

Received February 24, 2020, accepted April 27, 2020, date of publication May 6, 2020, date of current version May 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992746

Blind Stereo Image Quality Assessment Based on Binocular Visual Characteristics and Depth Perception

YONG CHEN¹, KAIXIN ZHU¹, AND LIU HUANLIN²

¹Key Laboratory of Industrial Internet of Things and Network Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

²School of Telecommunications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Yong Chen (chenyong@cqupt.edu.cn)

This work was supported by the Chongqing Education Science Committee of China under Grant KJ1400434.

ABSTRACT The quality prediction of stereo images has great challenges without reference images. In this paper, we propose a novel no-reference stereo image quality assessment (NR-SIQA) model based on binocular visual characteristics and depth perception, which can effectively evaluate the quality of symmetric distortion and asymmetric distortion images. To be specific, we discriminate the different binocular behaviors by analyzing binocular visual characteristics, and construct the corresponding cyclopean view instead of single cyclopean view to simulate different binocular behaviors. Then, we extract monocular and binocular visual features from the left view, the right view and the synthetic cyclopean view. Furthermore, in order to evaluate the depth quality of the stereo image accurately, we extract the depth perception features from the weighted disparity map and the longitudinal correlation coefficient map. Finally, we construct the mapping relationship model from quality perception feature domain to quality score domain by training an adaptive enhancement algorithm based on support vector regression (SVR). We evaluate the performance of the proposed algorithm on four stereo image databases. The experimental results show that compared with the state-of-the-art full reference(FR), reduced reference(RR) and NR-SIQA algorithms, the proposed algorithm achieves highly competitive performance for both symmetric and asymmetric distortions.

INDEX TERMS No reference, stereo image quality assessment, binocular visual characteristics, depth perception, longitudinal correlation coefficient map.

I. INTRODUCTION

With the rapid development of 3D technology, 3D movies and television have played an important role in daily life and attracted attention from all over the world. In the process of acquisition, compression, transmission and storage of 3D images, the left and right views may introduce different degrees and types of distortion, which affects the visual quality of experience. There are many image processing methods, such as image denoising [1] and deblurring [2], which can improve the image quality. And image quality assessment (IQA) plays an important role in image processing, because IQA can evaluate whether the method can improve the image quality. Moreover, compared with 2D images, there is certain disparity between the left and right views of a stereopair, which can provide additional depth perception

for the human visual system (HVS) to enable viewers to have realistic experience. But too large disparity also causes an uncomfortable experience for HVS, which also affects human eyes' judgment of image quality. Therefore, in order to provide a good visual experience to people, it is necessary to construct a stereo image quality assessment model consistent with the subjective score. Similarly to 2D-IQA metrics, 3D-IQA metrics can be classified into three categories according to the availability of the pristine reference image: the FR-SIQA methods [3]–[7] which use the original undistorted image as a reference to evaluate the quality of the stereopair; The RR-SIQA methods [8]–[10] which only utilize part of the pristine image when evaluating the quality of the image; and the NR-SIQA methods [11]–[17] which evaluate the quality of the image without any reference information. Compared with the first two categories, NR-SIQA method is more difficult, but it has wider application value in practical life.

The associate editor coordinating the review of this manuscript and approving it for publication was Peng Liu¹.

Different from traditional 2D visual perception, the visual perception quality of a stereopair depends on the qualities of the left and right views. The visual psychology research shows that the HVS can convert different views acquired by the left and right eyes into a single view, but in the process of conversion, it is not a simple addition of left and right eye stimulus to obtain a single view [18]. According to the degree of similarity/dissimilarity between binocular stimulus, different binocular phenomena may occur. Binocular fusion occurs when binocular stimulus are same or very similar [19]; binocular competition occurs when binocular stimulus are sufficiently dissimilar [20]. In [21], a study has showed that the fluctuation amplitude of the primary visual cortex (V1) activity during the binocular competition is 45-83% of the fluctuation range of the V1 activity caused by alternating the binocular stimulus. In other words, local interactions between V1 neurons can trigger sensory alternation during competition. While binocular suppression is a special case of binocular competition. When binocular stimulus are sufficiently dissimilar and important information contained in one view is sufficient to suppress the view acquired by another eye, the view contains more information dominates the perceived content. This phenomenon is called binocular suppression [22].

Most of the previous studies only constructed a single cyclopean image to simulate the final view of the left and right views fused in HVS, which ignores the different effects of different binocular behaviors on HVS perception. In [23], when evaluating the quality of stereoscopic video, the sum channel of binocular vision is constructed in the way of averaging left and right views directly for analyzing binocular visual characteristics. In [24]–[27], the Gabor filter responses of the left and right views are used to weight the left and right eye stimulus and synthesize a single cyclopean to simulate the complex binocular vision mechanism in HVS. In NR stereo image/video quality assessment metrics, although the performance of these algorithms is improved compared with the earlier methods of directly averaging the qualities of left and right views, these methods assume that one view dominates the fusion process of left and right views in HVS, and ignore the phenomenon that two views alternately dominate the fusion process in binocular competition. In addition, some SIQA algorithms [28]–[31] obtain the final qualities of stereo images by the weighted average of the qualities or features of left and right views. For example, in [31], the normalized binocular stimulus intensity is used as the weight of features extracted on the left and right views, and features are combined by weighted sum, and then the deep learning model is used to learn the quality prediction model of stereo images. Although these algorithms can improve the accuracy of stereo image quality prediction, they still cannot fully simulate the complex binocular vision mechanism.

In addition to binocular vision characteristics, another important factor that needs to be considered is depth information. Horizontal parallax is the main factor to provide human eyes with depth information, but excessive parallax can also

cause dizziness and discomfort, which affects the subjective judgment of image quality [32]. The current SIQA algorithms have limited research on depth quality measurement. In [33], Based on the free energy theory and the binocular vision mechanism, Chen *et al.* constructs the suppression map by the phase difference of the luminance channel in the left and right views, and extracts the disparity entropy feature based on autoregressive model to construct the depth quality metric model (DPQM) for stereo video. In [34]–[36], authors extract statistical features from disparity map to predict the depth quality of the stereo image, and consider that the contribution of disparity information to depth quality is same in all regions, which ignores the influence of the longitudinal depth information on the stereo image quality evaluation. The method of [37], Karimi *et al.* extract features from synthesized phase and contrast maps and then use a stacked neural network model to learn the predict model of stereo images. In [38], the prediction performance of the model using stacked auto-encoders (SAE) to predict stereoscopic image quality is accurate, but the complexity is high.

To solve the above problems, we propose a blind stereo image quality evaluation model based on binocular visual characteristics and depth perception. The main contributions of this algorithm are: (1) According to the degree of similarity/dissimilarity between binocular stimulus and the comparison of useful information in the left and right views, we construct the different binocular combination models instead of fixed cyclopean image to simulate different binocular behaviors, such as binocular fusion, binocular competition and binocular suppression. (2) In addition, we extract depth perception features from the weighted disparity map and the longitudinal correlation coefficient map between the left and right views, which considers the influence of horizontal and longitudinal depth information on SIQA.

II. PROPOSED METHOD

The framework of our method is shown in Fig.1. Firstly, we construct different binocular combination models based on binocular vision characteristics to simulate the complex visual mechanism. Then we divide the extraction of quality perception features into two parts: (1) We extract statistical distribution features and texture features from the left and right views as a set of image content quality perception features, and extract statistical distribution features from binocular combination model as another set of image content quality perception features. Then we combine the same features extracted from the left and right views by averaging. (2) We extract depth perception features from the weighted disparity map and the longitudinal correlation coefficient map between the left and right views. Finally, we use SVR to train the regression model from feature domain to quality domain.

A. SYNTHETIC CYCLOPEAN MAP

In [39], several subjective experiments show that when the distortion type of stereo pairs is asymmetric information-loss distortion (ILD), such as asymmetric Gaussian blur distortion,

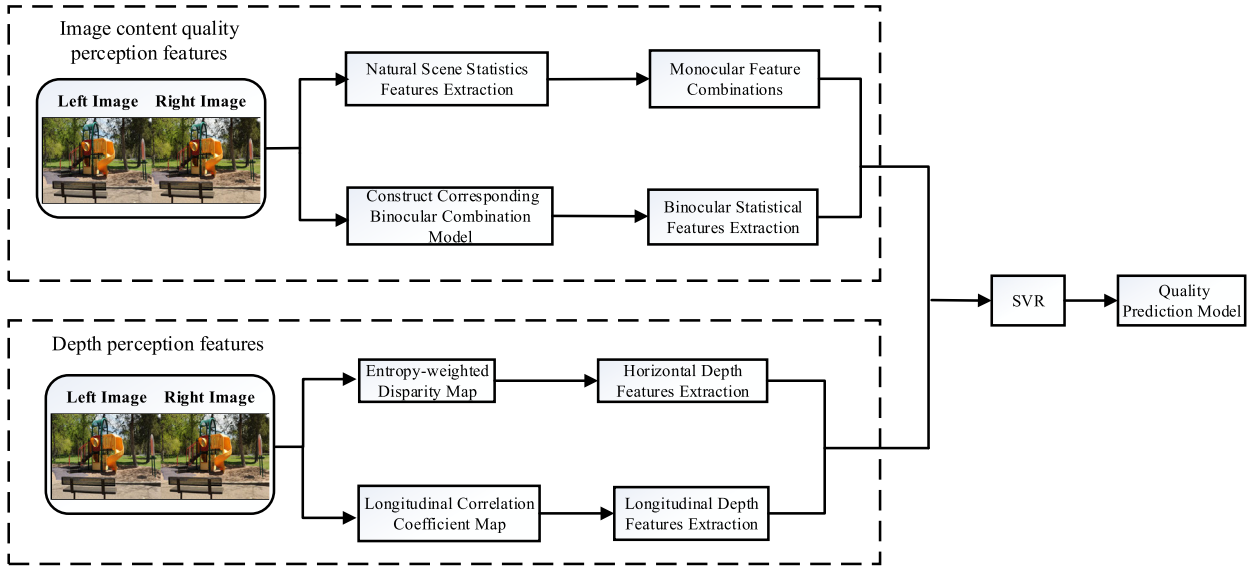


FIGURE 1. The framework of the proposed SIQA model. The monocular and binocular features are extracted from the left, right views and the cyclopean image. The depth perception features are extracted from the weighted disparity map and the longitudinal correlation coefficient map. SVR is used to train a prediction model from features to quality score.

the perceived quality of a stereoscopic image pair is dominated by the high-quality component of stereo pairs because a high-quality component containing sufficient information can complement the lost information of the other view. On the other hand, for information-additive distortion (IAD), such as asymmetric Gaussian noise, the perceptual quality of a stereoscopic image pair is dominated by the low-quality component of stereo pairs since IAD cannot be removed by the low-quality component. As mentioned in [40], the perceptual quality of a stereoscopic image pair is dominated by the view with more information, which is consistent with the psychophysical findings described in [22]. Therefore, in this paper, we select the view that contains more information as the domination view in binocular combination process. Compared with the reference [41], the proposed method does not need to classify distortion types, which reduces the complexity of the algorithm. Meanwhile the proposed method is also applicable to multiply-distorted images which cannot be classified, and reduces the error which may be caused by the classification.

To discriminate and simulate different binocular behaviors in HVS, we construct different cyclopean maps to simulate the binocular combination process. In the calculation process, we use the structural similarity model (SSIM) [42] based on stereo matching algorithm to calculate the similarity between left and right views and select the domination view in binocular combination process by comparing the amount of visual information of the left and right views. In this paper, we measure the amount of visual information by computing the entropy.

$$EL = 1 - \frac{1}{N} \cdot \sum_{k=1}^N \sum_{i=0}^L p(x_{ki,l}) \cdot \log_2\left(\frac{1}{p(x_{ki,l})}\right) \quad (1)$$

where EL is the amount of information contained in the left view, N and L are the number of channels in color space and the maximum pixel value, respectively, and subscripts k and i represent the k th channel in RGB color space and the i th gray scale, $p(x_{ki,l})$ is the probability when pixel value x in left view is equal to i . Similarly, the amount of information contained in the right view can be calculated as (1).

$S_{L,R}$ denotes the similarity value between left and right views, we calculate the similarity value of each pixel between two views based on SSIM [42], and then calculate the structural similarity value $S_{L,R}$ between two views by averaging similarity values of all pixels. In addition, reference [19], [20] and [22] pointed out that different binocular behaviors can be distinguished according to the degree of similarity/dissimilarity between binocular stimulus, so we set structural similarity threshold T_1 and in [39], S. Ryu et al demonstrated the view containing more information dominates the perceived quality of stereo image, so we select the main view by setting the threshold of visual information T_2 . The specific calculation is as follows.

1) BINOCULAR FUSION

When $S_{L,R} \geq T_1$, that indicates that the binocular stimulus are sufficiently similar. We select the left view as the domination view which is often used in previous studies. The binocular fusion model is calculated as follows.

$$CI_f(x, y) = W_L(x, y) \cdot I_L(x, y) + W_R(x+d, y) \cdot I_R(x+d, y) \quad (2)$$

2) BINOCULAR SUPPRESSION

When $S_{L,R} < T_1$ and $EL - ER \geq T_2$, that indicates that the binocular stimulus are sufficiently dissimilar and the left view contains more information. So the left view dominates the

perceptual quality of a stereoscopic image pair, and binocular suppression model is calculated as follows.

$$CI_{SL}(x, y) = W_L(x, y) \cdot I_L(x, y) + W_R(x+d, y) \cdot I_R(x+d, y) \quad (3)$$

When $S_{L,R} < T_1$ and $ER - EL \geq T_2$, that indicates that the binocular stimulus are sufficiently dissimilar and the right view contains more information. So the right view dominates the binocular combination process, and binocular suppression model is calculated as follows.

$$CI_{SR}(x, y) = W_L(x+d, y) \cdot I_L(x+d, y) + W_R(x, y) \cdot I_R(x, y) \quad (4)$$

3) BINOCULAR COMPETITION

When $S_{L,R} < T_1$ and $|EL - ER| < T_2$, that indicates that the binocular stimulus are sufficiently dissimilar and the left or right view does not contain sufficient information to suppress the other view, so the left and right views alternately dominate the fusion process, rather than simply taking a view as the domination view. Based on above analysis, we recombine the two suppression maps whose domination views are the left and right views respectively to simulate this alternating phenomenon. The binocular competition model is calculated as follows.

$$CI_R(x, y) = W_{SL}(x, y) \cdot CI_{SL}(x, y) + W_{SR}(x+d, y) \cdot CI_{SR}(x+d, y) \quad (5)$$

Weight W_L, W_R, W_{SL}, W_{SR} are calculated as follows.

$$W_L(x, y) = \frac{E_L(x, y)}{E_L(x, y) + E_R(x+d, y)} \quad (6)$$

$$W_R(x+d, y) = \frac{E_R(x+d, y)}{E_L(x, y) + E_R(x+d, y)} \quad (7)$$

$$W_{SL}(x, y) = \frac{E_{SL}(x, y)}{E_{SL}(x, y) + E_{SR}(x+d, y)} \quad (8)$$

$$W_{SR}(x, y) = \frac{E_{SR}(x+d, y)}{E_{SL}(x, y) + E_{SR}(x+d, y)} \quad (9)$$

where $I_L(x, y)$ and $I_R(x, y)$ are the left and right views, respectively, $d(x, y)$ is the horizontal disparity value in the (x, y) location. In this paper, the disparity value is calculated by SSIM-based algorithm utilized in [6]. $E_L(x, y)$ and $E_R(x, y)$ are 2D Gabor energy responses in all scales and directions of the left and right views, respectively, $E_{SL}(x, y)$ and $E_{SR}(x, y)$ are 2D Gabor energy responses in all scales and directions of the binocular suppression maps with the left or right view as the domination view respectively, the specific calculation is as follows.

2D-Gabor filter is defined as follows.

$$g(x, y, \lambda, \theta) = \frac{1}{2\pi\sigma} \exp\left[-\frac{1}{2} \left(\frac{x^2 + y^2}{\sigma^2}\right)\right] \times \exp\left[i \left(\frac{2\pi x}{\lambda}\right)\right] \quad (10)$$

where $x = x_0 \cos \theta + y_0 \sin \theta$ and $y = -x_0 \sin \theta + y_0 \cos \theta$, (x_0, y_0) is center of the filter, λ is the wavelength, which

controls the scale of the Gabor filter, θ is the orientation, σ is the standard deviation of an elliptical Gaussian envelope along x and y axes. Since the simple and complex cells in primary visual cortex have receptive fields at different scales, we utilize a multi-scale set of Gabor filter banks with five scales of frequency domain and four orientations, $\sigma = 0.5\lambda$, $\lambda \in (1, \sqrt{2}, 2\sqrt{2}, 4)$, $\theta \in (1, \pi/4, \pi/2, 3\pi/4)$.

2D-Gabor energy response is defined as follows.

$$GE_{\theta,\lambda}(x, y) = I(x, y) \otimes g(x, y, \lambda, \theta) \quad (11)$$

where \otimes represents a convolution operation, $I(x, y)$ is input image.

B. IMAGE CONTENT QUALITY PERCEPTION FEATURES

Although distinctive in contents, natural images inherently obey a particular statistical characteristic [25], which is measurably modified by the presence of distortions. Motivated by this, we extract the brightness statistical distribution features from the left and right views and cyclopean map calculated according to section 2.1.

Firstly, we use the mean subtracted contrast normalized method (MSCN) to preprocess images. Fig. 2(a)–(c) give the left and right views from LIVE 3D Phase I database [43] and the corresponding synthetic cyclopean view. The statistical distribution histograms of the MSCN coefficients of these images are more consistent with the generalized Gaussian distribution (GGD), as shown in Fig. 2(d)–(f). Therefore, in this paper, we use the GGD model to fit these statistical distributions. Given an image I with the size of $M \times N$, its MSCN coefficients are calculated by:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + 1}, \quad i \in \{1, 2, \dots, M\}, \quad j \in \{1, 2, \dots, N\} \quad (12)$$

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L \omega_{k,l} I_{k,l}(i, j) \quad (13)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L \omega_{k,l} (I_{k,l}(i, j) - \mu(i, j))^2} \quad (14)$$

where $I(i, j)$ denotes the pixel value in the (i, j) spatial location; $\omega = \{\omega_{k,l} | k = -K, \dots, K, l = -L, \dots, L\}$ is a 2D circular symmetric gaussian weighting function. K and L determine the size of 2D Gaussian kernel ($K = L = 3$).

A zero mean GGD is calculated by:

$$f(x, \alpha, v^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-(|x|/\beta)^2\right) \quad (15)$$

$$\beta = v\sqrt{\Gamma(1/\alpha) / \Gamma(3/\alpha)} \quad (16)$$

with gamma function $\Gamma(\cdot)$ defined as:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \alpha > 0 \quad (17)$$

where α and v^2 reflecting the image naturalness, control the shape and variance of distribution, respectively. x is the

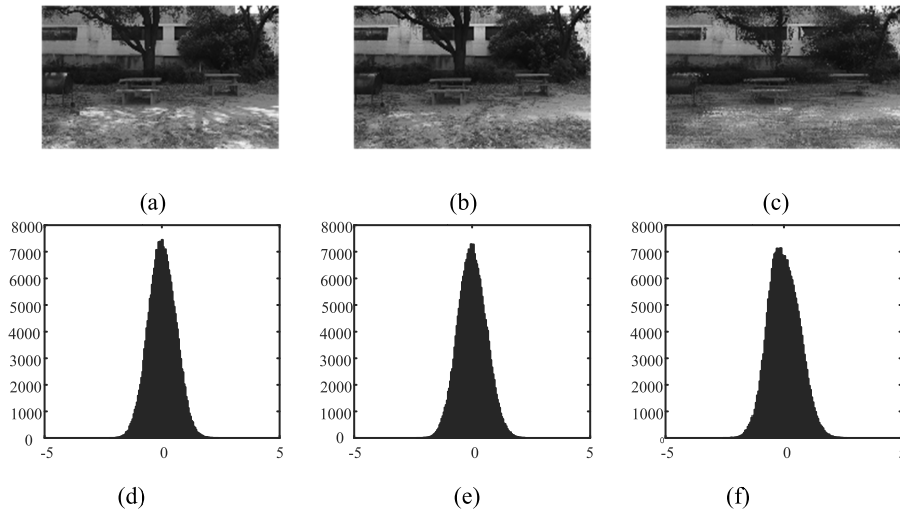


FIGURE 2. (a)–(c) are left and right views and corresponding synthetic cyclopean view. (d)–(f) are statistical distributions of MSCN coefficients corresponding to (a)–(c).

MSCN coefficient. In this paper, we use the GGD model to fit MSCN distributions of left and right views and the corresponding cyclopean view, and the parameters α and ν^2 are regarded as the quality-sensitive features $f_1 = \{\alpha, \nu^2\}$.

In addition to extracting statistical distribution features in spatial domain, we also extract texture features from left and right views as another set of image content quality perception features.

Local binary pattern (LBP) is an efficient local texture descriptor, which has the significant advantages of small computation, no training and light invariance. In this paper, we use rotation-invariant and uniform LBP ($LBP_{r,p}^{riu2}$) to extract texture features from left and right views, respectively.

$$LBP_{r,p}^{riu2} = \begin{cases} \sum_{n=0}^{p-1} s(x_{r,p,n} - x_c), & U(LBP_{r,p}) \leq 2 \\ p + 1, & otherwise \end{cases} \quad (18)$$

where x_c is the gray value of the center pixel of the local neighborhood, $x_{r,p,n}$ is the gray values of p equally spaced pixels on a circle of radius $r(r > 0)$ that form a circularly symmetric neighbor set, $s(x)$ is a symbolic function, $U \leq 2$ represents the two binary values that are adjacent to each other on a the circularly symmetric neighbor set change by 0/1(1/0) no more than twice, which is in the form of

$$U(LBP_{r,p}) = |s(x_{r,p,p-1} - x_c) - s(x_0 - x_c)| + \sum_{n=1}^{p-1} |s(x_{r,p,n} - x_c) - s(x_{n-1} - x_c)| \quad (19)$$

In this paper, we choose $p = 8$ and $r = 1$. And then we use the magnitude of the MSCN coefficient to weight the LBP

histogram to obtain 10-dimensional texture features f_2 .

$$f_2(k) = \sum_{j=1}^N w_j \delta(LBP_{r,p}(j), k) \quad (20)$$

$$\delta(x, y) = \begin{cases} 1, & x = y \\ 0, & otherwise \end{cases} \quad (21)$$

where N is the number of pixels, $k \in [0, p + 1]$ is LBP pattern, w_j is the magnitude of the MSCN coefficient.

C. DEPTH PERCEPTION FEATURES

Binocular disparity plays an important role in the depth perception of stereoscopic images. When human view stereoscopic images/video content, disparity provides depth perception for human eyes and improves the quality of the experience. However, excessive disparity tends to give viewers phenomena including eye strain, dizziness, reduced visual sensitivity, etc. Therefore, it is necessary to statistically analyze the disparity information of the stereo image pairs. However, not all regions of disparity information have the same impact on image quality assessment. As mentioned in [39], the view with more visual information will attract more attention of HVS. Therefore, based on the characteristics of human visual perception, we use the amount of visual information to weight the disparity map, so as to emphasize the influence of disparity value in the attention region of HVS on the estimation of depth quality.

Firstly, we use the visual information of dominant view to weight the disparity map to obtain a weighted disparity map D_e . Then since the existence of distortion will change the statistical characteristics of the disparity map, we extract some simple statistical features such as kurtosis and skewness from the weighted disparity map as depth

features $f_3 = \{S, K\}$.

$$D_e(i, j) = w_e(i, j) \times d(i, j) \quad (22)$$

$$S = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (D_e(i, j) - m)^3 / v^{\frac{3}{2}} \quad (23)$$

$$K = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (D_e(i, j) - m)^4 / v^2 \quad (24)$$

where $w_e(i, j)$ is the weight of the disparity at (i, j) , we apply equation (1) to a 7×7 image block centered on (i, j) , and calculate the amount of visual information of the image block as the weight value $w_e(i, j)$, $d(i, j)$ is the disparity at (i, j) which is calculated by SSIM-based algorithm utilized in [6], S, K, m and v are skewness, kurtosis, mean and variance of the weighted disparity map, respectively, $M \times N$ is the size of weighted disparity map.

In addition, although horizontal disparity is the main reason for generating depth information, it reflects the horizontal position difference between the left and right retinal projections of a given point in space, thereby ignoring the influence of longitudinal differences between the left and right views on the depth quality assessment. And the longitudinal mismatch information between left and right views caused by asymmetric distortion will also affect the depth perception, and cause binocular visual discomfort, which will affect the human eye's judgment on depth quality [44]. Therefore, in this paper, we calculate the degree of linear correlation of longitudinal changes between left and right views, as a supplement to horizontal depth information.

Firstly, the disparity compensation maps of the left and right views are calculated as follows.

$$I_{dl}(i, j) = I_l(i, j) + d_r(i, j) \quad (25)$$

$$I_{dr}(i, j) = I_r(i, j) + d_l(i, j) \quad (26)$$

where $d_r(x, y)$ corresponds to the disparity values which allows to find for each pixel in the left view its the maximum similarity one in the right, $d_l(x, y)$ corresponds to the disparity values which allows to find for each pixel in the right view its the maximum similarity one in the left.

Then, the mean subtracted contrast normalized operation is performed on the left and right views and their corresponding disparity compensation maps. Finally, the longitudinal correlation coefficient map between the left view and the right disparity compensation map is calculated as follows.

$$C = corr \{N [I_l(i, j)], \{N [I_{dl}(i, j)]\} \quad (27)$$

where $corr(X, Y)$ is a correlation function provided by MATLAB. $corr(X, Y)$ calculates the pairwise correlation coefficient between each pair of columns in the matrices X and Y , $N(\cdot)$ is the mean subtracted contrast normalized operation.

Fig. 3(a)-(d) show the left and right views from LIVE Phase II and the statistical distributions of the two longitudinal correlation coefficient maps calculated by the above method. We can see from Fig. 3(c)(d), the statistical distributions of

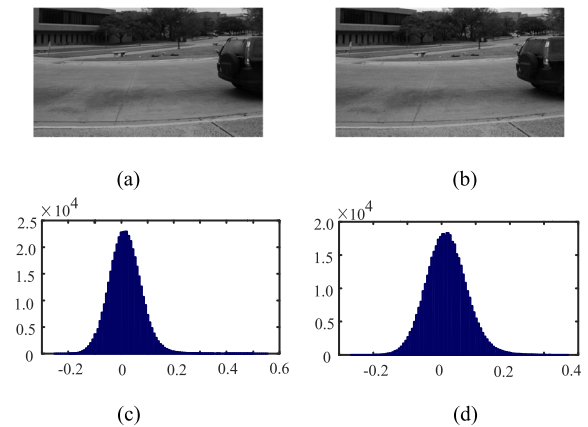


FIGURE 3. (a)-(b) are left and right views from LIVE Phase II database, (c) statistical distributions of the longitudinal correlation coefficient between the left view and right disparity compensation map, (d) longitudinal correlation coefficient between the right view and left disparity compensation map.

the two longitudinal correlation coefficient maps are consistent with the non-zero mean asymmetric generalized gaussian distribution AGGD.

Therefore, this paper uses the non-zero mean AGGD model to fit the longitudinal correlation coefficient map between the left view and right disparity compensation map. The non-zero mean AGGD model are calculated as follows.

$$f(x; \alpha, \beta_l, \beta_r, \mu)$$

$$= \begin{cases} \frac{\alpha}{(\beta_l + \beta_r)\Gamma(1/\alpha)} \exp\left(-\left(\frac{x - \mu}{\beta_l}\right)^\alpha\right), & x < \mu \\ \frac{\alpha}{(\beta_l + \beta_r)\Gamma(1/\alpha)} \exp\left(-\left(\frac{x - \mu}{\beta_r}\right)^\alpha\right), & x > \mu \end{cases} \quad (28)$$

$$\beta_l = \sigma_l \sqrt{\Gamma(1/\alpha) / \Gamma(3/\alpha)} \quad (29)$$

$$\beta_r = \sigma_r \sqrt{\Gamma(1/\alpha) / \Gamma(3/\alpha)} \quad (30)$$

where α is the shape parameter and controls the shape of AGGD distribution, β_l, β_r are the distribution variances of left and right respectively. μ is the mean value. When $\mu = 0$, it indicates that the distribution is zero mean AGGD.

Fig. 4(a) shows the statistical distributions of the longitudinal correlation coefficient maps between the left views and right disparity compensation maps. These stereo image pairs are from the LIVE Phase I. Fig. 4(b) shows the statistical distributions of the longitudinal correlation coefficient maps of three asymmetric distortion images in the LIVE Phase II. It can be seen from Fig. 4 that the correlation coefficient distributions corresponding to images with different distortion types have different statistical characteristics and are consistent with AGGD. Therefore, we extract the AGGD parameters of the longitudinal correlation coefficient map between the left view and right disparity compensation map as depth perception features $f_4 = \{\alpha, \beta_l, \beta_r, \mu\}$.

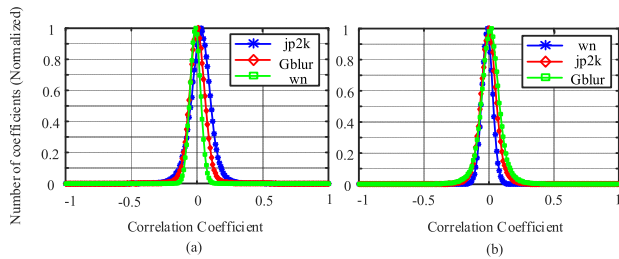


FIGURE 4. Statistical distribution of longitudinal correlation coefficients of stereo pairs with different distortion types.

D. QUALITY ESTIMATION

In this paper, we extract 44-dimensional features from each stereo image pairs. Among them, we extract $2 \times 3 = 6$ dimensions statistical distribution features f_1 and $10 \times 3 = 30$ dimensions texture features f_2 from the left and right views on three scales, respectively. In order to reduce the feature dimension and simulate the interaction between simple cells in the visual cortex, we combine the same features extracted from the left and right views by averaging and obtain 36-dimensional feature combinations; Then we extract 2-dimensional binocular statistical features f_1 from the cyclopean map, 2-dimensional depth features f_3 and 4-dimensional depth features f_4 .

After obtaining quality-sensitive features, we use support vector regression (SVR) to construct the mapping relationship from feature domain to quality domain. SVR has been proved to be an effective tool for solving prediction or non-linear fitting problems. During training stage, we first extract quality-sensitive features of training database and then use the SVR to learn a prediction function from training feature vector to subjective ratings (e.g., MOS or DMOS). During testing stage, we first extract the feature vector and then predict image quality scores of testing images by feeding feature vector into the well trained prediction function.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In order to test the performance of the proposed algorithm, our method is compared with some state-of-the-art NR-SIQA and FR-SIQA methods on four publicly available S3D IQA databases: LIVE 3D Phase I database [43], LIVE 3D Phase II database [6], Waterloo-IVC 3D Phase I database [45], Waterloo-IVC 3D Phase II database. Table 1 gives the specific parameters of each database.

A. OVERALL PERFORMANCE COMPARISON

In order to verify the effectiveness of the SIQA model proposed in this paper, the proposed method is compared with some state-of-the-art SIQA methods, including FR-SIQA metrics (Lin et al. [4], Khan and Channappayya [5], Chen et al. [6], SSIM [42], Jiang et al. [46]), RR-SIQA metrics (Ma et al. [8]) and NR-SIQA metrics (SINQ [15], Yang et al. [31], Zhou et al. [32], Karimi et al. [37], Yang et al. [38], Fezza et al. [41], BRISQUE [47]).

TABLE 1. The specific parameters of each database.

Database	Distortion Types	Number of Reference Images	Number of Distorted Images
LIVE Phase I	JP2K, JPEG, WN, FF, Blur	20	365
LIVE Phase II	JP2K, JPEG, WN, FF, Blur	8	360
IVC Phase I	JPEG, WN, Blur	6	330
IVC Phase II	JPEG, WN, Blur	10	460

In this paper, we utilize three commonly used criteria to quantify and verify the performance of the proposed method as well as all competing metrics. They are Spearman rank-order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), and Root mean square error (RMSE). Among them, SROCC and PLCC evaluate prediction monotonicity and consistency of prediction performance, respectively. Whereas, RMSE reflects the prediction error. For a perfect metric, the match value between the predicted scores and associated subjective ratings is close to 1 for PLCC and SRCC but close to 0 for RMSE. In this paper, the dataset requires to be randomly split into two non-overlapping subsets: training subset and testing subset. This study takes 80% dataset as training subset, while the resting is regarded as testing subset. After training the model on the training subset, the prediction performance is measured on the test subset. To avoid the performance bias, the random training-testing split is repeated 1000 times and the performance is reported in form of median value.

Tables 2-3 show the overall performance of our method compared with other SIQA methods on the four publicly available S3D IQA databases. Top three performance is highlighted in boldface for readers' convenience. Compared with the comparison algorithms, the overall performance of our method is ranked in the top three on the four S3D IQA databases and is close to the best performance. The SROCC, PLCC and RMSE values of the proposed algorithm on the four S3D IQA databases are significantly better than most of the comparison algorithms, and some values achieve the most superior results on LIVE 3D Phase I database and IVC 3D Phase II database. It is proved that the proposed method can effectively evaluate the quality of symmetric and asymmetric distortion stereo images. In particular, it can be seen from table 3 that the proposed algorithm can be effectively used to evaluate complex asymmetric distortions with different types.

B. PERFORMANCE ON INDIVIDUAL TYPES

An excellent IQA metric should not only show its powerful ability on entire database, but also effectively cope with quality assessment task on each individual distortion type. In this section, an experiment is further conducted to investigate and verify the superiority of the proposed method on estimating the perceived quality of individual distortion

TABLE 2. Overall performance comparison on LIVE Phase I and Phase II.

Database		LIVE Phase I			LIVE Phase II		
Method	Type	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Lin[4]	FR	0.9366	0.9314	5.7440	0.9113	0.8935	4.6477
Khan[5]	FR	0.9272	0.9163	-	0.9323	0.9222	-
Chen FR[6]	FR	0.8337	0.9157	6.2681	0.7701	0.8883	4.8922
Jiang[46]	FR	0.9459	0.9336	5.2763	0.9162	0.9030	4.5233
Ma[8]	RR	0.9301	0.9292	6.0241	0.9213	0.9175	4.3903
SINQ[15]	NR	0.9556	0.9359	4.7814	0.9367	0.9312	3.9595
Yang[31]	NR	0.9364	0.9289	5.2575	0.9131	0.8747	5.0994
Zhou[32]	NR	0.9410	0.9210	5.5400	0.9230	0.9190	4.2620
Karimi[37]	NR	0.9560	0.9400	4.9980	0.9230	0.9130	4.4360
Yang[38]	NR	0.9613	0.9497	-	0.9386	0.9284	-
Fezza[41]	NR	-	-	-	0.9250	0.9080	3.0180
BRISQUE[47]	NR	0.9100	0.9010	6.7930	0.7820	0.7700	7.0380
Proposed	NR	0.9595	0.9430	4.8380	0.9365	0.9228	3.6673

TABLE 3. Overall performance comparison on IVC Phase I and IVC Phase II.

Database		IVC Phase I			IVC Phase II		
Method	Type	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Khan[5]	FR	0.9304	0.9253	-	0.9097	0.9053	-
Chen FR[6]	FR	0.4575	0.4147	11.9912	0.3407	0.3302	16.2554
SSIM[42]	FR	0.5427	0.5372	11.1710	0.4080	0.4029	16.3354
SINQ[15]	NR	0.9561	0.9366	4.9602	0.9223	0.9117	7.3766
Karimi[37]	NR	0.9050	0.8840	-	0.8250	0.7890	-
Yang[38]	NR	0.9407	0.9114	-	0.8991	0.8666	-
Fezza[41]	NR	0.9040	0.8980	-	0.8900	0.8660	-
Proposed	NR	0.9310	0.9235	5.9890	0.9252	0.9226	7.1192

type. Given that Waterloo-IVC databases are more complex and contain mixed distortion types, therefore, they are not suitable for this task. Tables 4-6 tabulate the experimental results on LIVE Phase I and Phase II. Similarly, the best result for SIQA comparisons is also highlighted in bold. It can be seen from tables 4-6, although the algorithm proposed in this paper do not work well on all individual distortions, it shows the best performance on some distortion types and the top three on others. Moreover, compared with FR metrics, the proposed method shows powerful competitiveness with FR metrics. To this end, these observations verify that the proposed method can be fully competent to tackle quality assessment problem of S3D images.

C. PARAMETER SETTING

Section 2.1 analyzes binocular visual characteristics, and discriminates and simulates the binocular combination models of stereo image pairs in HVS by setting thresholds. In order to select appropriate thresholds, this paper analyzes the impact

of similarity threshold T_1 and amount of information threshold T_2 on the performance of the proposed model, where $T_1 \in \{0.5, 0.6\}$, $T_2 \in \{0.1, 0.2, 0.3\}$. It can be seen from table 7, when $T_1 = 0.5$, $T_2 = 0.3$, the PLCC and SROCC values of the proposed algorithm on the LIVE Phase I database are the maximum value; The performance on the LIVE Phase II database is not the best, but it is also the top three, and the gap with the best performance is small. In addition, when the value of T_1 , T_2 are increased, the performance of our model does not change much. Therefore, we set $T_1 = 0.5$, $T_2 = 0.3$.

D. CONTRIBUTE OF DIFFERENT FEATURE COMBINATIONS

According to different binocular behaviors, the algorithm proposed in this paper constructs the corresponding cyclopean image by thresholds judgment, and then extracts the quality perception features on the cyclopean image and the weighted disparity map to realize the quality prediction of stereo image. In order to analyze the effect of different feature

TABLE 4. Performance comparison of the proposed model and other metrics on each individual distortion type in terms of PLCC.

Database		LIVE Phase I					LIVE Phase II				
Method	Type	WN	JP2K	JPEG	GBLur	FF	WN	JP2K	JPEG	GBLur	FF
Lin[4]	FR	0.9266	0.9520	0.7546	0.9583	0.8620	-	-	-	-	-
Jiang[46]	FR	0.9503	0.9412	0.7074	0.9681	0.8677	0.9174	0.9419	0.9242	0.9204	0.9194
Ma[8]	RR	0.9347	0.9399	0.7200	0.9356	0.8427	0.9325	0.8795	0.7645	0.9131	0.9063
SINQ[15]	NR	0.9700	0.9480	0.7340	0.9670	0.8980	0.9700	0.9200	0.8710	0.9840	0.9380
Karimi[37]	NR	0.9550	0.9390	0.7710	0.9590	0.8820	0.9660	0.8970	0.8660	0.9570	0.9180
Fezza[41]	NR	-	-	-	-	-	0.9530	0.9360	0.9050	0.9740	0.9570
Proposed	NR	0.9584	0.9549	0.7878	0.9693	0.8619	0.9871	0.9338	0.8754	0.9884	0.9055

TABLE 5. Performance comparison of the proposed model and other metrics on each individual distortion type in terms of SROCC.

Database		LIVE Phase I					LIVE Phase II				
Method	Type	WN	JP2K	JPEG	GBLur	FF	WN	JP2K	JPEG	GBLur	FF
Lin[4]	FR	0.9289	0.9127	0.7164	0.9332	0.8286	-	-	-	-	-
Jiang[46]	FR	0.9246	0.8842	0.6092	0.9221	0.7957	0.8725	0.9093	0.8888	0.8645	0.8865
Ma[8]	RR	0.9283	0.9073	0.6595	0.9207	0.7919	0.9541	0.8680	0.7912	0.9229	0.9438
SINQ[15]	NR	0.9540	0.8970	0.6560	0.8830	0.8410	0.9570	0.9090	0.8390	0.9090	0.9240
Zhou[32]	NR	0.9410	0.8710	0.6870	0.9050	0.7850	0.9420	0.8970	0.8020	0.9070	0.9200
Karimi[37]	NR	0.9450	0.9107	0.7500	0.9190	0.8370	0.9530	0.8750	0.8320	0.8740	0.9070
Fezza[41]	NR	-	-	-	-	-	0.9470	0.9270	0.8860	0.9280	0.9520
Proposed	NR	0.9477	0.9130	0.6872	0.9174	0.8586	0.9545	0.8917	0.8416	0.9319	0.8948

TABLE 6. Performance comparison of the proposed model and other metrics on each individual distortion type in terms of RMSE.

Database		LIVE Phase I					LIVE Phase II				
Method	Type	WN	JP2K	JPEG	GBLur	FF	WN	JP2K	JPEG	GBLur	FF
Lin[4]	FR	6.2569	3.9631	4.2912	4.1371	6.2993	-	-	-	-	-
Jiang[46]	FR	-	-	-	-	-	4.1840	4.0210	3.5180	3.7680	4.6810
SINQ[15]	NR	3.8340	3.9850	4.0490	3.5540	5.0180	2.5190	3.4630	3.4760	2.4810	3.8030
Karimi[37]	NR	5.0170	4.6440	4.2900	4.4580	5.9970	2.9360	5.0830	4.0710	4.5810	4.9740
Fezza[41]	NR	-	-	-	-	-	3.2400	3.4430	3.4260	3.1370	3.3140
Proposed	NR	5.4754	3.9200	4.1697	4.0770	5.3135	3.2958	3.9596	2.9195	2.5033	4.7659

combinations on the performance, table 8 shows the performance comparison of these algorithms with different feature combinations. In table 8, “Mon” represents the model that extracts shape parameters of GGD and texture features from the left and right views as monocular features; “Mon + Bin” represents the model that extracts monocular and binocular features from the left, right views and corresponding cyclopean map; “Mon + EDM” represents the model that extracts monocular and horizontal depth features; “Bin + EDM” represents the model that extracts binocular and horizontal depth features; “EDM” represents the model that extracts horizontal depth features from the weighted disparity map in addition to monocular and binocular features. In table 8, the performance comparison between “Mon” and “Mon + Bin”,

and the performance comparison between “Mon + EDM” and “Bin + EDM” prove that the synthetic cyclopean image proposed in this paper can improve the performance of SIQA model and can be used to simulate the fusion process. The performance comparison between “Mon” and “Mon + EDM” proves the effectiveness of depth features. And the overall performance of the algorithm has been improved by extracting binocular features and depth features which confirms the effectiveness of the features extracted by our method.

E. GENERALIZATION PERFORMANCE

In order to verify the generalization performance of the algorithm in this paper, a cross-validation experiment

TABLE 7. The influence of different thresholds on the overall performance of the model.

Database		LIVE Phase I			LIVE Phase II		
T_1	T_2	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
0.5	0.1	0.9505	0.9417	4.9918	0.9295	0.9187	3.9852
	0.2	0.9573	0.9414	4.9875	0.9301	0.9200	3.5892
	0.3	0.9595	0.9430	4.8380	0.9365	0.9228	3.6673
0.6	0.1	0.9478	0.9340	5.0605	0.9272	0.9154	3.6395
	0.2	0.9480	0.9350	4.5916	0.9265	0.9096	3.6503
	0.3	0.9484	0.9349	4.9954	0.9280	0.9065	3.6007

TABLE 8. Performance comparison of models with different feature combinations on LIVE Phase I and LIVE Phase II.

Databases		LIVE Phase I			LIVE Phase II		
Methods	PLCC	SROC C	RMSE	PLCC	SROCC	RMSE	
Mon	0.8690	0.8508	8.7012	0.7866	0.7643	6.3102	
Mon+Bin	0.9351	0.9288	6.9057	0.9117	0.8966	5.0943	
Mon+EDM	0.9283	0.9136	5.7063	0.8907	0.8637	7.3054	
Bin+EDM	0.9196	0.9018	6.2351	0.9189	0.9008	4.3257	
EDM	0.9504	0.9409	5.0230	0.9305	0.9147	3.8632	
Proposed	0.9595	0.9430	4.8380	0.9365	0.9228	3.6673	

TABLE 9. Performance of model training on LIVE Phase I and testing on LIVE Phase II.

	JP2K	JPEG	WN	GB	FF	ALL
PLCC	0.8432	0.6210	0.6937	0.9085	0.7995	0.8124
SROCC	0.8090	0.6626	0.6402	0.8929	0.8207	0.7938
RMSE	5.2191	5.2567	6.6231	5.8189	6.2231	6.9216

is performed. The LIVE 3D Phase I and Phase II databases are used as training sets and test sets, respectively. The proposed model is trained on LIVE Phase I and tested on LIVE Phase II. Meanwhile, we train the model on the LIVE Phase II and test it on LIVE Phase I. The results are listed in Tables 9 and 10. Because the IVC Phase I and Phase II databases contain mixed distortion images with different distortion types, table 11 does not show cross-database testing for individual distortion. It can be seen that the performance of training on one database (i.e., Phase I) and testing on another database (i.e., Phase II) is not good because the distortion images contained in the training set and the test set differ greatly in image content, distortion type and distortion degree. Comparatively speaking, the prediction model trained by the algorithm in this paper has good adaptability to different test databases.

F. PERFORMANCE ON SYMMETRIC AND ASYMMETRIC DISTORTIONS

In order to further test the performance of our algorithm, we divided the LIVE Phase II database into two separate

TABLE 10. Performance of model training on LIVE Phase II and testing on LIVE Phase I

	JP2K	JPEG	WN	GB	FF	ALL
PLCC	0.9069	0.5821	0.7559	0.8302	0.8046	0.8277
SROCC	0.8915	0.5269	0.8506	0.8213	0.7571	0.8065
RMSE	5.7485	5.0570	9.8280	7.0693	8.8167	9.0943

TABLE 11. Performance of model on different distortions across databases(training database /testing database).

	LIVE I/ LIVE II	LIVE II/ LIVE I	IVC I/ IVC II	IVC II/ IVC I
PLCC	0.8124	0.8277	0.8063	0.8469
SROCC	0.7938	0.8065	0.7968	0.8302

TABLE 12. Performance on symmetric and asymmetric distortions on LIVE Phase II.

	PLCC		SROCC	
	Sym	Asym	Sym	Asym
Chen FR [6]	0.8134	0.7701	0.9173	0.8883
SINQ[15]	0.9314	0.9171	0.9287	0.9069
SSIM[42]	0.8490	0.7170	0.8430	0.6770
Fezza [41]	0.9350	0.9090	0.9280	0.8820
Proposed	0.9409	0.9328	0.9221	0.9268

TABLE 13. Computational time of evaluating a stereopair in LIVE Phase I (s).

SSIM [42]	Chen FR [6]	SINQ [15]	Karimi [37]	Yang [38]	Proposed
0.107	18.756	6.569	9.511	19.882	8.148

subsets of only symmetric and asymmetric distortions and tested the performance of our method on both subsets. The results are listed in Table 12. It is difficult to simulate the fusion process happening in the HVS, because of the complex visual mechanisms. So most IQA models perform significantly worse on asymmetric distortions than on symmetric ones, but our model performs well on both.symmetric and asymmetric distortions.

G. COMPUTATIONAL COMPLEXITY

In order to evaluate the time complexity, we calculate the time it takes for our method to evaluate a stereopair in LIVE Phase I. The complexity of SSIM is very low, because it does not need to synthesize cyclopean map, and the accuracy is not as high as ours. It can be seen from Table 13, compared with reference [6], [37], [38], the complexity of our algorithm is the lowest, because our method extracts spatial domain features directly, which have lower computational complexity than the transform domain features. Moreover, we use our method to evaluate image quality without classifying the distortion type of images. The models in [37] and [38] use CNN and stack auto-encoders to predict image quality, although the overall performance of the models is competitive with ours, the complexity of this paper is lower than theirs. Compared

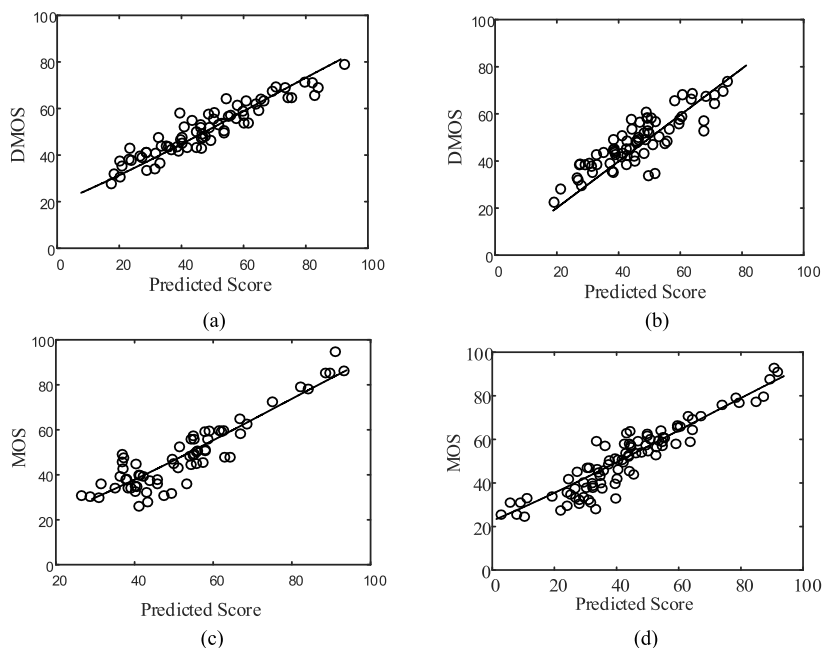


FIGURE 5. Scatter plots of objective scores versus subjective data (DMOS). (a-b) LIVE Phase I and Phase II. (c-d) IVC Phase I and Phase II.

with [15], the performance of our algorithm is better than it on LIVE Phase I and IVC Phase II and is competitive with it on other image databases, but the complexity is higher than it. In the future work, we should further reduce the algorithm complexity.

H. ALGORITHM CONSISTENCY

The scatter map of quality prediction value and subjective score of the proposed algorithm is shown in figure 5. It can be seen that the scatter distribution of the algorithm in this paper is closely clustered, indicating that the prediction results of the algorithm in this paper are highly consistent with human subjective evaluation.

IV. CONCLUSION

In this paper, we propose a no-reference stereo image quality assessment model based on binocular visual characteristics and depth perception. More precisely, we first discriminate different binocular behaviors based on thresholds and construct corresponding cyclopean images to simulate complex binocular visual mechanism. Then we extract image content quality-aware features from the left and right views and the corresponding cyclopean map and then extract the depth-aware features from the weighted disparity map and the longitudinal correlation coefficient map. Finally, we use SVR to construct a stereo image quality evaluation model from feature domain to quality fraction domain. Experimental results show the effectiveness of the proposed 3D quality assessment technique compared to the recent state-of-the-art methods.

REFERENCES

- [1] J. Lu, X. Luo, and S. Qi, "Improved non-local means image denoising algorithm using visual features based on fuzzy metric," *J. Chongqing Univ. Posts Telecommun., Natural Sci. Ed.*, vol. 30, no. 3, pp. 408–415, 2018.
- [2] X. Xu, H. Liu, and Y. Li, "Image deblurring with blur kernel estimation in RGB channels," *J. Chongqing Univ. Posts Telecommun., Natural Sci. Ed.*, vol. 30, no. 2, pp. 216–221, 2018.
- [3] X. Geng, L. Shen, K. Li, and P. An, "A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property," *Signal Process., Image Commun.*, vol. 52, pp. 54–63, Mar. 2017.
- [4] 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.
- [5] S. Khan and S. S. Channappayya, "Estimating depth-salient edges and its application to stereoscopic image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5892–5903, Dec. 2018.
- [6] 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.
- [7] X. Liu, K. Kang, and Y. Liu, "Stereoscopic image quality assessment based on depth and texture information," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2829–2838, Dec. 2017.
- [8] J. Ma, P. An, L. Shen, and K. Li, "Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation," *IEEE Access*, vol. 6, pp. 2768–2780, 2018.
- [9] L. Ma, X. Wang, Q. Liu, and K. N. Ngan, "Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment," *Neurocomputing*, vol. 215, pp. 21–31, Nov. 2016.
- [10] F. Qi, D. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2338–2344, Dec. 2015.
- [11] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1077–1084, Jun. 2016.
- [12] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, and Q. Dai, "Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 730–743, Mar. 2016.
- [13] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4923–4936, Oct. 2017.
- [14] Y. Yao, L. Shen, and P. An, "Bivariate analysis of 3D structure for stereoscopic image quality assessment," *Signal Process., Image Commun.*, vol. 65, pp. 128–140, Jul. 2018.

- [15] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Process., Image Commun.*, vol. 58, pp. 287–299, Oct. 2017.
- [16] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685–1699, May 2015.
- [17] M. Karimi, M. Nejati, S. M. R. Soroushmehr, S. Samavi, N. Karimi, and K. Najarian, "Blind stereo quality assessment based on learned features from binocular combined images," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2475–2489, Nov. 2017.
- [18] R. P. Rutstein, R. J. Fullard, J. A. Wilson, and A. Gordon, "Aniseikonia induced by cataract surgery and its effect on binocular vision," *Optometry Vis. Sci.*, vol. 92, no. 2, pp. 201–207, Feb. 2015.
- [19] M. Kytö, A. Mäkinen, T. Tossavainen, and P. Oittinen, "Stereoscopic depth perception in video see-through augmented reality within action space," *J. Electron. Imag.*, vol. 23, no. 1, Mar. 2014, Art. no. 011006.
- [20] L. F. Lago-Fernández and G. Deco, "A model of binocular rivalry based on competition in IT," *Neurocomputing*, vols. 44–46, pp. 503–507, Jun. 2002.
- [21] 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, Nov. 2000.
- [22] E. L. Smith, K. Fern, R. Manny, and R. S. Harwerth, "Intercular suppression produced by rivalry stimuli: A comparison of normal and abnormal binocular vision," *Optometry Vis. Sci.*, vol. 71, no. 8, pp. 479–491, Aug. 1994.
- [23] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 341–353, Jun. 2018.
- [24] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.
- [25] G. Yue, C. Hou, Q. Jiang, and Y. Yang, "Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry," *Signal Process.*, vol. 150, pp. 204–214, Sep. 2018.
- [26] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder," *Pattern Recognit.*, vol. 76, pp. 242–255, Apr. 2018.
- [27] Y. Li and X. Hu, "No-reference stereoscopic image quality assessment using natural scene statistics," in *Proc. 2nd Int. Conf. Multimedia Image Process.*, Wuhan, China, Jan. 2017, pp. 123–127.
- [28] J. Wang, K. Zeng, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic images from single views," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Chengdu, China, Jul. 2014, pp. 1–6.
- [29] Y. Ding and Y. Zhao, "No-reference stereoscopic image quality assessment guided by visual hierarchical structure and binocular effects," *Appl. Opt.*, vol. 57, no. 10, pp. 2610–2621, Apr. 2018.
- [30] F. Shao, Z. Zhang, Q. Jiang, W. Lin, and G. Jiang, "Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 573–585, Mar. 2018.
- [31] J. Yang, B. Jiang, H. Song, X. Yang, W. Lu, and H. Liu, "No-reference stereoimage quality assessment for multimedia analysis towards Internet-of-Things," *IEEE Access*, vol. 6, pp. 7631–7640, 2018.
- [32] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognit.*, vol. 71, pp. 207–217, Nov. 2017.
- [33] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 721–734, Feb. 2018.
- [34] S. Lee and H. S. Yang, "Quantisation-free binary feature aggregation for fast image representation," *Electron. Lett.*, vol. 53, no. 22, pp. 1470–1471, Oct. 2017.
- [35] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh, "No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models," *Signal Process., Image Commun.*, vol. 54, pp. 107–117, May 2017.
- [36] Z. A. Khan, M. Kaaniche, F. A. Cheikh, and A. Beghdadi, "Joint statistical models for no-reference stereoscopic image quality assessment," in *Proc. 7th Eur. Workshop Visual Inf. Process.* Tampere, Finland: Institute of Electrical and Electronics Engineers, Nov. 2018, pp. 1–5.
- [37] M. Karimi, N. Soltanian, S. Samavi, K. Najarian, N. Karimi, and S. M. R. Soroushmehr, "Blind stereo image quality assessment inspired by brain sensory-motor fusion," *Digit. Signal Process.*, vol. 91, pp. 91–104, Aug. 2019.
- [38] J. Yang, K. Sim, W. Lu, and B. Jiang, "Predicting stereoscopic image quality via stacked auto-encoders based on stereopsis formation," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1750–1761, Jul. 2019.
- [39] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.
- [40] D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery," *J. Experim. Psychol., Appl.*, vol. 7, no. 2, pp. 143–153, Jun. 2001.
- [41] S. A. Fezza, A. Chetouani, and M.-C. Larabi, "Using distortion and asymmetry determination for blind stereoscopic image quality assessment strategy," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 115–128, Nov. 2017.
- [42] 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.
- [43] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Sep. 2013.
- [44] S.-H. Cho and H.-B. Kang, "An analysis of visual discomfort caused by watching stereoscopic 3D content in terms of depth, viewing time and display size," *J. Imag. Sci. Technol.*, vol. 59, no. 2, pp. 205031–2050316, Mar. 2015.
- [45] 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.
- [46] G. Jiang, H. Xu, M. Yu, T. Luo, and Y. Zhang, "Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 269–279, Jul. 2017.
- [47] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.



YONG CHEN received the M.S. and Ph.D. degrees from Chongqing University, China, in 1995 and 2003, respectively. His research interest includes image processing.



KAIXIN ZHU is currently pursuing the M.E. degree with the Chongqing University of Posts and Telecommunications, China. Her research interest includes image processing.



LIU HUANLIN received the M.S. degree from the Chongqing University of Posts and Telecommunications, China, in 2000, and the Ph.D. degree from Chongqing University, China, in 2008. Her research interests include information acquiring and processing.

...