573    | 0.6132     | 0.5158    | 0.9550   |
| ALL   | 0.9043   | 0.9310    | 0.9232     | 0.8641    | 0.9441   |

**TABLE 5** SROCC performance of different schemes for different distortions

|       | Scheme-I | Scheme-II | Scheme-III | Scheme-IV | Proposed |
|-------|----------|-----------|------------|-----------|----------|
| JP2K  | 0.9206   | 0.9387    | 0.9250     | 0.8731    | 0.9247   |
| JPEG  | 0.5993   | 0.6681    | 0.6659     | 0.5220    | 0.7315   |
| WN    | 0.9241   | 0.9315    | 0.9430     | 0.9380    | 0.9209   |
| Blur  | 0.9533   | 0.9497    | 0.9174     | 0.9024    | 0.9530   |
| FF    | 0.7642   | 0.7726    | 0.7252     | 0.6726    | 0.9548   |
| ALL   | 0.9149   | 0.9237    | 0.9190     | 0.8570    | 0.9426   |

**FIGURE 6** Local phase maps extracted from the original and distorted images

**FIGURE 7** Local amplitude maps extracted from the original and distorted images
### TABLE 6  Overall performance on Live 3D image database

| Methods | LIVE 3D Phase I | | LIVE 3D Phase II | |
|---|---|---|---|---|
| | PLCC | SROCC | RMSE | PLCC | SROCC | RMSE |
| Conventional models | | | | | |
| SSIM [43] | 0.8725 | 0.8767 | 10.6841 | 0.7467 | 0.7292 | 8.0074 |
| MS-SSIM [47] | 0.9261 | 0.9239 | 8.2486 | 0.7824 | 0.7774 | 8.0217 |
| FR models | | | | | |
| Fezza [14] | 0.9250 | 0.9280 | 9.0500 | 0.8812 | 0.8726 | 8.3115 |
| Shao [21] | 0.9245 | 0.9217 | 6.2522 | 0.7585 | 0.7451 | 7.3554 |
| Lin [26] | 0.8645 | 0.8559 | 8.2424 | 0.6584 | 0.6375 | 8.4956 |
| Chen [27] | 0.9167 | 0.9157 | 6.5503 | 0.9010 | 0.8930 | 4.9870 |
| Ma [28] | 0.9319 | 0.9251 | - | 0.9222 | 0.9179 | - |
| Liu [15] | 0.9430 | 0.9402 | 5.4238 | 0.8417 | 0.8317 | 6.0946 |
| Benoit [81] | 0.8899 | 0.8901 | 7.4786 | 0.7642 | 0.7475 | 7.2806 |
| Chen [29] | 0.9390 | 0.9230 | 6.1580 | 0.8650 | 0.8590 | 6.2630 |
| NR models | | | | | |
| Kim [35] | 0.9410 | 0.9380 | - | - | - | - |
| Chen [37] | 0.8910 | 0.8950 | 7.2470 | 0.8800 | 0.8950 | 5.1020 |
| Yue [40] | 0.9370 | 0.9120 | 5.6620 | 0.9140 | 0.9060 | 4.4490 |
| Liu [41] | 0.9412 | 0.9356 | 5.5483 | 0.9504 | 0.9492 | 3.5260 |
| Oh [51] | 0.9430 | 0.9360 | - | 0.8630 | 0.8710 | - |
| Proposed model | 0.9441 | 0.9426 | 5.3128 | 0.9270 | 0.9240 | 6.1939 |

### TABLE 7  The performance comparison of different distortions on LIVE 3D Phase I

| Criteria | Fezza [14] | Shao [21] | Lin [26] | Chen [27] | Ma [28] | Liu [15] | Benoit [81] | Chen [29] | Kim [35] | Chen [37] | Yue [40] | Liu [41] | Oh [51] | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| JP2K | PLCC | 0.8798 | 0.9238 | 0.8381 | 0.9166 | 0.9344 | 0.9423 | 0.8897 | 0.8260 | 0.9100 | 0.9070 | 0.9340 | 0.9250 | 0.9130 | 0.9450 |
| | SROCC | 0.8783 | 0.8752 | 0.8388 | 0.8954 | 0.8981 | 0.9040 | 0.8701 | 0.8350 | 0.8850 | 0.8630 | 0.8320 | 0.9008 | 0.8850 | 0.9247 |
| JPEG | PLCC | 0.4698 | 0.6563 | 0.6654 | 0.6356 | 0.7046 | 0.7315 | 0.5597 | 0.5210 | 0.7680 | 0.6950 | 0.7440 | 0.7128 | 0.7670 | 0.7456 |
| | SROCC | 0.4963 | 0.6148 | 0.1960 | 0.5632 | 0.6443 | 0.6952 | 0.5189 | 0.5300 | 0.7630 | 0.6170 | 0.5950 | 0.7616 | 0.7620 | 0.7315 |
| WN | PLCC | 0.9422 | 0.9410 | 0.9280 | 0.9353 | 0.9302 | 0.9463 | 0.9360 | 0.9520 | 0.9100 | 0.9170 | 0.9620 | 0.9298 | 0.9100 | 0.9330 |
| | SROCC | 0.9400 | 0.9431 | 0.9284 | 0.9376 | 0.9347 | 0.9468 | 0.9147 | 0.9230 | 0.9220 | 0.9190 | 0.9320 | 0.9264 | 0.9210 | 0.9209 |
| FF | PLCC | 0.5630 | 0.8403 | 0.7086 | 0.7583 | 0.7865 | 0.8658 | 0.7514 | 0.9410 | 0.9570 | 0.7350 | 0.8540 | 0.8574 | 0.9540 | 0.9609 |
| | SROCC | 0.7155 | 0.7814 | 0.6581 | 0.6882 | 0.7391 | 0.8108 | 0.6142 | 0.9070 | 0.9450 | 0.6520 | 0.7790 | 0.8344 | 0.9440 | 0.9530 |
| Blur | PLCC | 0.8823 | 0.9513 | 0.8249 | 0.9418 | 0.9412 | 0.9530 | 0.9256 | 0.9390 | 0.9510 | 0.9170 | 0.9710 | 0.9481 | 0.9500 | 0.9550 |
| | SROCC | 0.9191 | 0.9375 | 0.7910 | 0.9283 | 0.9175 | 0.9294 | 0.8967 | 0.9460 | 0.9320 | 0.8780 | 0.8570 | 0.8912 | 0.9300 | 0.9548 |

In Table 6, for both the datasets, overall performances including 12 existing models and our proposed model are displayed. Also, Tables 7 and 8 have shown their performances for each of distortions including JP2K, JPWG, WN, FF and Blur in different phases. In the tables, the bold value of each performance index means the best result. From the tables, it can be found that conventional methods SSIM and MS-SSIM have lower performances than the proposed model in both Phase I and Phase II, because some binocular features are not taken into account. For eight FR models, Fezza’s method uses RSLM to divide images into different regions. And in Section 1 and Section 4.1 we have introduced the limitation of RSLM. Therefore it cannot have satisfied results due to a poor region division way. For Shao’s method, the strategy to divide different regions may have the similar weakness with Fezza’s and therefore its overall result is not well performed. Lin’s method cannot reach the expected result and that may be caused by the lack of disparity maps. Chen’s method [27] and Benoit’s method [81] have a high dependence on stereo matching. Ma’s method and Liu’s method are based on conventional algorithm SSIM and MS-SSIM. And the unexpected result may be attributed to the difference between 2D and 3D images. Chen’s method [29] has taken some local and global features into account, however, some HVS characteristics may be lost. But it has an excellent performance in White Noise which are the same as Liu’s [15]. In Table 6, for five NR methods, Kim’s [35] and Oh’s method [51] have a better...
performance than other NR ones in Phase I, which can be attributed to the strength of CNN to extract features. For the same reason, Kim’s method has the best performance for JPEG distortion than others. Compared with Yue’s [40] and Liu’s metric [41], the proposed method achieves much better assessment results on LIVE Phase I, which can be ascribed to the sufficient consideration of both global and local features. But Liu’s method has a better performance than the proposed model in Phase II, which may be caused by the individual calculation of different distortions. For Oh’s method [51], due to the lack of binocular features, it cannot gain expected results compared with the proposed metric. However, the main research content is FR models. Apart from NR models, the performance of the proposed model is better than others in Phase I and Phase II (see Table 6). And the performance of the model for JP2K, FF and Blur are all better than other models in Phase I, and for JPEG and WN distortion the performance of the proposed method has a similar result with others. Besides, the performance of the proposed model in Phase II is better than most of other FR models. What is mentioned above can prove the effectiveness of RSDE, PSGE pooling method and the proposed model.

5.6 Time complexity tests

In this subsection, a series of experiments are conducted to analyze the time complexity of the proposed model. In Table 9, the mean calculation time of each 3D image pair in LIVE 3D Phase I of different methods is displayed. From the table, we can find that compared with Kim’s and Lin’s methods, the proposed metric has a lower time complexity, which shows the superiority in running speed. Kim’s method computes the final quality score based on local patches, therefore it needs more computational time. For Lin’s method, the extraction of complicated features leads to higher computational time than the proposed metric. Liu’s model [15] has a lower running time, but its performance is lower than the proposed metric due to insufficient consideration of both monocular and binocular features.

6 CONCLUSION

This paper has presented an FR stereoscopic image quality model considering both monocular and binocular characteristics and local features. The advantages of our work can be shown as follows: (1) The proposed region segmentation method (RSDE) divided the stereo images into OC and NOC regions stably, ensuring a better balance between monocular and binocular vision; (2) the PSGE pooling strategy can well reflect visual contributions of OC and NOC regions; (3) extracted local features provided more local information to assist the visual evaluation. Compared with some existing SIQA models, our model has a better performance than them, which means that the proposed model has higher consistency with subjective scores.

ACKNOWLEDGEMENTS

This work was supported by Key Research & Development Plan of Shandong Province (2019GGX101021), and in part by the China Postdoctoral Science Foundation under Grant Number: 2017M622136.

ORCID

Jianwei Si https://orcid.org/0000-0002-6071-4262
Huan Yang https://orcid.org/0000-0001-5810-0248

REFERENCES

1. Alatan, A., et al.: Scene representation technologies for 3DTV—a survey, IEEE Trans Circuits Syst Video Technol, 17(11), 1587–1605 (2007)
2. Yuan, H., Calic, J., Kondoz, A.: Quality of interaction experience in stereoscopic 3d tv. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 748–752. IEEE (2014)
3. Voo, K., Bong, D.: Quality assessment for stereoscopic images with JPEG compression errors. In: 2015 IEEE International Conference on Consumer Electronics, pp. 220–221 (2015)

4. Wang, J., Wang, S., Wang, Z.: Depth perception of distorted stereoscopic images. In: 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MSPSP), pp. 1–6. IEEE (2015)

5. Wang, J., et al.: Quality prediction of asymmetrically distorted stereoscopic 3D images. IEEE Trans. Image Process. 24(11), 3400–3414 (2015)

6. Chetouani, A.: A fusion-based blind image quality metric for blurred stereoscopic images. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSP), pp. 1–5. IEEE (2017)

7. Chen, Z., Zhou, W., Li, W.: Blind stereoscopic video quality assessment: from depth perception to overall experience. IEEE Trans. Image Process. 27(2), 721–734 (2018)

8. Geng, X., et al.: A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property. Signal Process. Image Commun. 52, 54–63 (2017)

9. Reema, M., et al.: Wavelet based sharp features (wash): an image quality assessment metric based on HVS. In: 2013 2nd International Conference on Advanced Computing, Networking and Security, Mangalore, pp. 79–83. IEEE (2013)

10. Yang, B., Liang, L., Yang, J.: HVS-based structural image quality assessment model. In: 7th World Congress on Intelligent Control and Automation, pp. 8497–8500 (2008)

11. Wang, B., et al.: HVS-based structural similarity for image quality assessment. 2008 9th International Conference on Signal Processing, pp. 1194–1197. IEEE (2008)

12. Liu, A., et al.: Just noticeable difference for images with decomposition model for separating edge and textured regions. IEEE Trans. Circuits Syst. Video Technol. 20(11), 1648–1652 (2010)

13. Zhao, Y., et al.: Binocular just noticeable difference model for stereoscopic images. IEEE Signal Process Lett. 18(1), 19–22 (2011)

14. Fezza, S., Larabi, M., Faraoun, K.: Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 2002–2006. IEEE (2014)

15. Liu, Y., Kong, F., Zhen, Z.: Toward a quality predictor for stereoscopic images via analysis of human binocular visual perception. IEEE Access 7, 69283–69291 (2019)

16. Fan, Y., Larabi, M., Cheikh, F.: Full-reference stereoscopic image quality assessment accounting for binocular combination and disparity information. In: 2017 IEEE International Conference on Image Processing, pp. 760–764 (2017)

17. Sun, G., et al.: Stereoscopic image quality assessment by considering binocular visual mechanisms. IEEE Access 6, 51337–51347 (2018)

18. Li, Y., et al.: No-reference stereoscopic image quality assessment based on visual attention and perception. IEEE Access 7, 46706–46716 (2019)

19. Li, Y., Hu, X.: No-reference stereoscopic image quality assessment using natural scene statistics. In: 2017 2nd International Conference on Multimedia and Image Processing (ICMIP), pp. 123–127. IEEE (2017)

20. Li, X., Wang, Y., Zhao, D.: Joint just noticeable difference model based on depth perception for stereoscopic images. In: 2011 Visual Communications and Image Processing. pp. 1–4. IEEE (2011)

21. Shao, F., et al.: Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. IEEE Trans. Image Process. 22(5), 1940–1953 (2013)

22. Campisi, P., Callet, P., Martini, E.: Stereoscopic images quality assessment. 2007 15th European Signal Processing Conference, Poznan, pp. 2110–2114. IEEE (2007)

23. Yasakhteh, S., et al.: Quality analysis for 3D video using 2D video quality models. IEEE Trans. Consum. Electron. 54(4), 1969–1976 (2008)

24. Kuo, T-Y., et al.: A novel method for stereo image quality assessment. In: 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–4. IEEE (2018)

25. Wang, Z., Simoncelli, E.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. Proc. SPIE 5666, 149–159 (2005)

26. Lin, Y-H., Wu, J-L.: Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. IEEE Trans. Image Process. 23(4), 1527–1542 (2014)

27. Chen, M-J., et al.: Full-reference quality assessment of stereopairs accounting for rivalry. Signal Process. Image Commun. 28(9), 1143–1155 (2013)

28. Ma, J., et al.: SSIM-based binocular perceptual model for quality assessment of stereoscopic images. In: 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, pp. 1–4. IEEE (2017)

29. Chen, L., Zhao, J.: Quality assessment of stereoscopic 3D images based on local and global visual characteristics. In: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 61–66. IEEE (2017)

30. Wang, X., et al.: Quaternion representation based visual saliency for stereoscopic image quality assessment. Signal Process. 145, 202–213 (2018)

31. Wang, Z., Gu, K., Zhao, D.: Reduced reference stereoscopic image quality assessment using sparse representation and natural scene statistics. IEEE Trans. Multimedia 22(8), 2024–2037 (2020)

32. Zhou, W., et al.: Reduced-reference stereoscopic image quality assessment based on view and disparity zero-watermarks. Signal Process. Image Commun. 29(1), 167–176 (2014)

33. Zhang, Y., Phan, T., Chandler, D.: Reduced-reference image quality assessment based on distortion of local perceived sharpness. Signal Process. Image Commun. 55, 130–145 (2017)

34. Zhou, W., et al.: Blind quality estimator for 3D images based on binocular combination and extreme learning machine. Pattern Recognit. 71, 207–217 (2017)

35. Kim, J., et al.: CNN-based blind quality prediction on stereoscopic images via patch to image feature pooling. In: 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, pp. 1745–1749. IEEE (2019)

36. Ding, Y., Li, S., Chang, Y.: Stereoscopic image quality assessment weighted guidance by disparity map using convolutional neural network. In: 2019 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE (2019)

37. Chen, M-J., Cormack, L., Bovik, A.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. 22(9), 3379–3391 (2013)

38. Shen, L., et al.: No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information. IEEE Trans. Emerging Top. Comput. Intell. 3(1), 59–72 (2019)

39. Hachicha, W., Kaaniche, M., Beghdadi, A.: No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models. Signal Process. Image Commun. 54, 107–117 (2017)

40. Yue, G., et al.: Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry. Signal Process. 150, 204–214 (2018)

41. Liu, T-J., et al.: Blind stereoscopic image quality assessment based on hierarchical learning. IEEE Access 7, 8058–8069 (2019)

42. Liu, Y., et al.: Blind stereoscopic image quality assessment accounting for human monocular visual properties and binocular interactions. IEEE Access 8, 33666–33678 (2020)

43. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)

44. Wang, Z., Bovik, A.: A universal image quality index. IEEE Signal Process. Lett. 9(3), 81–84 (2002)

45. Carnec, M., Callet, P., Barba, D.: An image quality assessment method based on perception of structural information. In: Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429), Vol. 3, p. 185 (2003)

46. Wang, Z., Li, Q.: Information content weighting for perceptual image quality assessment. IEEE Trans. Image Process. 20(5), 1185–1198 (2011)

47. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: The Thirtieth Asilomar Conference on Signals, Systems Computers. Vol. 2, pp. 1398–1402 (2003)

48. Benoit, A., et al.: Using disparity for quality assessment of stereoscopic images. In: The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. Vol. 2, pp. 1398–1402 (2003)

49. Fan, Y., et al.: Stereoscopic image quality assessment based on the binocular properties of the human visual system. 2017 IEEE International Conference on Image Processing (ICIP), pp. 325–328. IEEE (2017)
