

Research Article

No-Reference Stereoscopic Image Quality Assessment Based on Binocular Statistical Features and Machine Learning

Peng Xu,¹ Man Guo,² Lei Chen ¹, Weifeng Hu,¹ Qingshan Chen ¹ and Yujun Li ¹

¹School of Information Science and Engineering, Shandong University, Qingdao 266237, China

²Torch High Technology Industry Development Center, Ministry of Science & Technology, Beijing 100045, China

Correspondence should be addressed to Lei Chen; leiguanglg@hotmail.com and Yujun Li; liyujun@sdu.edu.cn

Received 26 September 2020; Revised 16 December 2020; Accepted 30 December 2020; Published 29 January 2021

Academic Editor: Jia Wu

Copyright © 2021 Peng Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Learning a deep structure representation for complex information networks is a vital research area, and assessing the quality of stereoscopic images or videos is challenging due to complex 3D quality factors. In this paper, we explore how to extract effective features to enhance the prediction accuracy of perceptual quality assessment. Inspired by the structure representation of the human visual system and the machine learning technique, we propose a no-reference quality assessment scheme for stereoscopic images. More specifically, the statistical features of the gradient magnitude and Laplacian of Gaussian responses are extracted to form binocular quality-predictive features. After feature extraction, these features of distorted stereoscopic image and its human perceptual score are used to construct a statistical regression model with the machine learning technique. Experimental results on the benchmark databases show that the proposed model generates image quality prediction well correlated with the human visual perception and delivers highly competitive performance with the typical and representative methods. The proposed scheme can be further applied to the real-world applications on video broadcasting and 3D multimedia industry.

1. Introduction

During the past few decades, there has been an exponential increase of stereoscopic images and videos in 3D display market [1]. However, due to various 3D quality factors [2, 3] including binocular rivalry, visual comfort, and depth perception, the visual quality assessment of stereoscopic images is much more complex and relatively less researched than the traditional 2D image quality evaluation. To address these challenges, we require a deeper understanding of binocular vision mechanisms and interactions for the quality prediction of distorted stereoscopic images.

There are mainly two groups of methods on 3D image quality assessment (IQA): subjective quality evaluation by human observer [4] and objective quality evaluation by devised metric used to simulate human perceptual judgments [5]. Since the human eyes are the final receiver of visual information, the subjective evaluation can directly reflect the human visual perception and is accurate and effective to evaluate the visual quality. However, the

subjective evaluation involves many participants in the course of experiments, which is time-consuming and costly. Therefore, it is unrealistic to implement it in many scenarios like real-time evaluation [6]. As a result, it is in urgent demand to propose objective methods that can effectively evaluate the human perceptual quality of stereoscopic images.

Based on the volume of accessible information in the images, existing objective quality assessment metrics can be generally divided into three categories: full-reference (FR) [7, 8], reduced-reference (RR) [9], and no-reference/blind (NR) methods [10, 11]. When the reference contents are accessible, the FR method can offer more accurate quality assessment. Early approaches for FR 3D-IQA directly stemmed from 2D quality metrics [12]. Conventionally, a straightforward way is to apply the 2D-IQA metrics to both views of a 3D image independently and then integrate the two 2D quality scores into a final 3D quality score. Several 3D-IQA methods [13, 14] were proposed by introducing the associated disparity or depth map into the 3D image quality

model. These research findings indicated that a satisfactory result can be obtained if the disparity images and reference images are combined appropriately. Afterwards, more sophisticated algorithms were developed based on the binocular vision properties. For example, Lin and Wu [15] revisited the physiological discoveries of binocular vision and incorporated the binocular integration into the existing 2D quality metrics for measuring the quality of stereoscopic images. Shao et al. [16] classified the stereoscopic images into noncorresponding, binocular fusion, and binocular suppression regions. Each region was evaluated individually according to its binocular perception property. In our previous work [17], we proposed a full-reference quality evaluator by considering the local and global qualities of 3D images. The experimental results showed its good performance in terms of stereoscopic image quality assessment.

Since pristine reference images are rarely available in practical applications [18, 19], the NR algorithms are potentially much more feasible solutions. They can give quality evaluation without any information extracted from the corresponding pristine image. The NR 3D-IQA is still preliminary, and a limited number of blind 3D-IQA algorithms have been developed. Inspired by the human visual system, Chen et al. [20] proposed a no-reference binocular image quality assessment method for natural stereopairs. The proposed method extracted both 2D and 3D natural statistical features from a stereopair and utilized these statistical features and the binocular rivalry for 3D image quality prediction. Ryu and Sohn [21] investigated the relationship between visual information and binocular quality perception and developed an NR quality evaluation algorithm for 3D images. The scores of perceptual blockiness and blurriness were combined into an overall quality index based on the binocular perception models. Shen et al. [22] devised a no-reference quality scheme for stereoscopic images based on the visual perceptual characteristics. Three types of features relating to image distortion, depth perception, and binocular disparity were used to map the human opinion scores. Other relevant works can be found in references [23–25].

Recently, machine learning techniques have achieved great success and been widely applied to various research fields [26–28]. One of the advantages of applying machine learning to quality evaluation is that it can directly take original image data as input and then combine feature learning with quality regression in the training procedure [29, 30]. Kang et al. [31] applied convolution neural network (CNN) to image quality assessment. They devised a shallow network which extracts quality-predictive features from image patches. Several NR algorithms for 3D-IQA using deep learning have been developed. Oh et al. [32] reported a novel deep learning method for NR 3D-IQA in terms of local to global feature aggregation. Zhou et al. [33] proposed a dual-stream interactive network for stereoscopic image quality assessment. In our previous work [34], we developed a no-reference quality prediction scheme for 3D images based on binocular features and support vector regression (SVR). The scheme showed its effectiveness, but the performance in terms of prediction accuracy and time

complexity needs to be further improved. More efficient methods for stereoscopic image quality assessment should be explored to address these limitations.

It is challenging for the NR algorithms to have the assessment accuracy as good as can be obtained with the FR quality evaluation methods. Moreover, 3D image quality databases generally lack large-scale training images with subjective quality scores, which limit the performance of these algorithms using deep neural networks. Other techniques should be explored and worthy of further research. We are motivated to tackle these limitations for 3D image quality assessment. In this paper, inspired by the research findings on the human binocular visual system, we try to simulate the perceptual mechanism of binocular vision. We primarily work on extracting certain types of binocular features from distorted stereoscopic image and constructing a statistical regression model to map these quality-aware features to the human perceptual judgements. The main contributions of this work are as follows:

- (1) Different from other related studies [33, 35], the novelty of our work lies in that we propose to adopt the effective binocular statistical features from the fusion and difference maps of a stereopair for stereoscopic image quality prediction.
- (2) We have demonstrated that appropriate combination of binocular features and binocular energy can greatly promote the performance of 3D image quality evaluation.
- (3) Compared with other typical and representative methods, the proposed scheme achieves higher consistent alignment with human subjective assessment and has lower time complexity. The experimental results show that our scheme can accurately estimate the perceptual quality of distorted stereoscopic images and has promising generalization ability.

The remainder of this paper is organized as follows. Section 2 introduces some fundamental knowledge about binocular visual perception. Section 3 presents the proposed quality assessment scheme for stereoscopic images in detail. Section 4 gives the experimental results and performance analysis of the proposed scheme and the comparison with other related algorithms. Finally, Section 5 concludes the paper with possible ideas for future work.

2. Foundation for Binocular Visual Perception

It has been known that the binocular vision is a complex visual process that requires the brain and both eyes working together to produce clear vision. Figure 1 describes a simplified framework of two important visual neural pathways for the binocular visual system. The ventral stream starts from the primary visual cortex V1 and goes through V2 and V3 to V4 area. The functions of ventral stream are about the recognition and perception behaviors. The dorsal stream begins from the V1 area, goes through V2 and V3 to V5 area. The visual information-guided interactions occur in dorsal

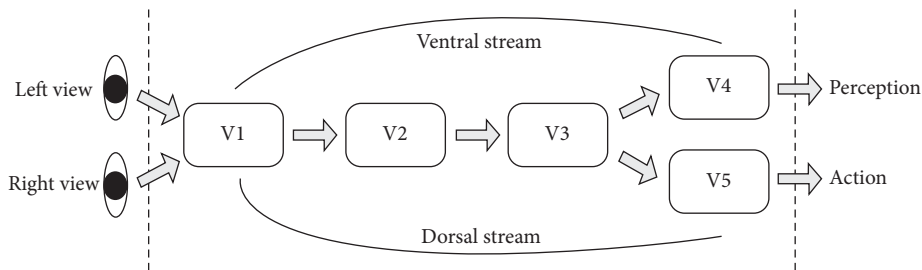


FIGURE 1: A simplified illustration of visual cortex with ventral stream and dorsal stream.

stream [15]. Regarding the detailed functions of each visual area, please refer to the binocular visual perception book [36] for further information.

The visual cortex plays an important role in our binocular visual perception, and it has been demonstrated that the primary visual cortex (V1) is mainly responsible for the human visual system (HVS) [37]. In the V1, simple and complex receptive fields are usually characterized to understand the behavior of visual perception. According to visual psychophysical study, two visual phenomena usually occur in the process of binocular visual response: binocular rivalry and binocular fusion. When the two eyes view mismatched images at the same retinal location, one experiences binocular rivalry. As a result of competition between the eyes, binocular rivalry involves reciprocal inhibition between the monocular channels. When two slightly different retinal signals can be perceived by two eyes, one experiences binocular fusion. During fusion, two retinal points are integrated into one single perception, superimposing and combining similar contents from the two views. Therefore, the binocular vision can be generally considered as a combination of binocular rivalry and binocular fusion.

As a significant content of primitives in V1, image structural information is closely related to image visual quality. And the degradation of perceptual quality can be reflected via the change of image structural information. Previous studies [29, 38] highlighted the significance of image structural information for image quality assessment. The gradient magnitude (GM) and Laplacian of Gaussian (LOG) are basic elements that are commonly used to represent image semantic structures [39]. More importantly, during 3D visual stimuli processing, binocular fusion and disparity responses are primitively formed in the V1 cortical area. The visual signals from the binocular summation and subtraction channels are multiplexed, and then each neuron in V1 receives a weighted sum of the visual stimuli from these two channels [40]. Motivated by these research results, in this paper, we extract the GM and LOG features from a stereopair and its fusion and difference maps as binocular features. In the following section, we will describe our proposed quality prediction model for distorted stereoscopic images in detail.

3. The Proposed No-Reference Quality Assessment Scheme

Figure 2 illustrates the architecture of the proposed scheme for stereoscopic image quality prediction. Given an original stereopair, we first generate the fusion map and difference map and extract the binocular statistical features from them as basic feature vectors. Then, we calculate the binocular energy responses from the local amplitude and local phase of the stereopair as quality-aware features. Finally, we employ an extreme learning machine method to map these features of distorted stereopair to its human perceptual quality score.

3.1. Binocular Feature Extraction. As can be seen from Figure 3, the fusion maps and difference maps of the left- and right-view images with different distortion types are discriminative, which can be utilized for extracting effective quality features. Specifically, the fusion map reflects the fusion ability of the left and right stereo-halves, while the difference map reveals the disparity information of a stereopair.

As discussed in Section 2, the gradient magnitude (GM) and Laplacian of Gaussian (LOG) features can be adopted to build the basic elements of image semantic structures, and they are hence closely related to the perceptual quality of natural images. The Gaussian derivative functions can model the receptive field responses of neurons along the visual pathway [41]. Therefore, we compute the GM and LOG maps using the first- and second-order derivatives of a circularly symmetric 2D Gaussian function G defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (1)$$

where x and y represent the horizontal and vertical directions, respectively. The parameter σ is the standard deviation. Then, we calculate the first-order partial derivative of $G(x, y, \sigma)$ with respect to x or y by

$$\frac{\partial G(x, y, \sigma)}{\partial d} = -\frac{1}{2\pi\sigma^2} \frac{d}{\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (2)$$

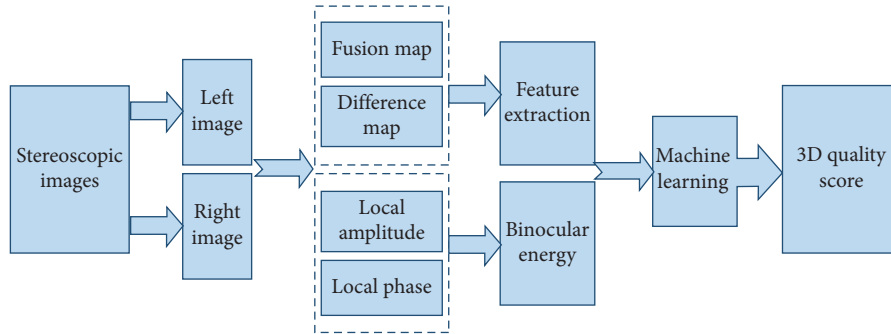


FIGURE 2: The architecture of our proposed quality assessment scheme for stereoscopic images.

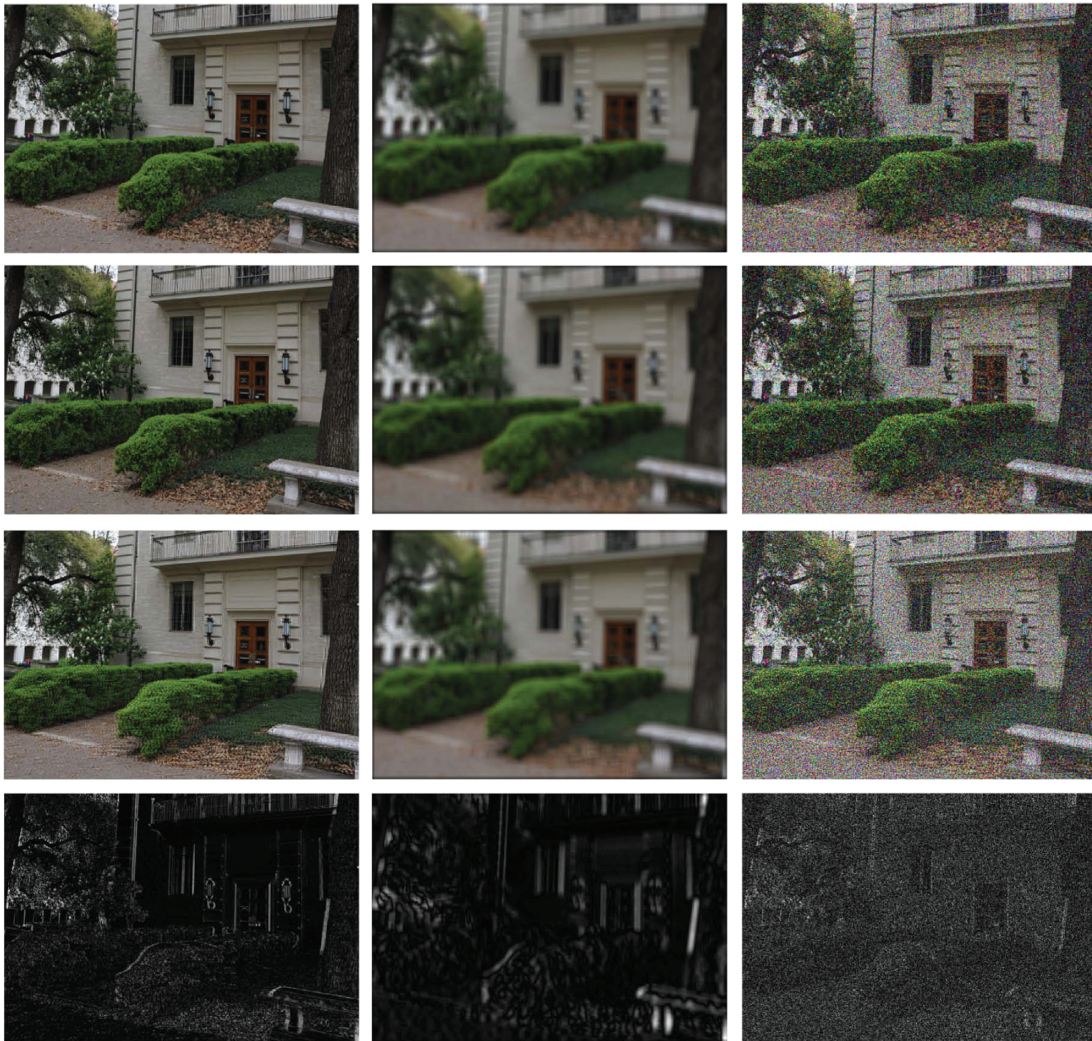


FIGURE 3: Examples of distorted stereoscopic images with different distortion types and the corresponding fusion and difference maps: the original stereoscopic image (left), distorted stereoscopic image with Gaussian blur (middle), and distorted stereoscopic image with white Gaussian noise (right). The first to the last rows show the left-view images, right-view images, fusion maps, and difference maps, respectively.

where $d \in \{x, y\}$ is the Gaussian partial derivative filter applied along the horizontal x or vertical y direction. An image is denoted by I ; thus, the GM map of the image I can be obtained by

$$GM_v = \sqrt{\left(I_v \otimes \frac{\partial G}{\partial x}\right)^2 + \left(I_v \otimes \frac{\partial G}{\partial y}\right)^2}, \quad (3)$$

where the symbol \otimes represents the convolution operation. $v \in \{l, r\}$, where l and r refer to the left and right views of a stereopair, respectively. Similarly, the LOG filter, corresponding to the second-order Gaussian partial derivative, is defined as follows:

$$\begin{aligned} h_{\text{LOG}}(x, y, \sigma) &= \frac{\partial^2 G(x, y, \sigma)}{\partial^2 x} + \frac{\partial^2 G(x, y, \sigma)}{\partial^2 y} \\ &= \frac{1}{2\pi\sigma^2} \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \end{aligned} \quad (4)$$

Accordingly, we estimate the LOG map of the left and right views by

$$\text{LOG}_v = I_v \otimes h_{\text{LOG}}. \quad (5)$$

Subsequently, a joint adaptive normalization procedure [39] is employed to normalize the GM and LOG coefficients for stable statistical image representations. Previous works [42, 43] have revealed that the overall quality of a distorted stereopair cannot be accurately calculated by directly averaging the qualities of the left- and right-view images, especially for an asymmetrically distorted 3D image. In practical application, to simulate the binocular rivalry (BR) phenomenon, the basis of weighting factors for the quality-predictive feature vectors of a stereopair can be defined as follows:

$$\begin{aligned} w_l &= \frac{(e_l)^\alpha}{[(e_l)^\alpha + (e_r)^\alpha]}, \\ w_r &= \frac{(e_r)^\alpha}{[(e_l)^\alpha + (e_r)^\alpha]}, \end{aligned} \quad (6)$$

where w_l and w_r denote the weights for the distorted left and right views, which can reflect the binocular contrast combination to a certain extent. e_l and e_r represent the local energy variances of the left and right images for a stereopair, respectively. The intensity adjusting parameter α is empirically set to 3 in the experiment. Therefore, the basic feature vectors of the gradient magnitude and Laplacian of Gaussian responses for a stereoscopic image can be calculated by

$$\begin{aligned} S_{\text{GM}} &= \sum_{x,y} w_l \text{GM}_l(x, y) + \sum_{x,y} w_r \text{GM}_r(x, y), \\ S_{\text{LOG}} &= \sum_{x,y} w_l \text{LOG}_l(x, y) + \sum_{x,y} w_r \text{LOG}_r(x, y), \end{aligned} \quad (7)$$

where $\text{GM}_{l/r}(x, y)$ and $\text{LOG}_{l/r}(x, y)$ denote the gradient magnitude and Laplacian of Gaussian for the left and right image, respectively. The features of GM and LOG responses are utilized to represent the visual semantic structures of the first-order and second-order binocular combination.

Finally, combined with the GM and LOG features of the fusion and difference maps, the binocular feature vectors used for further data training can be expressed by

$$V = [S_{\text{GM}}, S_{\text{LOG}}, F_{\text{GM}}, F_{\text{LOG}}, D_{\text{GM}}, D_{\text{LOG}}], \quad (8)$$

where $F_{\text{GM/LOG}}$ and $D_{\text{GM/LOG}}$ are the GM/LOG features of the fusion and difference maps, respectively.

3.2. Binocular Energy Response. The above extracted features are mainly utilized to indicate the visual sensitivity of binocular rivalry. Neurological research has reported that the human binocular vision phenomenon is a complicated process with combination of binocular rivalry, binocular fusion, and other factors [44]. The binocular fusion also contributes significantly to human visual perception besides the binocular rivalry. Previous research findings [45] indicated that the binocular energy responses play critical roles in representing binocular visual perception, especially for binocular fusion. In this paper, the binocular energy responses are obtained from the local magnitude and local phase of a stereopair.

In the proposed scheme, the left and right images of a stereopair are first processed using the log-Gabor filter. Here, we define $\alpha_{s,o}$ to represent the responses on different scales, where s is the spatial scale index. And we let $\beta_{s,o}$ denote the responses along different orientations, where o is the orientation scale index. The detailed description of this log-Gabor filter can be referred to the work in [46]. According to the given scale and orientation, the local amplitude at location x on scale s and along orientation o can be defined as

$$\text{LA}_{s,o}(x) = \sqrt{\alpha_{s,o}^2(x) + \beta_{s,o}^2(x)}. \quad (9)$$

With the sum of the local amplitudes on all the scales along the orientation o_m [46], the local amplitude can be calculated by

$$\text{LA}(x) = \sum_s \text{LA}_{s,o_m}(x), \quad (10)$$

where o_m is a parameter used to indicate the orientation with the maximum phase congruency value. Similar to the local amplitude, the local phase can be obtained by the angle along the orientation [46]:

$$\text{LP}(x) = \arctan\left(\frac{\sum_s \beta_{s,o_m}(x)}{\sum_s \alpha_{s,o_m}(x)}\right). \quad (11)$$

Based on previous works on binocular vision energy [16, 34], the left-view response and right-view response of a stereopair can be defined as follows:

$$\begin{aligned} C_l(x) &= \text{LA}_l(x) \cdot \exp(\text{LP}_l(x)), \\ C_r(x) &= \text{LA}_r(x) \cdot \exp(\text{LP}_r(x)). \end{aligned} \quad (12)$$

The right-view response $C_r(x)$ usually can be taken as a shifted transformation of the left-view response $C_l(x)$. The disparity d is defined as the difference between the locations of associated points in the left- and right-view responses. By considering a simple binocular cell with the left and right receptive fields, the binocular energy response E for a stereoscopic image pair can be calculated by

$$E = \sum_x \|C_l(x) + C_r(x+d)\|^2. \quad (13)$$

Finally, combined with the binocular statistical features, the overall quality-predictive features are $F = [V, E]$, which are fed into the following quality prediction for model learning. The weights among them are determined in the learning process.

3.3. Quality Prediction Model Learning. A number of training methods can be utilized to map the quality-predictive features of a stereopair to its corresponding subjective quality score, such as support vector regression (SVR) [47] and neural networks (NNs) [48]. SVR requires complex training algorithms and involves a quadratic programming problem. Neural networks have the difficulties of local minima, learning epochs, and slow convergence. An important question is that neural networks or training-based methods usually need large quantities of labeled training samples, while 3D image quality databases generally lack large-scale training images with subjective quality scores, which limits the performance of the methods using deep neural networks. In recent years, the extreme learning machine (ELM) [49] has attracted considerable attention and has been demonstrated as an effective and efficient technique in many applications, such as pattern recognition [50] and quality evaluation [51]. The ELM has advantages of faster learning speed, higher learning accuracy, and improved generalization. The weights between the input and hidden layers can be selected randomly and independent of the training data, and layer-by-layer back propagated tuning is not required [35]. Motivated by these unique properties, we try to employ the ELM for feature mapping and regression model learning in 3D image quality prediction.

For a given set of N arbitrary training samples (F_i, y_i) , where F_i represents the quality-predictive features of the i th pair of original/distorted images and y_i is the corresponding subjective quality score, our goal is to find a function which minimizes the deviation from the subjective quality score for all the training data. The function $f(F_i)$ with L hidden nodes can be mathematically modeled and expressed by

$$f(F_i) = \sum_{j=1}^L \beta_j \cdot g_j(F_i) = g(F_i) \cdot \beta, \quad i = 1, \dots, N, \quad (14)$$

where $g(F_i) = [g_1(F_i), \dots, g_L(F_i)]$ is the output vector of the hidden neuron and $\beta = [\beta_1, \dots, \beta_L]^T$ denotes the output weighting vector between the output node and the hidden layer of L nodes. The activation function $g(F_i)$ can approximate N training samples by minimizing the training error and can be formulated as

$$g_j(F_i) = g(w_j \cdot F_i + b_j), \quad (15)$$

where w_j is the weighing vector which connects the input layer and the j th hidden node and b_j denotes the corresponding threshold of the hidden node. In equation (14), β is the only parameter to be determined, which leads to fast learning for ELM [49]. For N training samples (F_i, y_i) , the

mathematical model for ELM (equation (14)) can be described as follows:

$$\mathbf{Y}_H \beta = \mathbf{Y}, \quad (16)$$

where \mathbf{Y} represents the target vector and \mathbf{Y}_H is called the hidden layer output matrix, which can be defined as

$$\mathbf{Y}_H = \begin{Bmatrix} g(w_1 \cdot F_1 + b_1) & \dots & g(w_L \cdot F_1 + b_L) \\ \vdots & & \vdots \\ g(w_1 \cdot F_N + b_1) & \dots & g(w_L \cdot F_N + b_L) \end{Bmatrix},$$

$$\beta = \begin{Bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{Bmatrix}, \quad (17)$$

$$\mathbf{Y} = \begin{Bmatrix} y_1 \\ \vdots \\ y_N \end{Bmatrix}.$$

The minimal norm least-squares method is used in ELM to minimize the norm of the output weights. Then, the vector of the output weights β can be predicted analytically and expressed by

$$\beta = \mathbf{Y} \mathbf{Y}_H^\dagger, \quad (18)$$

where \mathbf{Y}_H^\dagger denotes the Moore–Penrose (MP) generalized pseudoinverse of the hidden layer output matrix \mathbf{Y}_H . In practice, the orthogonal projection method [49] can be efficiently employed to calculate the Moore–Penrose inverse:

$$\mathbf{Y}_H^\dagger = \begin{cases} (\mathbf{Y}_H^T \mathbf{Y}_H)^{-1} \mathbf{Y}_H^T, & \text{if } \mathbf{Y}_H^T \mathbf{Y}_H \text{ is nonsingular,} \\ \mathbf{Y}_H^T (\mathbf{Y}_H \mathbf{Y}_H^T)^{-1}, & \text{if } \mathbf{Y}_H \mathbf{Y}_H^T \text{ is nonsingular.} \end{cases} \quad (19)$$

Based on the ridge regression theory, a positive value $1/\lambda$ is added to the diagonal of $\mathbf{Y}_H^T \mathbf{Y}_H$ or $\mathbf{Y}_H \mathbf{Y}_H^T$, which makes the solution more stable. Therefore, with this positive value $1/\lambda$, we can obtain

$$\beta = \begin{cases} \mathbf{Y}_H^T \left(\frac{I}{\lambda} + \mathbf{Y}_H \mathbf{Y}_H^T \right)^{-1} \mathbf{Y}, & \text{if } N \leq L, \\ \left(\frac{I}{\lambda} + \mathbf{Y}_H^T \mathbf{Y}_H \right)^{-1} \mathbf{Y}_H^T \mathbf{Y}, & \text{if } N > L, \end{cases} \quad (20)$$

where N represents the number of training samples and L denotes the number of hidden nodes. In this paper, the number of nodes L is selected to be equal to the number of training samples N . As a result, the output weight vector β is determined as $\mathbf{Y}_H^T ((I/\lambda) + \mathbf{Y}_H \mathbf{Y}_H^T)^{-1} \mathbf{Y}$ in the experiments. More details on the ELM can be found in [49].

4. Experimental Results and Analysis

In the experiments, we first describe the databases and criteria used for quality assessment. Then, we give the performance comparison with other related algorithms in terms of predicting the quality of distorted stereoscopic

images. Moreover, we show the evaluation results on individual distortion type. In addition, we investigate the effect of each component in the proposed metric. Finally, we perform the cross-database evaluation and analyze the time complexity in our experiments.

4.1. Experimental Databases and Protocols. In order to verify and compare the performance of our proposed quality assessment metric, three public and subject-rated benchmark 3D image databases were used as standards: LIVE 3D IQA database Phase I [52], LIVE 3D IQA database Phase II [20], and MCL-3D database [53].

- (1) LIVE 3D IQA database Phase I [52]: phase I contains 20 reference stereopairs and 365 symmetrically distorted stereopairs corresponding to five distortion types: JPEG compression, JPEG2000 (JP2K) compression, additive white noise (WN), Gaussian blur (GB), and a simulated fast-fading (FF) model. Each distorted stimulus has been evaluated by human observers and assigned a difference mean opinion score (DMOS) value. The lower DMOS values represent higher visual quality.
- (2) LIVE 3D IQA database Phase II [20]: phase II has 120 symmetrically distorted stimuli and 240 asymmetrically distorted stimuli generated from 8 pristine stereopairs. Each of the five distortion types (JPEG, JP2K, WN, GB, and FF) is symmetrically and asymmetrically applied to the pristine stereopairs at various degradation levels. The corresponding DMOS values are also given for the distorted stereopairs.
- (3) The MCL-3D database [53]: this database consists of 684 stereoscopic image pairs. Nine image-plus-depth sources are selected, and then a depth-image-based rendering technique is used to render 3D images. Four levels of distortions are applied to either the depth map or texture stereoscopic image prior to 3D image rendering. The distortion types are JPEG, JP2K, WN, Gaussian blur (GBLUR), downsampling blur (SBLUR), and transmission error (TERROR). Each distorted stimulus has been scored by human observers, and a pairwise comparison is used to obtain reliable mean opinion score (MOS) values.

To benchmark the performance of quality assessment metrics, three general performance indicators were employed to provide quantitative performance evaluations: (1) Pearson’s linear correlation coefficient (PLCC), which measures the linear dependence between the predicted quality scores and the ground truth targets, (2) Spearman’s rank-order correlation coefficient (SRCC), which serves as a measure of prediction monotonicity, and (3) Kendall’s rank-order correlation coefficient (KRCC), which is a nonparametric rank-order-based correlation metric. Higher values of PLCC, SRCC, and KRCC represent good consistency with human perceptual quality ratings. For the nonlinear regression, a five-parameter logistic function [54] was applied

to fit the predicted quality scores and provided quality scores.

In the experiments, we randomly split each database into two nonoverlapping subsets: a training subset and a test subset. A training process was required to calibrate the quality prediction model. In each train-test procedure, 80% of the database content was selected for training and the remaining 20% for test. After learning the statistical regression model using the training set, the quality prediction performance was evaluated using the test set. In specific, to avoid potential performance bias of the proposed scheme, the train-test iteration was repeated 1000 times, and the median values of PLCC, SRCC, and KRCC were chosen as the final validation results for performance evaluation. In the implementation, a unipolar sigmoidal function ($1/(1 + e^{-\lambda u})$) with $\lambda = 0.1$ was used as the ELM nonlinear activation function.

4.2. Overall Performance Comparison. To comprehensively investigate the effectiveness and robustness of the proposed scheme, we have conducted several different experiments to compare our scheme with the typical and representative methods. These mainly include two 2D-IQA methods (PSNR and multiscale structural similarity (MS-SSIM) [55]), two FR 3D-IQA methods (Benoit et al.’s method [13] and Chen et al.’s method [8]), and three NR 3D-IQA methods (Zhou and Yu’s method [56], Fan et al.’s method [41], and Shen et al.’s method [22]). For the previous two 2D-IQA approaches, the predicted quality score of a stereoscopic image was obtained by averaging the left and right image qualities. For Benoit et al.’s approach [13], the disparity distortion was the global correlation between the original and distorted disparity maps. For Chen et al.’s approach [8], we adopted the cyclopean metric in terms of multiscale SSIM described in their paper.

Figure 4 provides the scatter plots of predicted quality scores against subjective DMOS values for the proposed scheme and other compared methods on the LIVE 3D IQA database Phase I. In these figures, the horizontal axis represents the predicted quality scores and the vertical axis denotes the subjective DMOS values of the perceived distortions. Considering performance comparison, a straight-lined distribution of scatter points is better than other arbitrary shapes. For the PSNR and MS-SSIM [55] approaches, the performance is worse than most of other methods in general. The reason can be attributed to that these methods treat the left- and right-view images independently and binocular visual characteristics are not taken into account. For Benoit et al.’s approach [13], the quality evaluation accuracy is even lower than the 2D-IQA approaches under some distortions. One possible explanation is that the 2D image quality metric for disparity maps does not coincide with the human perception of disparity. Overall, the proposed scheme has better consistent alignment with human subjective judgements for stereoscopic 3D images on the database.

In order to further evaluate the performance comparison of quality assessment accuracy on the three databases, we

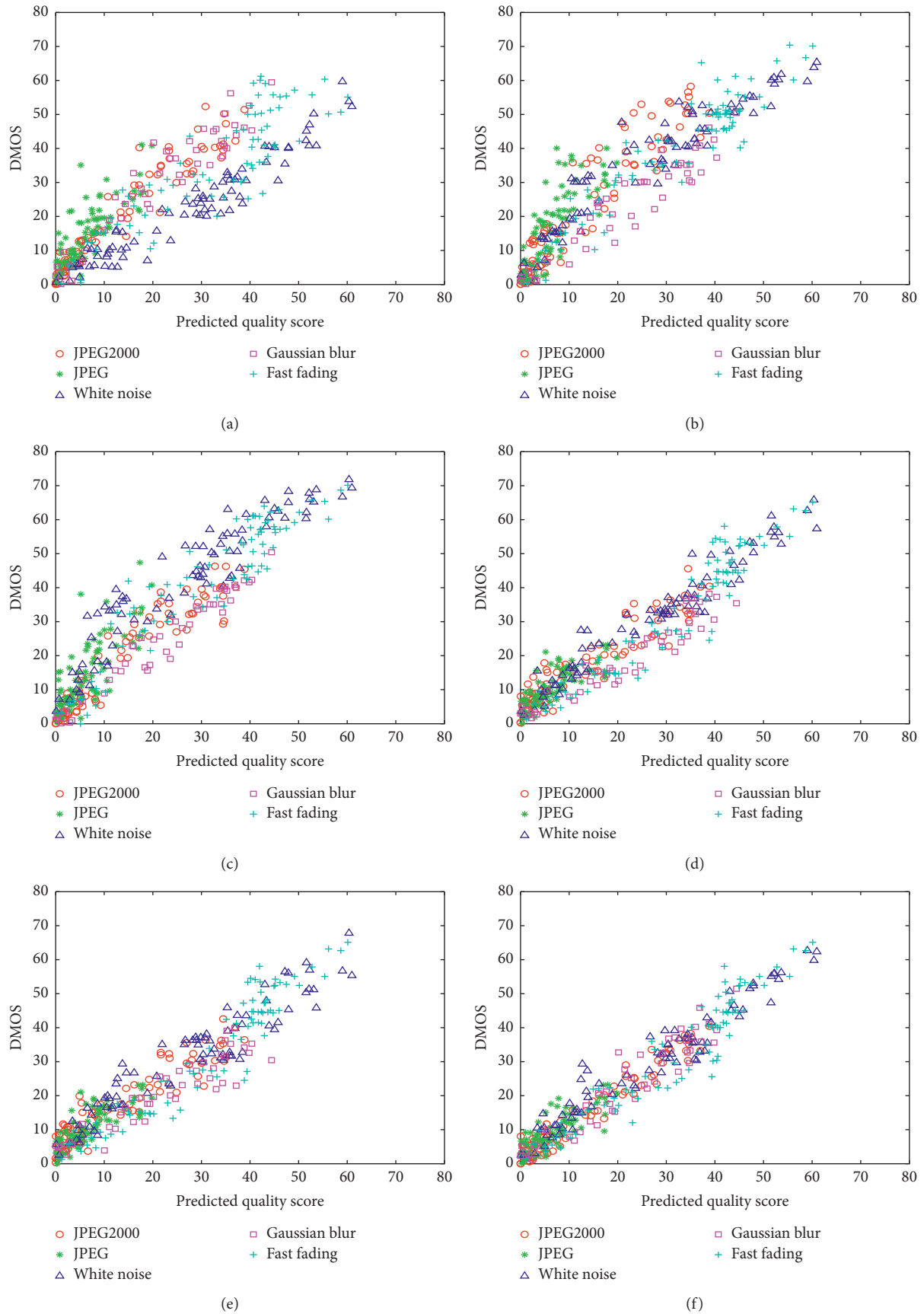


FIGURE 4: Continued.

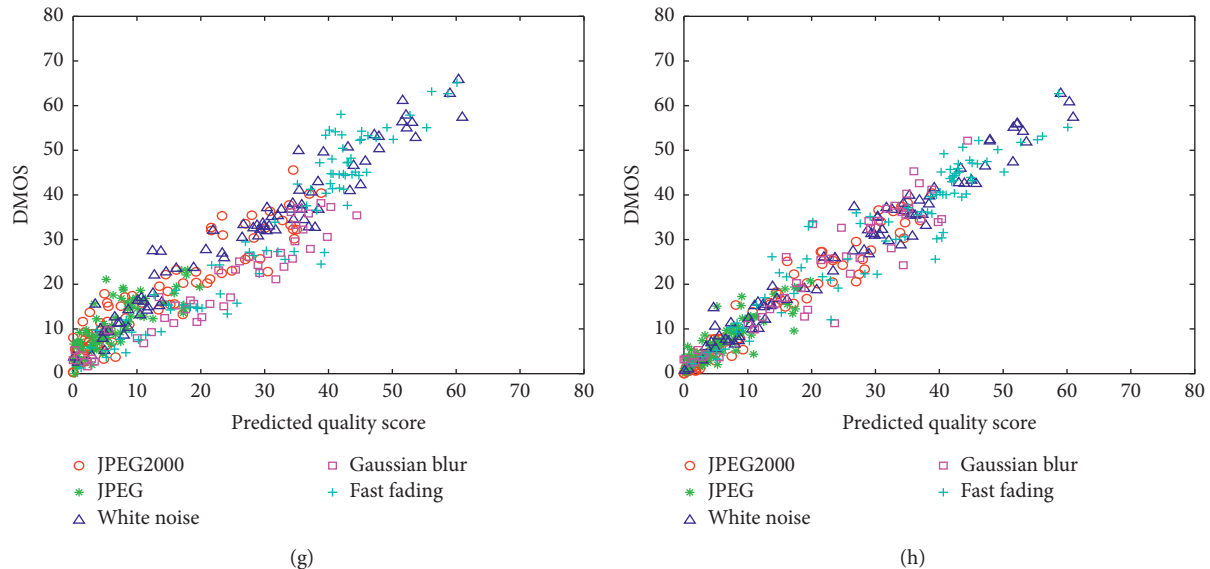


FIGURE 4: Scatter plots of the predicted quality scores against the subjective DMOS values for the eight methods. (a) PSNR. (b) MS-SSIM [55]. (c) Benoit et al.’s method [13]. (d) Chen et al.’s method [8]. (e) Zhou and Yu’s method [56]. (f) Fan et al.’s method [41]. (g) Shen et al.’s method [22]. (h) Proposed scheme.

have given the values of PLCC, SRCC, and KRCC between the provided and predicted quality scores for the proposed scheme and the compared methods. Table 1 presents the performance comparison results in terms of PLCC, SRCC, and KRCC on the three databases. In each case, the results of the best-performance metric are marked in bold. According to the experimental results in this table, Shen et al.’s method [22] performs best on asymmetrically distorted stereoscopic images in the LIVE 3D IQA database Phase II, and our proposed scheme achieves higher consistency with human opinion scores on the other two databases. Moreover, the PLCC and SRCC values for our scheme are above 0.912 and 0.907, respectively, on all databases, which demonstrate that the proposed scheme exhibits a good stability to quantify and predict the perceptual distortions of 3D images. On the whole, the proposed scheme has competitive performance and shows statistically superiority over other typical and representative methods for 3D image quality prediction.

4.3. Distortion-Specific Performance Evaluation. In this section, we have investigated the distortion-specific performance of the proposed scheme and other compared methods for each individual distortion type on the hybrid distortion databases. The PLCC, SRCC, and KRCC comparison results are summarized in Tables 2–4, respectively. For reasons of space and for brevity, $M[n]$ is used to represent the corresponding compared method proposed in paper [n]. The top two quality assessment metrics for each index (PLCC, SRCC, or KRCC) have been highlighted in bold. From these tables, we can find that our proposed metric achieves the highest hit-count for each index and is statistically superior to the compared methods. Some metrics may have high assessment accuracies for specific distortion types: Chen et al.’s method [8] shows strong

competitiveness on Gaussian blur, and Shen et al.’s method [22] has outstanding performance on JPEG compression. But our method is comparable to the best-performing metrics for these kinds of distortions. The proposed scheme generally outperforms the vast majority of compared methods by a certain margin for distortion-specific performance evaluation. From these experimental results, it is worth noting that the quality prediction of our scheme is basically independent of different sorts of distortions.

4.4. Contribution of Each Component in the Proposed Scheme. In this section, to understand the respective contributions of each component to the overall quality score in the proposed metric, we have devised three different schemes for comparison, denoted by scheme A, scheme B, and scheme C, respectively. For scheme A, the binocular features of GM response and the binocular energy were used to measure visual quality. For scheme B, the binocular features of LOG response and the binocular energy were adopted for quality prediction. For scheme C, the binocular energy was not included, and the binocular features of GM and LOG responses were considered for quality evaluation. The PLCC, SRCC, and KRCC results are reported in Table 5. As can be observed from this table, the binocular features of LOG response have the most important impact on quality prediction under distortions. It can be inferred that the binocular features of GM and LOG responses and the binocular energy are complementary, and only adopting one aspect of these features cannot obtain the best performance. In addition, according to the results in the table, scheme B has higher assessment accuracies than scheme A, which implies that the LOG features contain more useful visual information and contribute more to 3D quality prediction than the GM features. The experimental results have also

TABLE 1: Performance of the proposed scheme and other methods in terms of PLCC, SRCC, and KRCC, using the three databases.

IQA model	LIVE-I			LIVE-II			MCL-3D		
	PL	SR	KR	PL	SR	KR	PL	SR	KR
PSNR	0.852	0.846	0.752	0.809	0.790	0.685	0.837	0.816	0.741
MS-SSIM [55]	0.876	0.872	0.815	0.831	0.839	0.727	0.886	0.891	0.759
Benoit [13]	0.883	0.855	0.776	0.835	0.794	0.715	0.843	0.819	0.746
Chen [8]	0.917	0.916	0.737	0.900	0.889	0.730	0.881	0.884	0.717
Chen [17]	0.939	0.923	0.830	0.865	0.859	0.806	0.903	0.892	0.839
Zhou [56]	0.910	0.901	0.711	0.771	0.770	0.550	0.833	0.831	0.612
Fan [41]	0.928	0.887	0.724	0.861	0.823	0.627	0.865	0.830	0.751
Chen [34]	0.937	0.920	0.826	0.870	0.852	0.729	0.907	0.893	0.765
Shen [22]	0.932	0.925	0.839	0.928	0.920	0.856	0.930	0.922	0.835
Proposed	0.941	0.927	0.861	0.912	0.907	0.858	0.937	0.926	0.852

PL: PLCC, SR: SRCC, KR: KRCC. The results of the best-performance metric are marked in bold.

TABLE 2: PLCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.521	0.735	0.633	0.837	0.715	0.825	0.840	0.865
	JP2K	0.826	0.917	0.930	0.916	0.893	0.909	0.912	0.873
	WN	0.895	0.923	0.919	0.943	0.916	0.947	0.935	0.952
	GBLUR	0.913	0.936	0.924	0.951	0.912	0.950	0.942	0.958
	FF	0.740	0.861	0.893	0.915	0.878	0.882	0.918	0.916
LIVE-II	JPEG	0.726	0.865	0.694	0.820	0.770	0.692	0.871	0.863
	JP2K	0.783	0.868	0.826	0.872	0.611	0.735	0.904	0.855
	WN	0.905	0.940	0.743	0.902	0.796	0.904	0.920	0.927
	GBLUR	0.817	0.793	0.896	0.965	0.872	0.913	0.943	0.946
	FF	0.802	0.816	0.851	0.909	0.904	0.851	0.929	0.932
MCL-3D	JPEG	0.651	0.870	0.674	0.861	0.805	0.856	0.859	0.864
	JP2K	0.782	0.879	0.906	0.913	0.872	0.910	0.909	0.885
	WN	0.886	0.955	0.792	0.892	0.856	0.866	0.935	0.939
	GBLUR	0.867	0.914	0.839	0.867	0.757	0.915	0.951	0.954
	SBLUR	0.829	0.892	0.815	0.907	0.819	0.924	0.926	0.920
TERROR	0.813	0.867	0.790	0.905	0.743	0.769	0.893	0.927	

The top two quality assessment metrics have been highlighted in bold.

demonstrated that the quality assessment performance can be promoted by appropriate combination of binocular features and binocular energy.

4.5. Cross-Database Performance Evaluation. In the above experiments, the training and test subset have the same distortions selected from the databases. Since the proposed scheme is based on a learning framework, it is necessary to ascertain whether the performance is bound to a special training database on which it is trained. To verify the generalization ability and stability of our scheme, we have carried out cross-database experiments for performance evaluation. In the experiment, we examined whether satisfactory results could be obtained by applying the regression model trained on one database to the testing set from another database. For brevity, the SRCC results of cross-database performance evaluation are given in Table 6. It can be observed that the proposed metric has comparatively weak performance in comparison with the evaluation results in Table 3. The reason can be mainly attributed to that the training and test subsets have different types of distortions.

For instance, the LIVE Phase I database only has symmetrically distorted images, while the LIVE Phase II database contains both symmetric and asymmetric distortions. However, the values of corresponding indicators are still relatively high, which show that our framework can maintain a satisfying predictive capacity under different circumstances. Based on the above experimental results, it can be concluded that a larger training database with more comprehensive distortion types could probably promote the prediction accuracy of our scheme.

In this section, we have also compared the cross-database evaluation of the proposed scheme with other related methods. The SRCC results are provided in Table 7, where the top two metrics have been marked in bold. According to the experimental results in the table, no matter which training database is used, the cross-database evaluation of our scheme is usually stable, and it offers statistically better quality prediction in line with human perception than the compared methods on different databases. These facts demonstrate the generalization ability and effectiveness of the proposed scheme for stereoscopic image quality assessment.

TABLE 3: SRCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.507	0.725	0.652	0.816	0.569	0.614	0.845	0.842
	JP2K	0.818	0.906	0.897	0.877	0.812	0.824	0.909	0.853
	WN	0.902	0.925	0.912	0.895	0.940	0.915	0.937	0.942
	GBLUR	0.895	0.933	0.874	0.953	0.860	0.916	0.938	0.963
	FF	0.754	0.856	0.860	0.912	0.784	0.867	0.910	0.924
LIVE-II	JPEG	0.696	0.852	0.704	0.822	0.769	0.593	0.843	0.847
	JP2K	0.785	0.847	0.819	0.863	0.593	0.717	0.886	0.829
	WN	0.873	0.934	0.738	0.889	0.846	0.891	0.936	0.939
	GBLUR	0.809	0.803	0.904	0.956	0.862	0.903	0.951	0.932
	FF	0.806	0.821	0.836	0.895	0.935	0.891	0.901	0.909
MCL-3D	JPEG	0.645	0.861	0.719	0.857	0.781	0.848	0.863	0.868
	JP2K	0.809	0.873	0.843	0.902	0.875	0.927	0.908	0.897
	WN	0.891	0.942	0.758	0.899	0.823	0.861	0.912	0.932
	GBLUR	0.864	0.906	0.827	0.893	0.705	0.910	0.926	0.945
	SBLUR	0.832	0.880	0.796	0.920	0.818	0.951	0.933	0.904
	TERROR	0.815	0.852	0.807	0.903	0.609	0.676	0.895	0.907

The top two quality assessment metrics have been highlighted in bold.

TABLE 4: KRCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.434	0.692	0.558	0.736	0.535	0.601	0.757	0.826
	JP2K	0.685	0.826	0.804	0.727	0.793	0.706	0.859	0.803
	WN	0.823	0.829	0.811	0.736	0.912	0.835	0.820	0.915
	GBLUR	0.806	0.850	0.786	0.863	0.839	0.827	0.858	0.931
	FF	0.638	0.729	0.752	0.856	0.706	0.772	0.863	0.882
LIVE-II	JPEG	0.567	0.735	0.591	0.727	0.522	0.571	0.749	0.736
	JP2K	0.632	0.783	0.754	0.726	0.537	0.765	0.736	0.725
	WN	0.729	0.821	0.695	0.715	0.696	0.619	0.833	0.839
	GBLUR	0.656	0.806	0.827	0.862	0.729	0.820	0.881	0.894
	FF	0.694	0.715	0.709	0.739	0.853	0.783	0.872	0.851
MCL-3D	JPEG	0.656	0.719	0.682	0.709	0.607	0.552	0.753	0.772
	JP2K	0.727	0.753	0.817	0.763	0.715	0.863	0.827	0.846
	WN	0.825	0.851	0.786	0.685	0.616	0.639	0.849	0.895
	GBLUR	0.796	0.764	0.624	0.690	0.596	0.797	0.886	0.904
	SBLUR	0.809	0.766	0.732	0.817	0.775	0.812	0.825	0.813
	TERROR	0.751	0.705	0.724	0.761	0.579	0.537	0.758	0.819

The top two quality assessment metrics have been highlighted in bold.

TABLE 5: Performance of each component in the proposed scheme.

	LIVE-I			LIVE-II			MCL-3D		
	PL	SR	KR	PL	SR	KR	PL	SR	KR
Scheme A	0.832	0.754	0.681	0.724	0.658	0.593	0.761	0.696	0.605
Scheme B	0.879	0.826	0.735	0.803	0.717	0.642	0.829	0.740	0.652
Scheme C	0.851	0.797	0.714	0.750	0.664	0.607	0.785	0.728	0.637
Proposed	0.941	0.927	0.861	0.912	0.907	0.858	0.937	0.926	0.852

PL: PLCC, SR: SRCC, KR: KRCC.

4.6. Time Complexity Analysis. Time complexity is a significant indicator in evaluating the performance of the proposed scheme, to facilitate its use in real-time applications such as monitoring and adjustment. We have compared the computational complexity of our proposed scheme with other related methods. The experiment was performed in MATLAB R2014a on a Windows 10 PC with a

2.5 GHz Intel Core i7 processor and 8 GB RAM. The results of time consumption are given in Table 8, which presents the running time comparison on the LIVE Phase I database with 365 stereopairs. As can be seen from this table, the total processing time of the proposed scheme is 127 seconds, which indicates that it takes less than 0.35 seconds to predict a distorted stereopair. Although it is not the most efficient

TABLE 6: Cross-database performance evaluation in terms of SRCC.

Testing database	Training database	JPEG	JP2K	WN	GBLUR	FF	All
LIVE-I	LIVE-II	0.834	0.833	0.896	0.925	0.908	0.904
	MCL-3D	0.821	0.839	0.908	0.885	—	0.890
LIVE-II	LIVE-I	0.815	0.801	0.892	0.907	0.903	0.840
	MCL-3D	0.833	0.754	0.898	0.902	—	0.879
MCL-3D	LIVE-I	0.819	0.868	0.917	0.902	—	0.875
	LIVE-II	0.861	0.870	0.906	0.922	—	0.905

TABLE 7: SRCC comparison with other related methods for cross-database performance evaluation.

Testing	Training	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	LIVE-II	0.792	0.839	0.817	0.896	0.871	0.866	0.901	0.904
	MCL-3D	0.808	0.865	0.853	0.913	0.897	0.885	0.895	0.890
LIVE-II	LIVE-I	0.729	0.835	0.762	0.861	0.765	0.816	0.857	0.840
	MCL-3D	0.746	0.811	0.783	0.877	0.752	0.820	0.883	0.879
MCL-3D	LIVE-I	0.747	0.861	0.710	0.853	0.812	0.825	0.859	0.875
	LIVE-II	0.803	0.823	0.815	0.817	0.819	0.828	0.868	0.905

The top two metrics have been marked in bold.

TABLE 8: Performance comparison of time complexity on the LIVE Phase I database.

Metrics	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
Run time (s)	7105	9520	6814	5143	37	453	396	127

method, it has the best comprehensive performance in achieving the balance between accuracy and timeliness. The simulation results demonstrate that our proposed scheme has relatively lower computing complexity than the compared methods.

5. Conclusions

In this paper, we have presented a novel no-reference quality prediction method for stereoscopic images based on binocular statistical features and machine learning. The framework of the proposed scheme includes a feature extraction stage and a feature mapping stage. The gradient magnitude and Laplacian of Gaussian responses from a stereopair and its fusion and difference maps are utilized as quality-predictive features. With the extreme learning machine, a statistical regression model is established to map these binocular features of a stereopair to its corresponding perceptual quality score. The visual quality predictions by the proposed metric are highly correlated with subjective quality judgements for distorted image pairs of various distortion types. More importantly, our method achieves excellent performance and has a promising generalization ability. The proposed scheme can be applied to video broadcasting and 3D multimedia industry for its practicality.

For future work, how to explore deeper structure representation for a human visual system and how to design more efficient machine learning methods for visual quality prediction should be researched. In addition, more effective quality features can be considered to simulate the human perceptual vision. Other 3D quality factors such as depth perception and visual comfort still deserve further study.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported in part by the Key Research and Development Program of China under Grant no. 2018YFC0831000, the National Natural Science Foundation of China under Grant no. 62001267, the Natural Science Foundation of Shandong Province under Grant no. ZR2020QF013, the Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) under Grant no. 2019JZZY010119, and the Fundamental Research Funds of Shandong University under no. 2020HW017.

References

- [1] L. Chen and J. Zhao, "Quality assessment of stereoscopic 3D images based on local and global visual characteristics," in *Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 61–66, Hong Kong, China, July 2017.
- [2] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: visual comfort," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 335–346, 2011.

- [3] S. Khan and S. S. Channappayya, "Estimating depth-salient edges and its application to stereoscopic image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5892–5903, 2018.
- [4] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya, and A. C. Bovik, "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5027–5040, 2019.
- [5] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.
- [6] X. Ben, C. Gong, P. Zhang, X. Jia, Q. Wu, and W. Meng, "Coupled patch alignment for matching cross-view gaits," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3142–3157, 2019.
- [7] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
- [8] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.
- [9] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.
- [10] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
- [11] 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 Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2019.
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [13] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proceedings of the 15th IEEE International Conference on Image Processing*, pp. 389–392, San Diego, CA, USA, October 2008.
- [14] J. You, L. Xing, A. Perki, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proceedings of the Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM2010)*, pp. 61–66, Scottsdale, AZ, USA, November 2010.
- [15] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1527–1542, 2014.
- [16] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1940–1953, 2013.
- [17] L. Chen and J. Zhao, "Perceptual quality assessment of stereoscopic images based on local and global visual characteristics," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 12139–12156, 2019.
- [18] X. Ben, P. Zhang, Z. Lai, R. Yan, X. Zhai, and W. Meng, "A general tensor representation framework for cross-view gait recognition," *Pattern Recognition*, vol. 90, pp. 87–98, 2019.
- [19] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, and W. Meng, "Coupled bilinear discriminant projection for cross-view gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 734–747, 2020.
- [20] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.
- [21] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 591–602, 2014.
- [22] L. Shen, R. Fang, Y. Yao, X. Geng, and D. Wu, "No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 1, pp. 59–72, 2019.
- [23] L. Chen and J. Zhao, "No-reference quality assessment for stereoscopic 3D images based on binocular visual perception," in *Proceedings of the IEEE International Symposium on Haptic, Audio and Visual Environments and Games*, pp. 1–5, Dalian, China, September 2018.
- [24] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 341–353, 2018.
- [25] Z. Chen, J. Xu, C. Lin, and W. Zhou, "Stereoscopic omnidirectional image quality assessment based on predictive coding theory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 103–117, 2020.
- [26] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.
- [27] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.
- [28] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Yokohama, Japan, August 2020.
- [29] W. Zhou, W. Qiu, and M.-W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 404–415, 2017.
- [30] S. Li, X. Han, and Y. Chang, "Adaptive cyclopean image-based stereoscopic image-quality assessment using ensemble learning," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2616–2624, 2019.
- [31] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1733–1740, Columbus, OH, USA, June 2014.
- [32] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4923–4936, 2017.
- [33] W. Zhou, Z. Chen, and W. Li, "Dual-stream interactive networks for no-reference stereoscopic image quality

- assessment,” *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3946–3958, 2019.
- [34] L. Chen and J. Zhao, “No-reference perceptual quality assessment of stereoscopic images based on binocular visual characteristics,” *Signal Processing: Image Communication*, vol. 76, pp. 1–10, 2019.
- [35] 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 Recognition*, vol. 71, pp. 207–217, 2017.
- [36] D. Stidwill and R. Fletcher, *Normal Binocular Vision: Theory, Investigation and Partial Aspects*, Wiley, New York, NY, USA.
- [37] N. Kruger, P. Janssen, S. Kalkan et al., “Deep hierarchies in the primate visual cortex: what can we learn for computer vision?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.
- [38] J. Wu, W. Lin, and G. Shi, “Image quality assessment with degradation on spatial structure,” *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 437–440, 2014.
- [39] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, “Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features,” *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.
- [40] K. A. May and L. Zhaoping, “Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns,” *Current Biology*, vol. 26, no. 12, pp. 1571–1576, 2016.
- [41] Y. Fan, M. C. Larabi, F. A. Cheikh, and C. F. Maloigne, “No-reference quality assessment of stereoscopic images based on binocular combination of local features statistics,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 3538–3542, Athens, Greece, October 2018.
- [42] P. Seuntjens, L. Meesters, and W. Ijsselstein, “Perceived quality of compressed stereoscopic images,” *ACM Transactions on Applied Perception*, vol. 3, no. 2, pp. 95–109, 2006.
- [43] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, “Quality prediction of asymmetrically distorted stereoscopic 3D images,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.
- [44] K. Lee and S. Lee, “3D perception based quality pooling: stereopsis, binocular rivalry, and binocular suppression,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 533–545, 2015.
- [45] Q. Peng and B. E. Shi, “The changing disparity energy model,” *Vision Research*, vol. 50, no. 2, pp. 181–192, 2010.
- [46] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: a feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [47] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [48] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [49] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [50] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 42, no. 2, pp. 513–529, 2012.
- [51] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, “NMF-based image quality assessment using extreme learning machine,” *IEEE Transactions on Cybernetics*, vol. 47, pp. 232–243, 2017.
- [52] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, “Subjective evaluation of stereoscopic image quality,” *Signal Process: Image Communication*, vol. 28, no. 8, pp. 870–883, 2012.
- [53] R. Song, H. Ko, and C. C. J. Kuo, “MCL-3D: a database for stereoscopic image quality assessment using 2D-image-plus-depth source,” *Journal of Information Science and Engineering*, vol. 31, no. 5, pp. 1593–1611, 2015.
- [54] P. G. Gottschalk and J. R. Dunn, “The five-parameter logistic: a characterization and comparison with the four-parameter logistic,” *Analytical Biochemistry*, vol. 343, no. 1, pp. 54–65, 2005.
- [55] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” in *Proceedings of the IEEE The Thirtieth Asilomar Conference on Signals, Systems & Computers*, pp. 1398–1402, Pacific Grove, CA, USA, November 2003.
- [56] W. Zhou and L. Yu, “Binocular responses for no-reference 3D image quality assessment,” *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1077–1084, 2016.