Received August 21, 2019, accepted September 6, 2019, date of publication September 13, 2019,
date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941112

Learning a No-Reference Quality Predictor of Stereoscopic Images by Visual Binocular Properties

YUMING FANG[1], (Senior Member, IEEE), JIEBIN YAN[1], JIHENG WANG[2], (Member, IEEE), XUELIN LIU[1], GUANGTAO ZHAI[3], (Senior Member, IEEE), AND PATRICK LE CALLET[4], (Fellow, IEEE)

[1]School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330032, China
[2]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada
[3]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[4]Équipe Image, Perception et Interation, Laboratoire des Sciences du Numérique de Nantes, Université de Nantes, 44306 Nantes, France

Corresponding author: Guangtao Zhai (zhaiguangtao@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61822109 and Grant 61571212, and in part by the Henry Fok Education Foundation under Grant 161061.

ABSTRACT In this work, we develop a novel no-reference (NR) quality assessment metric for stereoscopic images based on monocular and binocular features, motivated by visual perception properties of the human visual system (HVS) named binocular rivalry and binocular integration. To be more specific, we first calculate the normalized intensity feature maps of right- and left-view images through local contrast normalization, where statistical intensity features are extracted by the histogram of the normalized intensity feature map to represent monocular features. Then, we compute the disparity map of stereoscopic image, with which we extract structure feature map of stereoscopic image based on local binary pattern (LBP). We further extract statistical structure features and statistical depth features from structure feature map and disparity map by histogram to represent binocular features. Finally, we adopt support vector regression (SVR) to train the mapping function from the extracted monocular and binocular features to subjective quality scores. Comparison experiments are conducted on four large-scale stereoscopic image databases and the results demonstrate the promising performance of the proposed method in stereoscopic image quality assessment.

INDEX TERMS Stereoscopic images, no reference, image quality assessment, asymmetric distortion, symmetric distortion.

I. INTRODUCTION

In recent years, the rapid development of stereoscopic imaging technologies makes stereoscopic images become the import medium of information transmission, stereoscopic image processing techniques have become a hot topic in both academic research and industry applications, and visual quality estimation of stereoscopic images plays an important role in various stereoscopic image processing techniques. There are many researchers investigating subjective [1]–[4] and objective quality assessment [5]–[16], [29], [30] for stereoscopic images. Subjective quality assessment methods invite subjects to provide perceptual quality scores of images by human ratings, while objective metrics estimate the perceptual quality scores of images automatically. As we know, subjective quality evaluation is complicated and time-consuming, and thus, it cannot be embedded in real-time systems. Objective quality assessment methods predicting image quality automatically can be easily applied in multimedia processing systems.

Over the past few years, there has been substantial effort made to develop objective image quality assessment (IQA) metrics. According to the available reference information, we can classify IQA metrics into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR) approaches. Among IQA methods, FR approaches compute the visual quality score of a distorted image by comparing the reference image with this distorted image, such as structure similarity (SSIM) [17], peak signal-to-noise ratio (PSNR), structure features and uncertainty weighting (SFUW) [18]. They require the complete reference

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

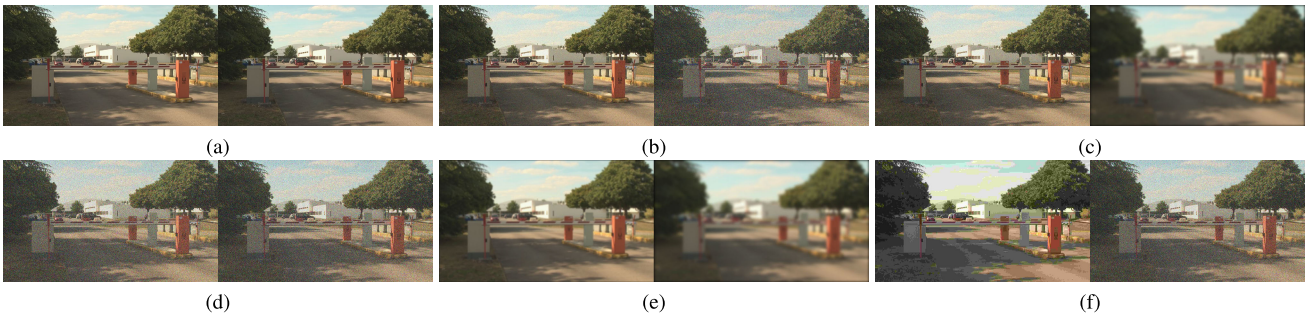

FIGURE 1. Examples of asymmetric distortion and symmetric distortion of stereoscopic images [28]. (a): reference stereoscopic image; (b) and (e): distorted stereoscopic images by additive Gaussian noise and Gaussian blur with different distortion levels;(c): distorted stereoscopic image by additive Gaussian noise for left view and Gaussian blur for right view; (d): distorted stereoscopic image by additive Gaussian noise with the same distortion level for both views; (f): distorted stereoscopic image is altered by additive Gaussian blur for right view and JPEG compression for left view.

information when predicting visual quality of images. RR approaches only need partial information of the reference image when estimating the visual quality of images [19]. Different from FR and RR approaches, NR approaches compute the visual quality of images without any reference information, such as natural image quality evaluator (NIQE) [20], blind image spatial quality evaluator (BRISQUE) [21], robust image sharpness evaluation (RISE) [22].

These aforementioned FR, RR and NR IQA approaches are developed for 2D image quality assessment (2D-IQA). For stereoscopic image quality prediction, one simple yet effective solution is to apply 2D-IQA methods to right- and left-view images directly, and then combine the predicted visual quality scores of these two views by certain pooling strategy. In [5], the authors pointed out that using 2D-IQA methods to predict visual quality of stereoscopic images would obtain poor performance in the case of asymmetric distortion and it only works well in the case of symmetrical distortion. In Fig. 1, we give some samples of stereoscopic images with asymmetrical and symmetrical distortion to illustrate the difference between these two distortion types. Asymmetrical distortion means that right- and left-view images suffer from different kinds of distortions or different distortion levels, while symmetrical distortion means that right- and left-view images suffer from the same amount of distortion [2].

As indicated by existing studies concerning binocular perception [23]–[27], the perceptual effect of binocular rivalry occurs for presented inconsistent signals, while binocular integration occurs for presented consistent signals. As revealed by the binocular perception mechanism [23]–[27], the binocular rivalry is a phenomenon of visual perception from different images presented to each eye. Only right- or left-view image is perceived by the HVS when binocular rivalry occurs in the HVS, while both right- and left-view images are viewed by the HVS simultaneously when binocular integration occurs in the HVS. Inspired by this, we develop a novel NR perceptual quality assessment metric of stereoscopic images based on monocular and binocular features. First, with the consideration of different influence of variations in right- and left-view images on visual perception, we use normalized intensity features to measure the distortion of right- and left-view images. The normalized intensity maps of stereoscopic images are extracted through local contrast normalization and we extract the statistical intensity features from normalized intensity maps in the form of histogram. The statistical intensity features computed from stereoscopic images are used as monocular features. Second, we calculate the disparity map from right- and left-view images. The statistical depth features are extracted from the disparity map by histogram. Meanwhile, we combine the disparity information and normalized intensity information to extract structure features. Two structure feature maps are computed corresponding to left- and right-view images based on local binary pattern (LBP). We extract statistical structure features by the histogram of the structure feature map. Here, we use statistical depth features and statistical structure features as binocular features in the proposed method.

The feature vector including statistical intensity, depth and structure features is used to learn a perceptual quality estimation model. We conduct validation experiments on large-scale public databases and the results show that the developed method can obtain promising performance in stereoscopic image quality assessment over other existing FR and NR methods. In sum, the main contributions of the proposed method include the following aspects.

- Considering the effect of binocular rivalry that only right- or left-view image is perceived, we extract intensity features as monocular features to capture the visual distortion of stereoscopic images simultaneously.
- In view of that binocular integration occurs when right- and left-view images are consistent during stereoscopic image viewing, we extract the binocular features of stereoscopic images including structure and depth features in the proposed method.
- Different from existing methods which employ the parameters of fitted functions as certain features to estimate image quality, we utilize the histograms of different features to represent the monocular and binocular features to design a novel blind IQA metric for stereoscopic images. Compared with feature extraction by fitting

functions, the feature extraction in the proposed method shows better generalization ability, since it does not have to make the assumption that the used features obey some specific distributions.

II. RELATED WORK

The study [31] demonstrates that the perceived depth and 2D image quality determines the overall stereoscopic quality of experience. Inspired by the study on binocular rivalry [24], a ''cyclopean'' algorithm [5] was proposed for stereoscopic image quality assessment and it performed significantly better than the baseline IQA metrics on stereoscopic image quality datasets including both asymmetric and symmetric distorted stereoscopic images. It have been widely accepted that binocular perceptual properties are of great importance in developing effective IQA algorithms for stereoscopic images [2], [5], [8].

Currently, considering binocular characteristics of the HVS [2], [5], many FR quality prediction methods have been proposed for stereoscopic images. In [28], Wang *et al.* proposed a multi-scale IQA method to estimate the visual quality of stereoscopic images inspired by binocular rivalry. The authors proposed an divisive normalization and information content-based pooling strategy for stereoscopic image quality assessment and developed a 2D-to-3D quality estimation metric to compute the overall quality score of the stereoscopic image. In [8], binocular integration behaviors are utilized as the basis to compute the perceptual quality of stereoscopic images. In [32], Bensalma *et al.* developed a binocular energy quality metric (BEQM) with the consideration of binocular fusion in the HVS where binocular energy is used to calculate the visual distortion of stereoscopic images. In [33], Cao *et al.* presented a primary visual processing mechanism (PVPM) based model for estimating the perceptual quality of stereoscopic images, in which the stereoscopic image was divided into binocular and monocular regions for quality estimation. Qi *et al.* proposed a RR IQA method for stereoscopic image quality assessment by using binocular visual features, which are calculated by the entropy of right- and left-view images [34].

Besides FR and RR IQA methods designed for perceptual quality estimation of stereoscopic images, there have also been some NR IQA models developed for perceptual quality evaluation of stereoscopic images [2], [6], [35], [36]. In [2], Chen *et al.* proposed a NR IQA method for stereoscopic images by natural scene statistics (NSS) features, including both 3D and 2D features. In [35], the authors employed local features and disparity to construct a NR quality assessment model for JPEG coded stereoscopic images. In [36], the authors proposed a unified NR quality evaluator for multiply and singly distorted stereoscopic images by learning binocular and monocular local perceptual primitives, which characterize the underlying binocular and monocular local receptive field properties of the visual cortex in response to stereopairs.

Ryu *et al.* designed a perceptual quality estimation metric for stereoscopic images by computing blockiness and blurriness scores in right- and left-view images [37]. In that study, the authors calculated blurriness by using spatial activity and wavelet diagonal coefficients, and measured the blockiness based on the marginal distribution of local wavelet coefficients. The overall quality score of the stereoscopic image is calculated by linear summation of blockiness and blurriness scores. In [38], Zhou *et al.* used binocular energy response (BER) and binocular rivalry response (BRR) of stereoscopic images to predict stereoscopic image quality. BER can be expressed as responses of a couple of monocular simple cells perceived by both eyes, while BRR is a perceptual effect that occurs when both left and right eyes view mismatched right- and left-view images at the same retinal location [38]. The authors used the local magnitude pattern and local directional pattern to extract local patterns of BER and BRR to evaluate the perceptual quality of stereoscopic images [38].

In [39], Zhang *et al.* introduced a 3-channel convolutional neural network (CNN) model to estimate visual quality of stereoscopic images, whose input includes the patches of left- and right-view images and the image patch difference between the corresponding right- and left-view images. The authors utilized subjective quality scores of the whole stereoscopic image to represent the ground-truth of the input image patches for training. Lv *et al.* designed a NR IQA model for stereoscopic images based on binocular self-similarity (BS) binocular integration (BI) [40]. In [40], the authors computed BS by measuring the similarity of original left-view and synthesized images, and calculated BI by the trained CNN. In [41], Zhou *et al.* proposed to use sparse coding to build stereoscopic image quality estimation model. In that study, the authors utilized K-Nearest Neighbor (KNN) method to construct stereoscopic image quality estimation metric based on the assumption that stereoscopic images have similar visual quality if they have similar quality-aware features.

Although many IQA models have been designed for stereoscopic images, predicting the perceptual quality of stereoscopic images without reference images is still challenging. In this paper, we propose a novel NR IQA method for stereoscopic images by monocular and binocular features. First, the statistical intensity features, regarded as monocular features, are extracted from both left- and right-view images; the statistical depth and structure features, regarded as binocular features, are computed from disparity map and single-view images. Then the quality prediction model is learned by support vector regression (SVR) from the feature vector including both monocular and binocular features to subjective quality scores.

III. PROPOSED METHOD

The framework of the proposed method is given in Fig. 2. First, we apply local contrast normalization to stereoscopic images which mimics the early behavior of the HVS and remove visual information redundancy [20], [21], and the

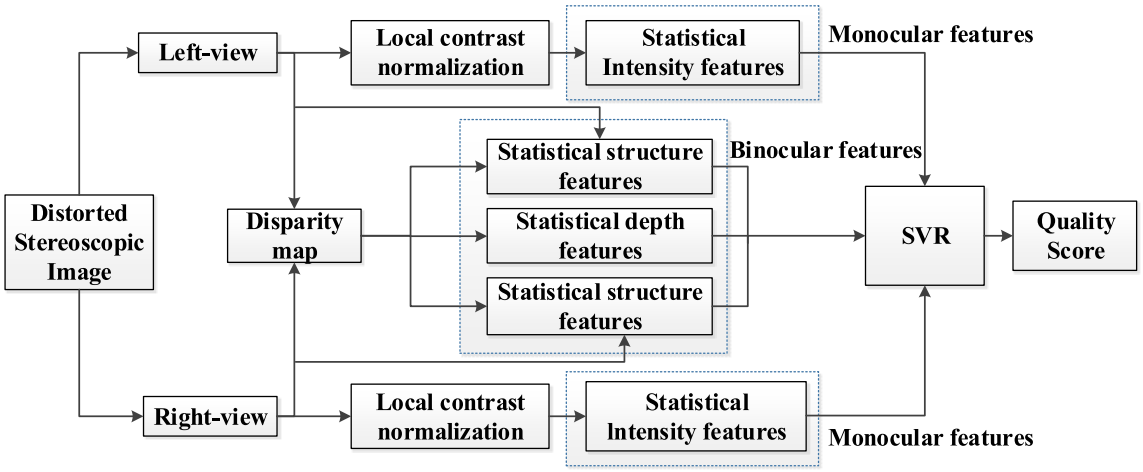

FIGURE 2. The framework of the proposed method.

statistical intensity features from the normalized intensity feature maps are then extracted. Second, we estimate the disparity map of the right- and left-view images, and further extract the depth features from the disparity map by histogram. Meanwhile, we combine the disparity map with the intensity maps of right- and left-view images to compute two corresponding structure feature maps by LBP. Finally, SVR is adopted to train the mapping function which maps the quality-aware features space including structure, intensity and depth features to subjective quality scores.

A. MONOCULAR FEATURE EXTRACTION

Local contrast normalization is a local non-linear process to log-contrast luminance, and the statistical dependence of input visual signals can be eliminated via this operation [46]. In this work, we adopt a normalization operation [20], [21] to right- and left-view images as:

$$\hat{\mathbf{I}}(j,i) = \frac{\mathbf{I}(j,i) - \mu_{\mathbf{I}}}{\sigma_{\mathbf{I}} + \mathbb{C}} \tag{1}$$

where $\mathbf{I}(j,i)$ and $\hat{\mathbf{I}}(j,i)$ denote the pixel values at the spatial location (j,i) in the original and normalized intensity maps, respectively. Here, $\mathbb{C}$ is used to avoid instability when denominator is close to zero, and we set $\mathbb{C}$ as 6.5025 [17]. $\mu_{\mathbf{I}}$ and $\sigma_{\mathbf{I}}$ denote the mean and standard deviation of the local region, which are computed as:

$$\mu_{\mathbf{I}} = \sum_{x=-X}^{Y}\sum_{y=-Y}^{Y} \varphi_{(x,y)}\mathbf{I}(j+x, i+y) \tag{2}$$

$$\sigma_{\mathbf{I}} = \sqrt{\sum_{x=-X}^{X}\sum_{y=-Y}^{Y} \varphi_{(x,y)}[\mathbf{I}(j+x, i+y) - \mu_{\mathbf{I}}]^2} \tag{3}$$

where $\{\varphi_{(x,y)}|x = -X, \ldots, X; y = -Y, \ldots, Y\}$ is a unit-volume Gaussian weighting window (X and Y are set as 3 [21]).

As introduced in [47], these normalized intensity values follow a Gaussian distribution, and NSS features extracted by these intensity values are widely adopted in existing 2D-IQA metrics [48]–[52]. Most existing studies of 2D image quality assessment [20], [21] and stereoscopic image quality assessment [1] fit a generalized Gaussian distribution (GGD) by normalized intensity values and the fitted GGD parameters are used as visual perceptual features. Different from these existing studies [1], [20], [21], we propose to employ histogram to extract intensity features in this work. It is advantageous to employ histogram to extract statistical features compared with the feature representation method by fitting functions [1], [20], [21]. The feature representation by histogram in the proposed method shows better generalization ability than that by fitting functions, since we don't have to make the assumption that the features obey some specific distributions. Here, we obtain the intensity features with 15 elements: $\{\mathbf{h}_1, \mathbf{h}_2, \cdots \mathbf{h}_{15}\}$ by setting the number of histogram bins as 15. The histogram of normalized intensity values is computed below:

$$\mathbf{h}_b = \frac{1}{HW}\sum_{b=1}^{B}\sum_{j=1}^{H}\sum_{i=1}^{W} \Omega(|\hat{\mathbf{I}}(j,i)|, G(b)) \tag{4}$$

$$\Omega(p,q) = \begin{cases} 1, & p \in q \\ 0, & otherwise \end{cases} \tag{5}$$

where b represents the bin index of the histogram; $G(b)$ is the interval of the b-th bin; W and H denote the width and height of the image, respectively.

In Fig. 3, we provide some visual samples of distorted stereoscopic images degraded by different distortion types. As shown in Fig. 3, due to the presence of visual distortion, the calculated feature maps are altered and different distortion levels can be clearly observed by feature maps. From Fig. 3 (e) and (n), we can observe that different distortion types (white noise and Gaussian blur) destroy the image content in different ways, which can be also reflected by characteristics of feature distributions. Thus, the statistical

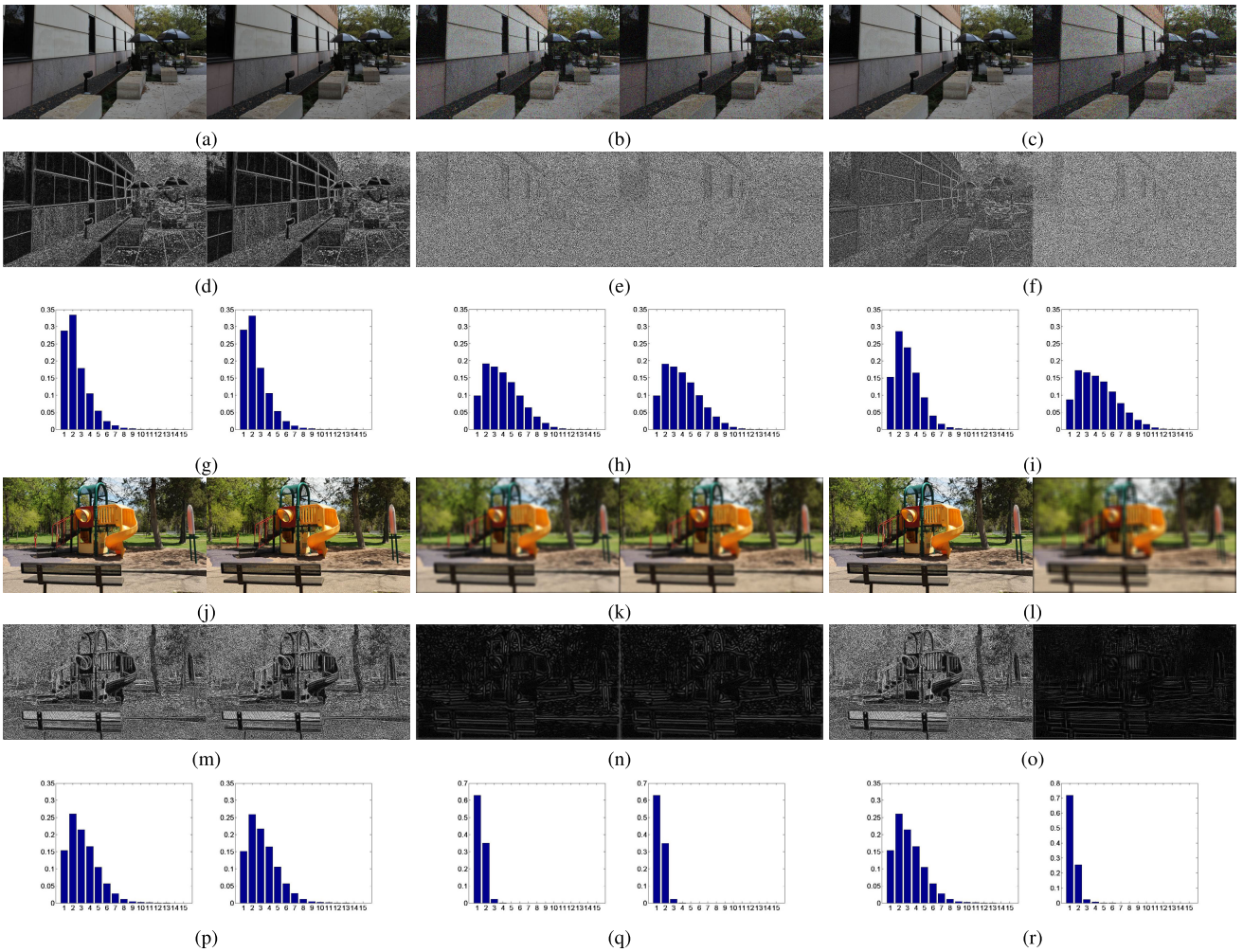

FIGURE 3. The visual examples of stereoscopic images and their corresponding feature maps. (a): a reference stereopair; (b) and (c): distorted stereopairs degraded by white noise symmetrically and asymmetrically, respectively; (d)(e)(f) and (g)(h)(i): the corresponding normalized intensity maps and distributions for stereoscopic images in (a), (b) and (c); (j): the reference stereoscopic image pair; (k) and (l): distorted stereoscopic images degraded by Gaussian blur symmetrically and asymmetrically, respectively; (m)(n)(o) and (p)(q)(r): the corresponding normalized intensity maps and distributions for stereoscopic images in (j), (k) and (l).

intensity features can effectively capture the visual distortion of single-view in stereoscopic images.

B. BINOCULAR FEATURE EXTRACTION

In addition to monocular features, we extract binocular features including structure and depth features. Here, we first compute the disparity map denoted as d by SSIM based stereo algorithm [5] between the right- and left-view images, and extract structure features by integrating disparity map into right- and left-view images.

The structure features are extracted as shown in Fig. 4, where $l(j, i)$ and $d(j, i)$ denote values at location (j, i) in the intensity map of left-view image and disparity map, respectively. In order to obtain the structure feature of each pixel, we calculate the Euclidean distance between the center pixel and its neighboring pixels:

$$D_k = \sqrt{(l(j', i') - l(j, i))^2 + (d(j', i') - d(j, i))^2} \quad (6)$$

$$D_m = \frac{1}{8}\sum_{k=0}^{7} D_k \quad (7)$$

where j' and i' represent spatial indices and their values are in the range of $(j-1, j+1)$ and $(i-1, i+1)$, respectively; D_k denotes the distance between feature descriptors of the k-th neighboring pixel ($k \in \{0, 1 \ldots, 7\}$) and the centre pixel $\{l(j, i), d(j, i)\}$; D_m represents the mean value of the set $\{D_0, D_1, \ldots, D_7\}$ in a local patch.

As indicated in [53], LBP can be used to capture structure information effectively, and it has been used in the field of image quality evaluation [54]–[60]. Here, we use LBP to estimate the variation of structure information of left-view image from the angle of stereo after the transformation by Eqs. (6) and (7) as below.

$$\mathbf{P}_{K,R} = \sum_{k=0}^{K-1} t(D_k - D_m)2^{K-1-k} \quad (8)$$

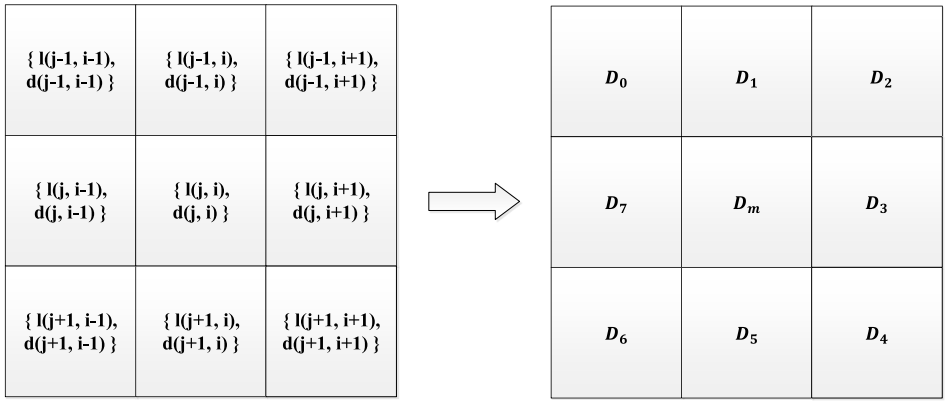

FIGURE 4. The example of combining left-view image with disparity map to extract the corresponding structure feature.

$$t(D_k - D_m) = \begin{cases} 1, & (D_k - D_m) \geq 0 \\ 0, & (D_k - D_m) < 0 \end{cases} \tag{9}$$

where R and K denote the radius of the neighborhood and the number of neighbors, respectively; $(D_0, D_1, \ldots, D_{(K-1)})$ represents the feature vector which is composed of K circularly symmetric neighborhood. As suggested in the study [57], LBP is effective in capturing the structure information when R and K are set to 1 and 8, respectively. When K increases, R would be larger and it would result in the increase of feature dimension. If R increases and K keeps a small value, the sampling points is insufficient and it makes the feature description unable to capture local structure features. Thus, we set $K = 8$ and $R = 1$ and a feature map are obtained by incorporating the left-view image with disparity map based on the aforementioned operations in Eqs. (8) and (9). Similarly, we can obtain another feature map by combining the disparity map and the intensity map of right-view image through the same operation. The statistical structure features of the right- and left-view images are extracted in the form of histogram as defined in Eqs. (4) and (5).

Meanwhile, we also extract the depth feature to capture the quality degradation of stereoscopic images. In the study [2], with the normalization operation defined in Eq. (1), the distribution of disparity can be modeled by a GGD and the parameters of GGD are used as quality-aware features. Here, different from the depth feature extraction method in [2], we first adopt the normalization operation defined in Eq. (1) to process the disparity map to obtain the normalized disparity map. Then, the histogram defined in Eqs. (4) and (5) is used to compute statistical depth features and the number of histogram bins is also set to 15.

Fig. 5 shows some original and distorted stereopairs, and their corresponding structure and depth feature distributions. By comparing the binocular feature distributions of the original and distorted stereopairs, it can be seen that the shapes of these distributions are altered more in the case of noise distortion than these in other distortion types, which suggests that the influence of distortion on binocular perception (fusion and rivalry) is more sensitive. This is consistent with previous findings on the role of noise contaminations with stereoscopic image quality [28] and stereoscopic depth quality [45]. With respect to depth feature distributions, it can be seen that a larger dynamic range of the distributions exists for noisy images than that for blurry images. This trend, i.e., the shapes of feature distribution change (large and narrow range), is consistent with the distortion type dependency found in previous subjective quality assessment studies [28], [45] that the influence of noise and blur on stereoscopic image quality are significantly different. Thus, the extracted binocular features, i.e., structure and depth features, can be used as effective features to represent influence of different distortion types on the perceived stereoscopic image and depth quality.

C. QUALITY PREDICTION

In this work, we extract statistical intensity features, structure features and depth features by employing the histogram. The bin number is set to 15. Thus, we can obtain 30 features for intensity information from right- and left-view images, 30 features for structure information from right- and left-view images, and 15 features for depth information of a stereoscopic image. In total, we can obtain 75 features for a distorted stereoscopic image pair. A regression model is used to learn a mapping function from feature space to quality scores. In our implementation, SVR with radial basis function (RBF) kernel [61] is used as the mapping tool for feature pooling from the feature vector to quality measure. The standard form of SVR with the parameters α and β is given by:

$$\min_{w,b,\delta,\delta'} \frac{1}{2} w^T w + \alpha \sum_{s=1}^{S} (\delta_s + \delta'_s) \tag{10}$$

$$subject\ to\ w^T \phi(h_s) + b - y_s \leq \beta + \delta_s \tag{11}$$

$$y_s - w^T \phi(h_s) - b \leq \beta + \delta'_s \tag{12}$$

$$\delta_s, \quad \delta'_s \geq 0, \ s = 1, \ldots, S. \tag{13}$$

where $\alpha > 0$ and $\beta > 0$; $K(h_i, h_j) = \phi(h_i)^T \phi(h_j)$ denotes the kernel function and RBF kernel $K(h_i, h_j) = exp(-\chi \left\| h_i - h_j \right\|^2)$. The parameters α and χ are set through

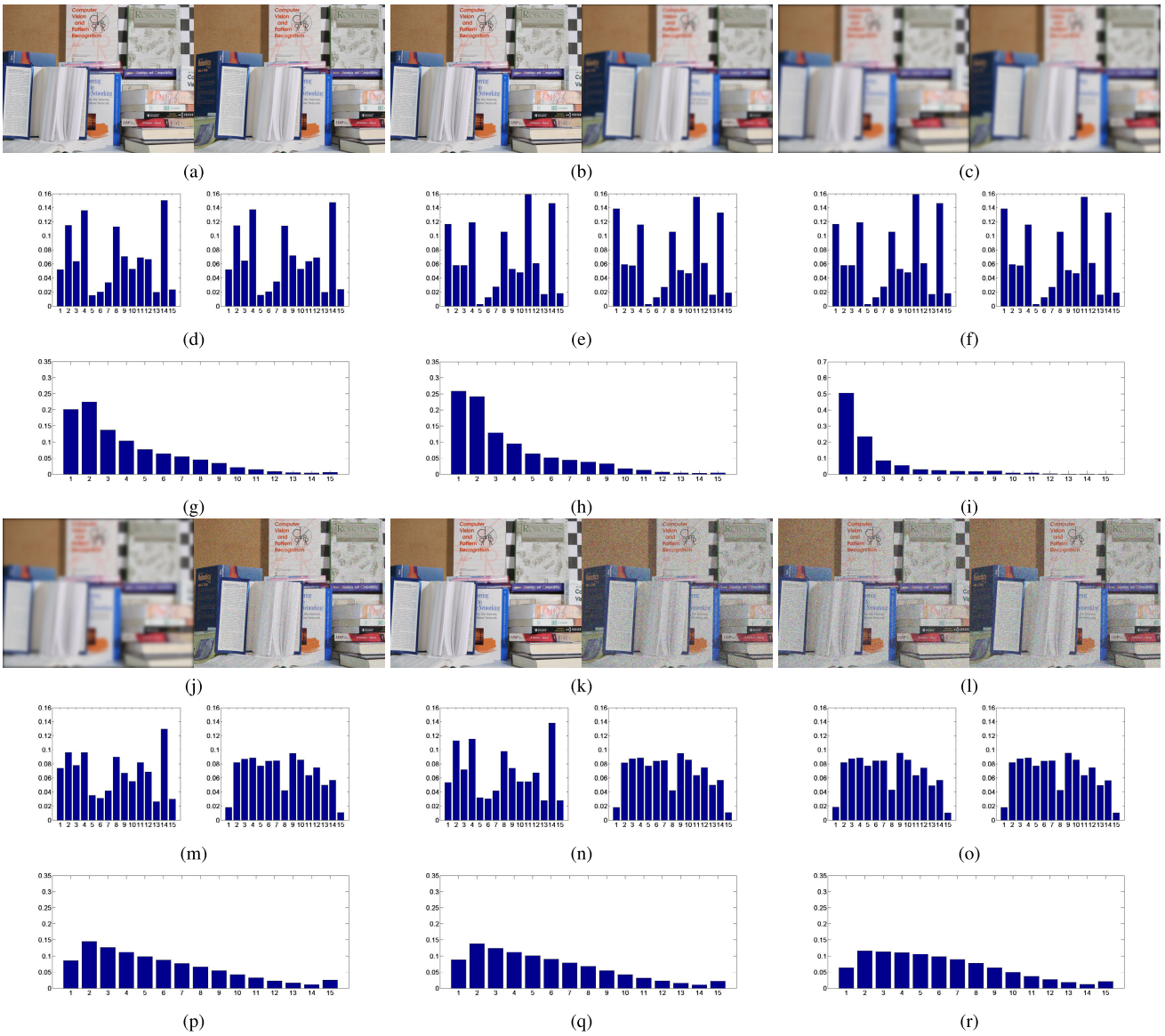

FIGURE 5. Original and distorted stereopairs, and their corresponding structure and depth feature distributions. (a): original pristine stereopair; (b): asymmetric distortion (original + bur); (c): symmetric distortion (blur + blur); (j): mixed distortion (blur + noise); (k): asymmetric distortion (original + noise); (l): symmetric distortion (noise + noise). (d), (e), (f), (m), (n), (o) and (g), (h), (i), (p), (q), (r) are their corresponding structure and depth feature distributions.

cross validation on the training data. We set χ to 1 which is fixed for all databases.

In the experiment, we select 80% images with the corresponding subjective scores randomly in each database for training and the remaining images of the same database are used for testing. This operation is repeated for 1000 times for each database where the training image and testing image sets change in repeated experiments, and the median performance [20], [21] is reported as the final result.

IV. EXPERIMENTAL RESULTS

A. STEREOSCOPIC IMAGE QUALITY DATABASES

We utilize four publicly available stereoscopic image quality databases to evaluate the proposed NR 3D-IQA algorithm, including Waterloo-IVC 3D Image Quality Database (Phase I and Phase II) [28] and LIVE 3D Image Quality Database (Phase I and Phase II) [1], [2], [5]. Please note that Waterloo-IVC Phase I/II and LIVE Phase II contain stereoscopic images with symmetrical and asymmetrical distortion, while LIVE Phase I only contains stereoscopic images with symmetrical distortion.

1) Waterloo IVC 3D PHASE I DATABASE [64]

There are 78 distorted single-view images and 330 distorted stereoscopic images in Waterloo IVC 3D Phase I Database. Three types of distortion with four levels of distortion, including Gaussian blur (GB), additive white Gaussian noise contamination (GN) and JPEG, are applied in single-view images

TABLE 1. Detailed information of stereoscoptic image quality databases.

Databases	Number	Spatial Resolution	Distortions	Subjects
Waterloo IVC 3D Phase I Database	330	1390*1080/1342*1080	WN, GB, JPEG	24
Waterloo IVC 3D Phase II Database	460	1920*1080	WN, GB, JPEG	24
LIVE 3D IQA Phase I Database	365	640*360	JPEG, JP2K, GB, WN, FF	32
LIVE 3D IQA Phase II Database	360	640*360	JPEG, JP2K, GB, WN, FF	33

of 6 pristine stereoscopic image pairs to produce the distorted images. This database contains stereoscopic images with symmetrical and asymmetrical distortion.

2) Waterloo IVC 3D PHASE II DATABASE [64]

Compared to Waterloo IVC 3D Phase I Database, there are more diverse image content in Waterloo IVC 3D Phase II Database. This database provides 130 distorted single-view images and 460 distorted stereoscopic images in total. Similar to Waterloo IVC 3D Phase I Database, the distorted images are generated by introducing the same types and levels of distortion in single-view images from 10 pristine stereoscopic image pairs. Both symmetrically and asymmetrically distorted stereoscopic images are included in this database.

3) LIVE 3D IQA PHASE I DATABASE [1]

This database consists of 20 reference images and 365 distorted images, which are degraded symmetrically by five common types of distortion, including GB, white noise (WN), Rayleigh fast-fading channel simulations (FF), JPEG2000 (JP2K) and JPEG. GB is used to create 45 distorted images, and the remaining types of distortion are used to generate 80 distorted images for each type of distortion. The corresponding subjective scores are available in the form of DMOS. Given a stereoscopic image, a higher DMOS represents a worse visual quality.

4) LIVE 3D IQA PHASE II DATABASE [2], [5]

Different from LIVE 3D IQA Phase I Database, LIVE 3D IQA Phase II Database consists of distorted stereoscopic images with asymmetrical and symmetrical distortion. The same types of distortion as LIVE 3D IQA Phase I Database are utilized to generate distorted images and each type of distortion produces 72 distorted stereoscopic images. There are a total of 120 symmetrically distorted and 240 asymmetrically distorted stereoscopic images in this database.

The detailed information of these databases is given in Table 1. As shown in Table 1, the spatial resolutions of images in Waterloo IVC 3D Phase I Database are different and the resolutions of images in Waterloo IVC 3D Phase II Database are the same; while the spatial resolutions of images in LIVE 3D IQA Phase I Database and LIVE 3D IQA Phase II Database are the same. The subjective scores for all these databases are not in the same scale. We extract the features for each stereoscopic image in gray-scale. When extracting features, we first convert original colorful stereoscopic images into gray-scale images. For each gray-scale image, each pixel is represented by 8 bits.

B. PERFORMANCE EVALUATION

To evaluate the performance of the algorithm, we use two performance evaluation metrics, including Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-order Correlation Coefficient (SRCC) [65]. PLCC and SRCC are correlation coefficients which measure the prediction accuracy of quality prediction metrics. Generally, the higher SRCC and PLCC values mean that the performance of the objective evaluation method is better. Given N images with subjective and objective scores, PLCC and SRCC can be calculated as follows.

$$PLCC = \frac{\sum_{i=1}^{N}(v_i - \overline{v})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(v_i - \overline{v}) * \sum_{i=1}^{N}(s_i - \overline{s})}} \tag{14}$$

$$SRCC = 1 - \frac{6\sum_{i=1}^{N}\epsilon_i^2}{N^3 - N} \tag{15}$$

where s_i and v_i are subjective and objective scores of the i-th image; $\overline{s}$ and $\overline{v}$ represent the mean values of subjective and objective scores, respectively; ϵ_i denotes the difference between the i-th image's ranks in subjective and objective results.

For fair performance comparison, the Video Quality Experts Group (VQEG) recommends reducing the nonlinearity of estimated image quality scores during performance evaluation [65]. Here, a five-parameter logistic regression function which is employed for nonlinearly regressing the quality scores into a common score space is defined as follows [65]:

$$F(x) = \Theta_1\left(\frac{1}{2} - \frac{1}{\exp(\Theta_2(x - \Theta_3)) + 1}\right) + \Theta_4 x + \Theta_5 \tag{16}$$

where $(\Theta_1, \ldots, \Theta_5)$ are fitted parameters via using objective and subjective quality scores.

C. EXPERIMENT COMPARISONS AND DISCUSSIONS

We compare the proposed method with state-of-the-art 2D and 3D IQA models, including PSNR, MS-SSIM [5], SSIM [17], IDW-SSIM [28], Chen [2], S3D-BLIND [9], Zhou [38], Zhang [39], Shao [62], and Shao [63]. Among these compared metrics, SSIM [17] and PSNR are designed for 2D image quality assessment, while the rest methods are designed for stereoscopic image quality assessment. PLCC

TABLE 2. Experimental results on LIVE 3D IQA DataBases, * means that the performance of the proposed is significantly different from that of the existing method (multi-comparison test, $p<0.05$). The best and second-best performance are denoted by boldface.

Databases	LIVE Phase I		LIVE Phase II	
Methods	PLCC	SRCC	PLCC	SRCC
PSNR	0.834*	0.834*	0.665*	0.665*
MS-SSIM [5]	0.917*	0.916*	0.900*	0.889*
SSIM [17]	0.873*	0.877*	0.802*	0.793*
IDW-SSIM [28]	0.873*	0.874*	0.916*	0.919*
Chen [2]	0.895	0.891	0.880	0.880
S3D-BLIND [9]	0.921	0.924	0.913	0.905
Zhou [38]	0.928	0.887	0.861	0.823
Shao [62]	0.926	0.928	**0.912**	**0.908**
Shao [63]	0.907	0.896	0.848	0.824
Zhang [39]	**0.947**	**0.943**	0.911	0.901
Proposed	**0.951**	**0.932**	**0.931**	**0.919**

and SRCC are used to evaluated to the performance of IQA methods.

In Table 2, we provide the performance of different IQA methods on LIVE 3D image quality Phase I and Phase II. As shown in Table 2, the quality prediction of the proposed NR method on both Phase I and Phase II is the best in terms of PLCC. The SRCC value of Zhang is higher than that of the proposed method due to its use of deep network for learning deep features. The consistent results of the proposed method from Phase I (including symmetric distortion only) and Phase II (including both symmetric and asymmetric distortion) indicate that the designed monocular and binocular features can reasonably account for the relationship between the binocular rivalry and binocular integration.

The experimental results on Waterloo IVC image quality databases are shown in Table 3. Please note that only experimental results of FR 2D-IQA methods PSNR and SSIM [17], and FR 3D-IQA methods MS-SSIM [5] and IDW-SSIM [28] are shown in Table 3, since we cannot get the source code of other related studies. From this table, the proposed method outperforms other metrics on both Phase I and Phase II database in terms of PLCC and SRCC. For [4], it is shown that the weighted averaging 2D-IQA estimation of right- and left-view images can remove the prediction bias by considering binocular rivalry mechanism. However, the simple combination of quality scores of right- and left-view images cannot yield good results. The interactions between multiple stereoscopic visual cues during stereoscopic visual perception are complex and non-intuitive. Here, the superiority of the proposed method suggests that high predicting accuracy in stereoscopic image visual quality assessment are obtained with introduced structure and depth features.

TABLE 3. Experimental results on the Waterloo IVC 3D IQA databases, * means that the performance of the proposed method is significantly different from that of the existing method (multi-comparison test, $p<0.05$). The best performance is denoted by boldface.

Databases	Waterloo Phase I		Waterloo Phase II	
Methods	PLCC	SRCC	PLCC	SRCC
PSNR	0.693*	0.521*	0.791*	0.495*
MS-SSIM [5]	0.734*	0.684*	0.613*	0.578*
SSIM [17]	0.711*	0.596*	0.728*	0.564*
IDW-SSIM [28]	0.930*	0.918*	0.892*	0.869*
Proposed	**0.953**	**0.950**	**0.936**	**0.922**

Moreover, in order to evaluate the generalization capabilities of the proposed method on LIVE 3D image quality databases, experimental results for individual distortion types in terms of PLCC are reported in Table 4. As shown in Table 4, 2D-IQA methods which are directly applied for stereoscopic image quality assessment works well on both LIVE 3D IQA Phase I and Phase II database in the case of white noise distortion. From Table 4, there are five types of distortion in the database. Compared with other IQA methods, the proposed method obtains better prediction performance for most distortion types. The main reason is that the effective monocular and binocular features used in the proposed method can measure the visual distortion from all these five distortion types. Zhang *et al.* [39] delivers higher performance than other methods in most cases since the structure and depth features of stereoscopic images are learned effectively and automatically by CNN model instead of handcrafted features in that study. Chen *et al.* [2] can obtain competitive performance on LIVE Phase II database, since both asymmetric- and symmetric-distorted stereopairs can be represented by the features of binocular rivalry used in that study. It also proves the effectiveness of binocular rivalry in stereoscopic image quality assessment.

We conduct the statistical test by using analysis of variance (ANOVA) followed by multi-comparison test and show some experimental results in Tables 2 and 3. From these experimental results, we conclude that the proposed method is significantly different from some compared IQA methods.

D. PARAMETER SETTINGS

We also conduct comparison experiments to test the parameter sensitivity of the proposed method. In our implementation, we separate each database into training samples and test samples without overlapping. We set the ratio of training samples to 50%, 60%, 70%, and 80%, and the median results of 1000 iterations are presented in Table 5. As shown in Table 5, the proposed method can obtain the best performance when the ratio of training samples is set to 80%. What's more, the performance decreases as the training samples decreases. When the proportion of training samples decrease to 50%, the proposed method can still obtain highly consistent results with subjective results, which proves the robustness of the proposed method. Here, we set the ratio of training samples as 80% for the optimal performance.

TABLE 4. Experimental results for individual distortion types in terms of PLCC on LIVE 3D IQA databases. The best and second-best performance are denoted by boldface for each distortion.

Databases	LIVE Phase I					LIVE Phase II				
Distortions	JPEG	JP2K	GB	WN	FF	JPEG	JP2K	GB	WN	FF
PSNR	0.219	0.785	0.916	0.935	0.703	0.491	0.597	0.690	0.919	0.730
SSIM [17]	0.487	0.865	0.919	0.939	0.721	0.678	0.704	0.838	0.922	0.834
Shao [62]	0.686	**0.939**	0.950	0.937	0.825	0.853	0.850	0.976	**0.956**	0.926
Chen [2]	0.695	0.907	0.917	0.917	0.735	**0.867**	**0.867**	0.900	0.950	**0.933**
Shao [63]	0.458	0.901	**0.952**	0.916	0.794	0.828	0.826	**0.984**	0.928	0.928
Zhang [39]	**0.740**	0.926	0.930	**0.944**	**0.883**	0.849	0.853	0.954	0.946	0.912
Proposed	**0.767**	**0.963**	**0.982**	**0.962**	**0.894**	**0.895**	**0.923**	**0.989**	**0.976**	**0.932**

TABLE 5. Experimental results with different ratios of training samples.

Databases	Waterloo Phase I		Waterloo Phase II		LIVE Phase I		LIVE Phase II	
Components	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
80%	0.953	0.950	0.936	0.922	0.951	0.932	0.931	0.919
70%	0.939	0.912	0.920	0.909	0.945	0.922	0.931	0.915
60%	0.937	0.916	0.909	0.892	0.937	0.917	0.921	0.915
50%	0.924	0.905	0.901	0.887	0.931	0.915	0.906	0.905

TABLE 6. Experimental results from different features in the proposed method.

Databases	Waterloo Phase I		Waterloo Phase II		LIVE Phase I		LIVE Phase II	
Components	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
Monocular	0.911	0.880	0.910	0.888	0.921	0.884	0.826	0.802
Binocular	0.925	0.891	0.861	0.833	0.942	0.919	0.906	0.895
Proposed	0.953	0.950	0.936	0.922	0.951	0.932	0.931	0.919

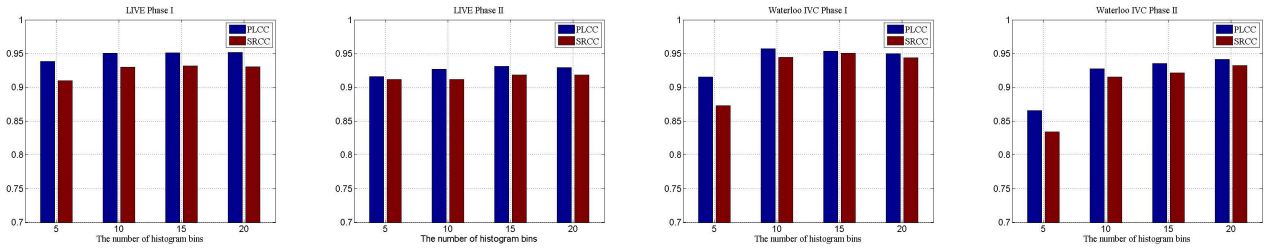

FIGURE 6. The experimental results of histograms with different numbers of bins on four stereoscopic image databases in terms of PLCC and SRCC.

To demonstrate the importance of the monocular and binocular features, we further conduct the comparison experiment and tabulate the experimental results in Table 6. As we can observe from this table, the proposed method by using binocular features can obtain better performance than using monocular features. Overall, the performance of the combination of monocular and binocular features is better than that with only one type of features, and the best performance is obtained when all the features are integrated.

In this work, we extract the features in the form of histogram. The performance of the proposed method might vary with the number of bins in feature histogram. In general, the feature histogram with a small number of bins cannot effectively represent the characteristics of feature distribution, while the feature histogram with a large number of bins is unstable. We perform a comparative experiment with different number of bins for histogram feature extraction. The experimental results are given in Fig. 6. As indicated in Fig. 6,

we set the number of histogram bins to 5, 10, 15 and 20, and the robust stereoscopic image quality prediction performance is obtained with different number of histogram bins. Considering the overall performance, the number of histogram bins is set as 15 to get stable performance in the proposed method.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel NR 3D-IQA algorithm for stereoscopic images based on monocular and binocular features, motivated by binocular perception mechanism including binocular rivalry and binocular integration. For monocular feature extraction, we first calculate the normalized intensity maps of right- and left-view images by local normalization, which are used to compute statistical intensity features. For binocular feature extraction, we first estimate the disparity map of stereopairs. Then, the estimated disparity map is combined with intensity maps of right- or left-view images to compute the structure features, and we also extract depth features from the estimated disparity map. After computing monocular and binocular features, we use SVR as the mapping tool for mapping the extracted perceptual feature vectors to subjective scores. Experimental results show that the prediction results of the proposed method are highly consistent with subjective ratings and the performance is better than most state-of-the-art methods, especially for asymmetrically distorted stereoscopic images.

Although the promising performance of the proposed method in visual quality assessment of stereoscopic images are achieved, there are still some aspects that we can investigate in the future. One is to design effective opinion-unaware methods for stereoscopic image quality assessment, which means that there is no subjective score in the training process. In the training stage, we can use FR objective quality assessment algorithms to calculate the visual quality scores of images which can be used as subjective ratings. We can also use the relative quality of image pairs to represent subjective ratings. Machine learning techniques can be further adopted to build stereoscopic image quality assessment models.

REFERENCES

[1] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Dec. 2013.

[2] 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.

[3] M. T. Pourazad, Z. Mai, P. Nasiopoulos, K. Plataniotis, and R. K. Ward, "Effect of brightness on the quality of visual 3D perception," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 989–992.

[4] J. Wang, K. Zeng, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic images from single views," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.

[5] 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, 2013.

[6] F. Shao, K. Li, W. Lin, G. Jiang, and Q. Dai, "Learning blind quality evaluator for stereoscopic images using joint sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2104–2114, Oct. 2016.

[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] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.

[9] 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.

[10] P. Hanhart, F. De Simone, and T. Ebrahimi, "Quality assessment of asymmetric stereo pair formed from decoded and synthesized views," in *Proc. IEEE Int. Conf. Qual. Multimedia Exper.*, Jul. 2012, pp. 236–241.

[11] X. Wang, L. Cao, L. Ma, Y. Zhou, and S. Kwong, "Complex singular value decomposition based stereoscopic image quality assessment," in *Proc. IEEE Conf. Vis. Commun. Image Process.*, Nov. 2016, pp. 1–4.

[12] Y. Zhu, G. Zhai, K. Gu, and M. Liu, "Blindly evaluating stereoscopic image quality with Free-Energy principle," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2016, pp. 2222–2225.

[13] Y. Fang, J. Yan, and J. Wang, "No reference quality assessment for stereoscopic images by statistical features," in *Proc. IEEE Int. Conf. Qual. Multimedia Exper.*, May/Jun. 2017, pp. 1–6.

[14] Q. Wu, H. Li, K. N. Ngan, and K. Ma, "Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2078–2089, Sep. 2018.

[15] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.

[16] Y. Zhou, L. Li, J. Wu, K. Gu, W. Dong, and G. Shi, "Blind quality index for multiply distorted images using biorder structure degradation and nonlocal statistics," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.

[17] 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.

[18] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, Apr. 2017.

[19] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.

[20] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[21] 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.

[22] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.

[23] L. Kaufman, "Sight and mind: An introduction to visual perception," *Amer. J. Psychol.*, vol. 87, no. 4, pp. 742–746, Dec. 1974.

[24] W. J. M. Levelt, "The alternation process in binocular rivalry," *Brit. J. Psychol.*, vol. 57, pp. 225–238, Nov. 1966.

[25] W. J. Tam, L. B. Stelmach, and P. J. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," *Proc. SPIE*, vol. 3295, pp. 226–235, Apr. 1998.

[26] 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, Jul. 2001.

[27] R. Blake, D. H. Westendorf, and R. Overton, "What is suppressed during binocular rivalry?" *Perception*, vol. 9, no. 2, pp. 223–231, Feb. 1980.

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

[29] A. K. Jain, A. E. Robinson, and T. Q. Nguyen, "Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 418–429, Mar. 2014.

[30] J. Wang, S. Wang, and Z. Wang, "Asymmetrically compressed stereoscopic 3D videos: Quality assessment and rate-distortion performance evaluation," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1330–1343, Mar. 2017.

[31] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 432–444, Jun. 2011.
[32] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, 2013.
[33] Y. Cao, W. Hong, and L. Yu, "Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2016, pp. 1–6.
[34] 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.
[35] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," *Proc. SPIE*, vol. 7524, Feb. 2010, Art. no. 75240T.
[36] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1866–1881, Apr. 2019.
[37] 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.
[38] 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.
[39] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.
[40] Y. Lv, M. Yu, G. Jiang, F. Shao, Z. Peng, and F. Chen, "No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network," *Signal Process., Image Commun.*, vol. 47, pp. 346–357, Sep. 2016.
[41] W. Zhou, W. Qiu, and M.-W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images," *IEEE Trans. Broadcast.*, vol. 63, no. 2, pp. 404–415, Jun. 2017.
[42] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4892–4904, Dec. 2013.
[43] H. C. Nothdurft, "Sensitivity for structure gradient in texture discrimination tasks," *Vis. Res.*, vol. 25, no. 12, pp. 1957–1968, 1985.
[44] J. Wang, S. Wang, and Z. Wang, "Depth perception of distorted stereoscopic images," in *Proc. 17th IEEE Int. Workshop Multimedia Signal Process.*, Oct. 2015, pp. 1–6.
[45] J. Wang, S. Wang, K. Ma, and Z. Wang, "Perceptual depth quality in distorted stereoscopic images," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1202–1215, Mar. 2017.
[46] S. Lyu and E. P. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[47] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[48] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[49] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[50] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
[51] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.
[52] Y. Zhang, J. Wu, X. Xie, L. Li, and G. Shi, "Blind image quality assessment with improved natural scene statistics," *Digit. Signal Process.*, vol. 57, pp. 56–65, Oct. 2016.
[53] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[54] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.
[55] Q. Li, W. Lin, and Y. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93–103, May 2016.
[56] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, Feb. 2015.
[57] J. Wu, W. Lin, G. Shi, and L. Xu, "Reduced-reference image quality assessment with local binary structural pattern," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2014, pp. 898–901.
[58] J. Wu, W. Lin, and G. Shi, "Image quality assessment with degradation on spatial structure," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 437–440, Apr. 2014.
[59] P. G. Freitas, W. Y. L. Akamine, and M. C. Q. Farias, "No-reference image quality assessment based on statistics of local ternary pattern," in *Proc. IEEE Int. Conf. Qual. Multimedia Exper.*, Jun. 2016, pp. 1–6.
[60] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.
[61] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
[62] F. Shao, W. Lin, G. Jiang, and Q. Dai, "Models of monocular and binocular visual perception in quality assessment of stereoscopic images," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 123–135, Jun. 2016.
[63] 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.
[64] J. Wang and Z. Wang, "Perceptual quality of asymmetrically distorted stereoscopic images: The role of image distortion types," in *Proc. Int. Workshop Video Process. Qual. Metrics Consumer Electron.*, 2014, pp. 1–6.
[65] *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*, VQEG, Mar. 2000.

YUMING FANG (M'13–SM'17) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, and computer vision. He also serves as an Associate Editor for the IEEE Access. He is on the Editorial Board of *Signal Processing: Image Communication*.

JIEBIN YAN received the B.E. degree from Beijing City University, Beijing, China, and the M.S. degree from the Jiangxi University of Finance and Economics, Nanchang, China, where he is currently pursuing the Ph.D. degree with the School of Information Technology. He was a Computer Vision Engineer with MTlab, Meitu, Inc. His research interests include visual quality assessment, saliency detection, and image composition.

JIHENG WANG (S'11–M'17) received the M.Math. degree in statistics computing and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011 and 2016, respectively. In 2013, he was with the Video Compression Research Group, Blackberry, Waterloo. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His current research interests include 3-D image and video quality assessment, perceptual 2-D and 3-D video coding, biomedical signal processing, statistical learning, and dimensionality reduction.

XUELIN LIU received the B.E. degree from the Jiangxi University of Finance and Economics, Nanchang, China, where he is currently pursuing the M.S. degree with the School of Information Technology. His research interests include image quality assessment and deep learning.

GUANGTAO ZHAI (M'10–SM'19) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2006 to 2007, he was a Student Intern with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Postdoctoral Fellow, from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. His research interests include multimedia signal processing and perceptual signal processing. He received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China, in 2012.

PATRICK LE CALLET (F'19) received the M.Sc. and Ph.D. degrees in image processing from the École Polytechnique de l'Université de Nantes. He was an Assistant Professor and a full-time Lecturer with the Department of Electrical Engineering, Technical Institute, University of Nantes, from 1997 to 1999 and 1999 to 2003, respectively. He led the Image and Video Communication Laboratory, CNRS, IRCCyN, from 2006 to 2016, and was one of the five members of the Steering Board of CNRS, from 2013 to 2016. Since 2015, he has been the Scientific Director of the Cluster Ouest Industries Cratives, a five-year program gathering over ten institutions (including three universities). Since 2017, he has been one of the seven members of the Steering Board of the CNRS LS2N Laboratory (450 researchers), as a Representative of Polytech Nantes. He is mostly involved in research dealing with the application of human vision modeling in image and video processing. His current research interests include the quality of experience assessment, visual attention modeling and applications, perceptual video coding, and immersive media processing.