IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 13, 2021, accepted November 18, 2021, date of publication November 22, 2021,
date of current version December 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3129814

Reduced-Reference Stereoscopic Image Quality Assessment Using Gradient Sparse Representation and Structural Degradation

JIAN MA[1,2], GUOMING XU[1], AND XIYU HAN[1]

[1]School of Internet, Anhui University, Hefei 230039, China
[2]School of Computer Science, Fudan University, Shanghai 200433, China

Corresponding author: Jian Ma (jianma@ahu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61906118, and in part by the Anhui Natural Science Foundation under Grant 2108085MF230.

ABSTRACT Reduced-reference stereoscopic image quality assessment (RRSIQA) models evaluate stereoscopic image quality degradation with partial information about the ''ideal-quality'' reference stereopair. On one hand, sparse representation in recent theoretical studies of visual cognition has been proved to resemble the strategy used to represent natural images in the primary visual cortex. On the other hand, the joint statistics of gradient magnitude (GM) and Laplacian of Gaussian (LOG) features are popularly utilized to form image semantic structures. Motivated by these findings, we present a new RRSIQA metric using gradient sparse representation and structural degradation in this paper. Concretely, the proposed metric is based on two main tasks: the first task extracts the distribution statistics of visual primitives by gradient sparse representation, while the second task measures structural degradation of stereoscopic image due to the presence of distortion by extracting the joint statistics of GM and LOG features. The former, so-called the binocular perceptual visual information (PVI), aims to effectively integrates the gradient map that is sparser than the image itself. Especially, the process of binocular fusion is simulated by using the mutual information of the gradient-based visual primitives between left and right view's images as binocular cue. Furthermore, the perceptual loss vectors are taken as the differences of binocular perceptual visual information and structural degradation between reference and distorted stereopairs. Finally, the perceptual loss vectors are utilized to calculate the quality score by a prediction function which is trained using kernel ridge regressing (KRR). The experiments are performed on the popular LIVE 3D IQA databases and Waterloo IVC 3D databases, and experimental results show highly competitive performance with the state-of-the-art algorithms. Moreover, in some challenging cases with particular asymmetric distortion types, the proposed metric can achieves the best quality prediction accuracy in LIVE 3D phase II and Waterloo IVC 3D Phase II.

INDEX TERMS Reduced-reference, stereoscopic image quality assessment, sparse representation, structural degradation, kernel ridge regressing.

I. INTRODUCTION

DURING the past two decades, various three-dimensional (3D) technologies (such as 3D image coding, reconstruction, enhancement, and monitoring, etc.) have advanced rapidly and drastically changed the way people viewed their world. However, since the current 3D technologies are still immature, various levels and types of distortion will inevitably be introduced into 3D content which may give rise to a degradation of 3D visual quality. For this reason, it is an urgent demand to establish an effective 3D content quality evaluation method. In general, the most direct and reliable method to estimate image quality is by subjective assessment. However, the subjective metrics are regarded as inconvenient, expensive, and time consuming [1]. These drawbacks provide the motivation for developing efficient and fast objective stereoscopic image quality assessment (SIQA) metrics.

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

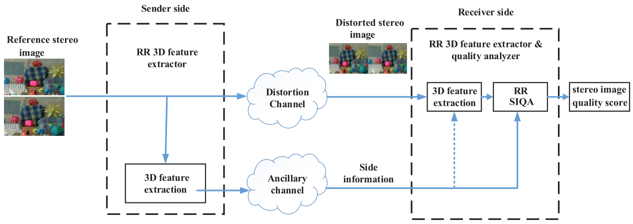


FIGURE 1. **The general framework for the SIQA system.**

Objective SIQA is a key link in 3D image processing systems. However, how to establish an effective SIQA metric has always been a difficult challenge. Particularly, when the two separate monocular images of a stereopair have different levels and types of distortion, it is called asymmetric distortion [2]. To address this issue, some existing researches [3]–[5] assume that the human visual system (HVS) may employ both the two monocular images' quality and the depth/disparity map quality to evaluate the quality of stereoscopic image. However, the ground truth depth/disparity maps are not always available, and meanwhile the depth/disparity information may not be directly related to the quality of stereoscopic image. In this case, there are still enormous spaces of research objective SIQA. In general, based on the reference information provided to the calculation model, SIQA methods can be divided into three categories: full reference (FR), reduced reference (RR) and no/blind reference (NR/blind) SIQA metrics. The FR metric operates on a distorted stereopair with a reference stereopair available for comparison, while the NR SIQA methods do not use reference information at all. As a compromise measure, the reduced reference (RR) metric uses only partial information or a handful of features extracted from the reference stereopair [3]. The general framework for the RRSIQA system is demonstrated in Fig. 1, which includes two parts, namely, sender side and receiver side, respectively. At the sender side, a feature extraction process is performed for reference stereopairs. Likewise, at the receiver side, the same procedure for distorted stereopairs at the sender side. In this study, we will focus on the RRSIQA method, which is widely used to guide the optimization of 3D content production.

Since the ultimate receiver of the image is the visual cortex of brain, the key point to objective image quality assessment (IQA) is to match the characteristics of HVS. In [6], Field and Olshausen showed that natural images can be sparsely unfolded by an overcomplete set of simple atoms. Furthermore, as a supplement to the distribution-based statistical description of natural images, the basic structure of natural images can be reflected in the field of retinal and cortical neurons [6]. In [7], Wang *et al.* validated that the human eyes are very sensitive to the change of structural information of input scenarios. Besides, the work of [8] declared that the efficiency of neural coding depended both on the transformations that map the input to the neural response and on the statistics of the input. Thusly, the evaluation of image quality by human eyes depends not only on the statistical characteristics of image but also on the visual characteristics of HVS. However, in previous studies, some SIQA metrics are mostly derived from the visual characteristics of HVS or the statistical characteristics of image, which do not assess the stereoscopic image quality accurately. Thusly, in this paper, we try to combine the visual characteristics of HVS with the statistical characteristics of images to overcome the shortcomings of a single strategy.

From the origin of sparse representation, it is directly related to compressed sensing (CS) [9]. The sparse representation theory proves that sparse or compressible signals can be accurately reconstructed from a small number of basic atoms onto a certain subspace [10]. With advancements in mathematics, sparse representation methods span a wide variety of applications, especially in the field of image processing, such as image segmentation [11], image denoising [12], visual tracking [13], and image super-resolution [14], etc. Meanwhile, sparse representation also shows great potential in dealing with the IQA issues [15]–[17]. Almost all existing sparse representation-based IQA methods follow a three-stage framework: dictionary learning (DL), quality-aware feature extraction, and regression model learning from subjective opinions. For DL, the K-SVD algorithm [18] is proven to be an effective method. In the stage of quality-aware feature extraction, the concept of entropy of primitive (EoP) has been proposed to measure the image visual information [19], [20]. Then, some typical SIQA metrics have been done based on the concept of EoP. For instance, Qi *et al.* [21] presented an RRSIQA metric by using binocular perceptual information. Wan *et al.* [22] proposed an RRSIQA method using sparse representation and natural scene statistics (NSS). Furthermore, in the regression analysis phase, the most common utilized regression model is support vector regression (SVR) with a radial basis function (RBF) kernel. Although these RRSIQA metrics achieve relatively well evaluation results, their performance is still limited. Therefore, an interesting question to consider is whether it is more effective to train the dictionary in a transform domain. A related work on this question is recently proposed by Liu *et al.*, where the dictionary is learned from the patches extracted in gradient-domain [23]. It suggests that the sparser the training samples/patches, the more powerful the learned dictionary. However, the works in [21], [22] are restricted to original pixel domain, and extending to more sparser gradient-domain is more appealing.

Apart from the sparse representation theory used for IQA issues, there have been a number of NSS-oriented IQA measurement theories in the last couple of years. For instance, the most extensively accepted method is to use the generalized Gaussian density (GDD) to model the marginal distributions of luminance wavelet coefficients [24]. However, due to the discrete wavelet transform (DWT) or discrete cosine transform (DCT) based manner applied in GDD, most existing NSS-based SIQA methods suffer from two drawbacks, namely, limited representation in image semantic structure and the use of computationally expensive image transformations, while these two issues are of great concern in IQA. Compared to the existing NSS-based IQA models, the low-order Gaussian derivative operators, such as GM and LOG, are very suitable for the design high performance NR IQA metrics [25]. The reasons why the GM and LOG features are so effective could be divided into two parts: a direct cause and an essential cause. Specifically, the direct cause is that the GM feature measures the strength of local luminance change, while the LOG operator responds to intensity contrast in a small spatial neighborhood. Furthermore, the essential cause is that the LOG operator is a good model of the receptive field of retinal ganglion cells [26], [27]. And in fact, such low-order Gaussian derivative operators have been widely applied in the applications of computer vision [28]–[30]. The effectiveness of the joint statistics of GM and LOG features in the work of [25] motivated us to introduce them into the task of RRSIQA metric.

Inspired by the above analysis, we propose a new RRSIQA metric using gradient sparse representation and structural degradation. Firstly, the binocular perceptual visual information (PVI) extracted by using gradient-based sparse representation. To be specific, the entropy of gradient primitives (EGP) of each view image is used as monocular cue, while the mutual information of gradient primitives (MIGP) between the two separate monocular images is regarded as binocular cue. Then, since HVS is very sensitive to the structural degradation of natural images, this paper considers the joint statistics of GM and LOG features to measure the structural degradation of each view image of the distorted stereopair, which is a supplement to the monocular cue EGP. Besides, compared with the SVR model, the kernel ridge regressing (KRR) fitting can be done in closed form and is usually faster for medium size datasets. Therefore, we use the KRR to establish the nonlinear relationship between quality-aware features and stereoscopic image quality index. A preliminary version of this study is published in [71], which does not consider the structural degradation of stereopair. Therefore, it is a sub-optimal model for all the available stereoscopic image information unemployed. In this paper, we have added some new insights and innovations to the initial version in the following ways so that the proposed model has higher accuracy and versatility.

The novelties of this study are generalized as follows.

(1) We propose a new RRSIQA model based on two complementary components: the sparsity properties of HVS and the joint statistics of image semantic structural degradation. These two complementary components are used to quantify the perceived quality degradation on each view image of stereoscopic image. With respect to the previous works, we demonstrate that the use of gradient sparse representation and joint statistics of GM and LOG features results in a higher consistent with subjective opinions.

(2) We introduce the concept of EGP and MIGP to achieve perceptual visual signal representation of the distorted stereoscopic image. More importantly, this study opens a new avenue to study how the sparse representation model in gradient domain can be used to RRSIQA framework design.

(3) Through a comprehensive verification, we find that the proposed model achieves a highly consistent with subjective scorings on both symmetric and asymmetrical distortions. Simultaneously, because the proposed model does not use the depth/disparity information of stereopair, the computational complexity of this model is low enough to meet the requirements of the real-time application.

The rest of this study is organized as follows. Section 2 provides an overview of the related work. Section 3 introduces the proposed RRSIQA model. Section 4 demonstrates and analyzes the experimental results. Section 5 summarizes this article.

II. RELATED WORK

With ready access to the booming markets of stereoscopic image based on a variety of 3D applications, the research of efficient and effective SIQA techniques are extremely essential. Currently, existing SIQA methods generally fall into four categories, named SIQA model extended from the typical 2D IQA models, SIQA method developed by simulating the characteristics of HVS, SIQA model designed by extracting the regularities of NSS, and SIQA model proposed based on deep learning.

A. SIQA MODEL EXTENDED FROM THE TYPICAL 2D IQA MODELS

As in the earlier studies, because a stereoscopic image consists of the left image and right image, some researchers attempt to extend the typical 2D IQA models to SIQA models. For simplicity, this kind of method usually processes each view image of stereopair independently, and combines the quality scores of the two views' image to yield the quality index of distorted stereoscopic image. For instance, You *et al.* [4], Benoit *et al.* [5], Campisi *et al.* [31], and Gorley and Holliman [32] extended the existing 2D IQA models to their SIQA models in a simple and direct manner. Furthermore, in [33], the relevance between subjective scores and three 2D IQA quality indexes, such as PSNR, video quality model (VQM) [34] and SSIM [7] for stereoscopic video were investigated. But obviously, since these models individually evaluate the quality of each view image without considering the binocular perceptual characteristics of HVS, the resulting performance most likely will not be satisfactory.

B. SIQA METHOD DEVELOPED BY SIMULATING THE CHARACTERISTICS OF HVS

As the study on SIQA moves forward, particularly with increasing cognition of the binocular perceptual mechanism, such as binocular fusion, binocular rivalry and depth perception behaviors of HVS, many scholars try to put these specific characteristics into SIQA design. In [35], a FR SIQA model based on the binocular fusion process was proposed. In [36], a FR SIQA model based on cyclopean image was presented. In [37], a FR SIQA metric by using binocular combination and binocular frequency integration was developed. The work of [38] simulated simple and complex cells in the primary visual cortex for SIQA design. In [39], a blind SIQA metric based on stacked auto-encoders (SAE) was proposed. In [40], a blind SIQA metric by simulating the whole visual perception route from the eyes to the frontal lobe was proposed. Li *et al.* [41] presented a FR SIQA index based on an adaptive cyclopean image by using ensemble learning. Li *et al.* [42] proposed an efficient general purpose blind SIQA model based on learned features from binocular combined images. Shao *et al.* [43] presented a blind asymmetrically distorted SIQA model based on supervised dictionary learning. The study of [44] presented a blind SIQA metric via using joint sparse representation. In our previous work, a FR SIQA model is proposed by learning binocular visual properties [45]. Liu *et al.* [46] proposed a FR SIQA metric by simulating binocular behaviors of HVS. In [47], Liu *et al.* [46] proposed a novel FR SIQA metric by considering the depth information and integral color information of 3D image under cloud computing environment. Galkandage *et al.* [48] designed a stereoscopic video quality index based on the motion sensitive HVS model. However, because of the complexity of HVS is unsurpassed in nature as we know, the above-mentioned SIQA models may not precisely response the change caused by different distortions.

C. SIQA METHOD DESIGNED BY EXTRACTING THE REGULARITIES OF NSS

In the past few year, with the development of the statistical properties of natural images, several NSS-oriented SIQA methods have been proposed. For instance, in [49], a blind SIQA metric was presented using 2D and 3D NSS features. The work of [50] used natural stereoscopic scene statistics for SIQA design. Su *et al.* [51] extracted both univariate and generalized bivariate NSS features for blind SIQA metric. In [52], a blind SIQA metric was developed based on joint wavelet decomposition and statistical models. The work of [53] designed a novel RRSIQA metric based on the Gaussian scale mixtures (GSM) model in contourlet domain. Ma *et al.* [54] proposed an RRSIQA metric by extracting NSS features in the reorganized discrete cosine transform (RDCT) domain. Appina *et al.* [55] proposed a blind video quality index by measuring the statistical dependencies between motion and disparity subband coefficients of stereoscopic video. Nevertheless, most of these SIQA models need to use the disparity/depth information, while the disparity estimation is still a fundamental, unsolved mystery in the field of stereo-related research. On the other hand, traditional NSS-based models usually require computationally expensive image transformations. Thusly, these shortcomings are the bottlenecks that restrict their widespread applications.

D. SIQA METHOD PROPOSED BASED ON DEEP LEARNING

Nowadays, with the fast development of advanced artificial intelligence (AI) technology, more and more scholars focus their studies on deep learning. Meanwhile, the research of SIQA has evolved hand-in-hand with the deep learning. For instance, Zhang *et al.* [56] proposed a blind SIQA metric by learning the structure of stereoscopic image with the three-column convolutional neural network (CNN). Lv *et al.* [57] presented a blind SIQA metric by using binocular self-similarity (BS) and deep neural networks (DNN). Oh *et al.* [58] developed a blind SIQA metric based on a deep CNN model. Jiang *et al.* [59] presented a FR SIQA model via hierarchical deep feature degradation fusion. Zhou *et al.* [60] designed an end-to-end dual-stream interactive network for SIQA. Sun *et al.* [61] utilized CNN to learn deeper local quality-aware structures for SIQA. However, although these approaches have achieved satisfactory performance, the major shortcoming of above approaches is that the structures of deep learning networks usually lack interpretability.

III. THE PROPOSED METHOD

As discussed previously, we explore the fusion features in which incorporated gradient-based sparse representation features and joint statistics of GM and LOG features to build an efficient RRSIQA model. The proposed framework for the RRSIQA metric is shown in Fig. 2. Firstly, taking the general framework for the RRSIQA system, at the sender side, the adaptive dictionary learning is trained offline in gradient domain to build sparse representation of stereoscopic images. Note that, the process of gradient-based dictionary learning is independent of testing stereopairs. Afterwards, the GM maps and the LOG responses of monocular image of reference and distorted stereopairs are calculated by using the low-order Gaussian derivative operators. For each reference and distorted stereopairs, binocular PVI and joint statistics of structural degradation are applied to quality prediction. The binocular PVI extracted by using gradient-based sparse representation. More specifically, each view's EGPs are calculated as monocular cue, and the left and right view's MIGP is derived as binocular cue. Moreover, the structural degradation is represented by the joint statistics of GM and LOG features. A perceptual loss vector is obtained by calculating binocular PVI and structural degradation differences between reference and distorted stereopairs. Finally, the perceptual loss vector is inputted into the KRR to build the maps between quality-aware features and objective quality scores of the test stereopairs. We describe our approach below.

A. GRADIENT SPARSE REPRESENTATION

The most popular interpretation of the sparse representation model is to assume that a natural signal represented by the vector $x \in R^n$, can be synthesized in term of a linear combination of only a few primitives or atoms, from a matrix $D \in R^{n\times k}$, termed a dictionary. Formally, sparse approximation can be represented by the formula: $\exists a \in R^k$ such that $x \approx Da$ and $\|a\|_0 \ll n$, note that, the $\|.\|$ is ℓ_0 norm, where the vector $a \in R^k$ is sparse: only a few of its entries are non-zeros. We typically assume $k > n$, implying that the D is redundant to x.

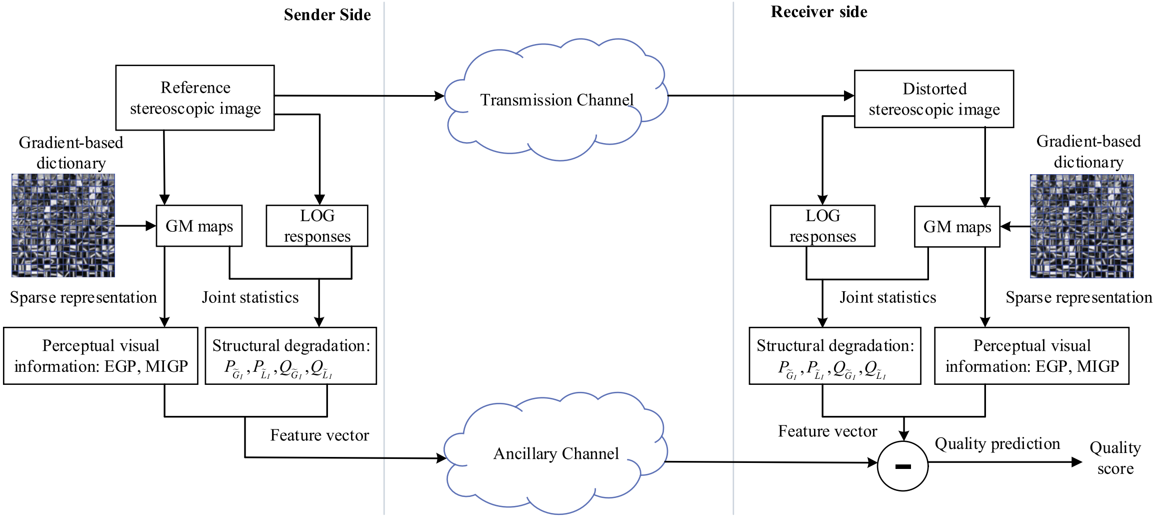

Sender Side
Receiver side
Reference stereoscopic image
Transmission Channel
Distorted stereoscopic image
Gradient-based dictionary
GM maps
LOG responses
Sparse representation
Joint statistics
Perceptual visual information: EGP, MIGP
Structural degradation: $P_{\tilde{G}_I}, P_{\tilde{L}_I}, Q_{\tilde{G}_I}, Q_{\tilde{L}_I}$
Feature vector
Ancillary Channel
LOG responses
GM maps
Gradient-based dictionary
Joint statistics
Sparse representation
Structural degradation: $P_{\tilde{G}_I}, P_{\tilde{L}_I}, Q_{\tilde{G}_I}, Q_{\tilde{L}_I}$
Perceptual visual information: EGP, MIGP
Feature vector
Quality prediction
Quality score
-

FIGURE 2. The proposed framework for the RRSIQA metric.

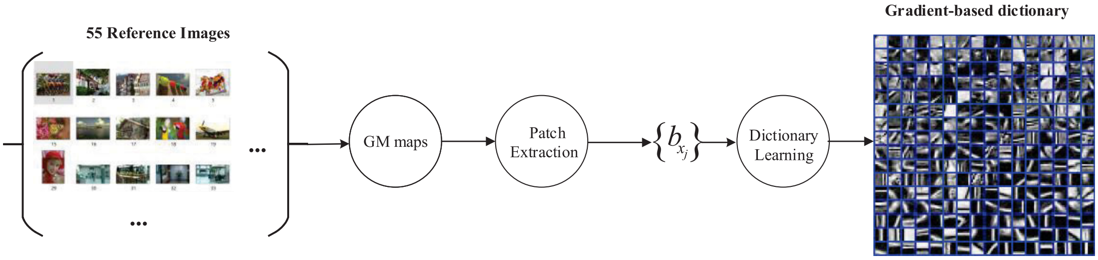


FIGURE 3. Illustration of gradient-based dictionary construction.

Since the gradients are sparser than the image itself, the learned dictionary in the gradient domain may have sparser representation than the pixel domain image [23]. This finding motivates us to learn the dictionary in the gradient domain. The process of gradient-based dictionary learning is shown in Fig. 3. Specifically, we choose 55 reference images from the LIVE 2D IQA dataset [7] and the IEEE Stereo IQA dataset [62] as image samples. For a given training image I, its gradient map I_{GM} can be defined by:

$$I_{GM} = \sqrt{[I \otimes h_x]^2 + \left[I \otimes h_y\right]^2} \tag{1}$$

where $\otimes$ refers to the linear convolution operator and h_d, $d \in \{x, y\}$, denotes the Gaussian partial derivative filter applied along the horizontal (x) or vertical (y) direction:

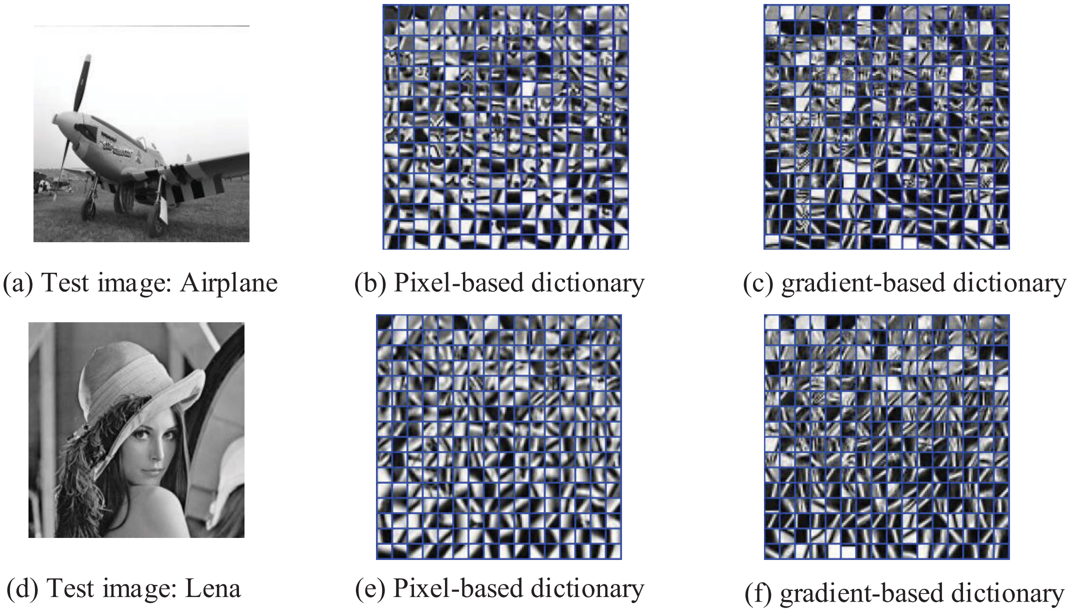
$$\begin{aligned} h_d\left(x, y \middle| \delta\right) &= \frac{\partial}{\partial d} f\left(x, y \middle| \delta\right) \\ &= -\frac{1}{2\pi\delta^2}\frac{d}{\delta^2}\exp\left(-\frac{x^2+y^2}{2\delta^2}\right)^{d\in\{x,y\},} \end{aligned} \tag{2}$$

where $f\left(x, y \middle| \delta\right) = \frac{1}{2\pi\delta^2}\exp\left(-\frac{x^2+y^2}{2\delta^2}\right)$ is the isotropic Gaussian function with scale parameter δ.

Given the gradient map I_{GM}, a set of k random, possibly overlapping patches with each of dimension $\sqrt{m} \times \sqrt{m}$ are extracted from I_{GM}. Then, every patch is verted to a vector of length m, and the patches are concatenated to form a matrix $B_x \in R^{m\times k}$. Furthermore, we learn an overcomplete dictionary $D_x \in R^{m\times n}$ that has n atoms ($m < n$) using the local patches in B_x as input. Our goal is to learn D_x such that each patch (column) $b_{x_j} \in B_x$ can be closely approximated as a linear superposition of a small number of atoms in D_x. This is achieved by solving the following sparse optimization problem:

$$\begin{cases} \min\limits_{\{D_x, a_x\}} \sum\limits_j \|a_{x_j}\|_p \\ s.t. \quad \forall i, \ \|b_{x_j} - D_x a_{x_j}\|_2 \le \varepsilon \end{cases} \tag{3}$$

where the vector $\mathbf{a}_{\mathbf{x_j}} \in R^n$ is the sparse representation of the patch $\mathbf{b}_{\mathbf{x_j}} \in R^n$. The value of p is typically 0 or 1, and ε refers to the reconstruction error controlled by the user. It is worth noting that the patch size of I_{GM} is set as 8*8, and the number of primitives in the trained gradient dictionary is set as 256. Besides, the classic K-SVD model [18] is utilized for computing the gradient dictionary D_x. To better understand the benefit of sparse representation in the gradient domain,


(a) Test image: Airplane

(b) Pixel-based dictionary

(c) gradient-based dictionary

(d) Test image: Lena

(e) Pixel-based dictionary

(f) gradient-based dictionary

FIGURE 4. Pixel-based dictionary and gradient-based dictionary comparison for test images Airplane and Lena.

one demonstration of visual inspection between traditional pixel-based dictionaries and gradient-based dictionaries is shown in Fig. 4. The learned dictionaries in the pixel domain depicted in Fig. 4 (b) and (e) and the learned dictionaries in the gradient domain illustrated in Fig. 4 (c) and (f), which are learned from test images Airplane and Lena. Compared to the pixel-based dictionaries in Fig. 4(b) and (e), As can be seen from Fig. 4 (c) and (f), the gradient-based dictionaries show a larger range of feature orientations, crisper features, and less redundancy.

For each patch (column) $b_{x_j} \in B_x$, the process of calculating its sparse representation vector a_{x_j} with respect to the dictionary D_x is called sparse coding, which can be formulated as follow:

$$\begin{cases} a_{x_j} = \arg\min \left\| b_{x_j} - D_x a_{x_j} \right\|_2^2 \\ s.t. \quad \left\| a_{x_j} \right\| < L \end{cases} \tag{4}$$

where L is the number of primitives used to represent the sparse level of each patch. Although the problem (4) is usually NP-Hard, it can be approximated by various techniques. In this study, because of the simplicity and effectiveness of the orthogonal matching pursuit (OMP) [63] algorithm, we use it to solve problem (4). According to [20], the EoP can be used to measure the amount of visual information in an image. In this section, in order to better show that the proposed EGP can represent the image visual information more sparsely, the EoP/EGP, PSNR and SSIM comparison curves for test images Airplane and Lena with regarding to image primitives and gradient primitives are shown in Fig. 5. As can be clearly seen from Fig. 5 (a) and (d), the two values of EoP and EGP converge almost simultaneously when the number of primitives L is equal to 60. We also observe that the value of EGP is less than the value of EoP. These findings are strong evidence that the extracted sparse representation vector with respect to the gradient dictionary is more sparsely, and it is of great benefit to measure image visual information. Furthermore, from Fig. 5 (b), (c), (e) and (f), it is very clear to see that the reconstructed image quality gets better and better for the two types of dictionaries, as the number of primitives to represent each patch L increases. Intriguingly, one important observation from Fig. 5 (b), (c), (e) and (f) is when $L = 60$, the SSIM values of the reconstructed images for the two types of dictionaries are equal, while the PSNR values of the reconstructed images with respect to gradient-based dictionaries are higher than the PSNR values of the reconstructed images with respect to pixel-based dictionaries. That means, as $L = 60$, the reconstructed image with respect to gradient-based dictionary is closer to the original image in visual perception.

B. PERCEPTUAL VISUAL INFORMATION EXPRESSION

During natural vision, the classical and nonclassical receptive fields function together to form a sparse representation of the visual world [64]. In [6], Field and Olshausen declared that the basis or primitive represented sparsely have characteristics of spatially localized, bandpass, and oriented, etc., which are closely related to the characteristics of the receptive fields of simple cell. Additionally, in [23], sparse representation in gradient domain provides a good solution for image recovery. Therefore, in this study, we can hypothesize that the visual gradient primitives is a good representation of the basic units of visual perception, which is also analogous the receptive fields of simple cells in the visual cortex.

Typically, in previous studies, the work of [19] first proposed the concept of EoP to measure the amount of image visual information. With this model, the coefficients of primitives are considered, known as l_0 norm based EoP. In [65], Shi *et al.* showed that the l_1 norm based EoP is superior to the l_0 norm based one in measuring image visual information. Moreover, Wan *et al.* [22] proposed the concept of the entropy of classified primitives (ECP) to measure the monocular visual information. In this paper, we further explore the concept of EoP and propose a new concept based on image gradient primitives, namely, entropy of the gradient primitives (EGP), which is used for measuring image visual information.

Generally, a stereoscopic image consists of the left view image I_L and right view image I_R. Given a gradient dictionary D_x, we can compute the sparse representation matrix A_{X_L} and A_{X_R} for I_L and I_R, respectively. Assume d_k is the k-th visual primitive of D_x. Afterwards, the probability density of visual primitive d_k for I_L is calculated by

$$p_L^k = \frac{\left\| A_{x_L}^k \right\|_1}{\sum_{s=1}^{n} \left\| A_{x_L}^k \right\|_1} \tag{5}$$

And then, based on the Shannon theory, the EGP of the left view image I_L of stereoscopic image can be calculated by

$$EGP(I_L) = -\sum_{k=1}^{M} p_L^k \log\left(p_L^k\right) \tag{6}$$

Similarly, the probability density of visual gradient primitive for the right view image of stereoscopic image can be calculated in the same process.

Since HVS relies on both monocular and binocular cues to obtain effective stereoscopic perception, both cues should be considered simultaneously in SIQA design. To meet this goal, we utilize the MIGP as the binocular cues. Then, the sum of coefficients that d_k is used to reconstruct both the i^{th} path in left view and the j^{th} path in the right view can be defined by

$$a_{d_k}(i,j) = \begin{cases} \left|a_{x_i}[k] + a_{x_j}[k]\right|, & \\ \qquad\qquad if\ a_{x_i}[k] \neq 0\ and\ a_{x_j}[k] \neq 0 & \\ 0 \qquad\qquad otherwise & \end{cases} \tag{7}$$

where $a_{x_i}[k]$ and $a_{x_j}[k]$, $(i, j = 1, \ldots, n/2)$ are the coefficients of in the path of left image and the path of right image, respectively. Then, the sum of coefficients that is utilized to reconstruct the patches both in the left image and right image can be defined by

$$a^k = \sum_{i=1}^{n/2} \sum_{j=(n/2+1)}^{n} a_{d_k}(i,j) \tag{8}$$

Thus, the joint probability density of visual primitive d_k for the left image I_L and the right image I_R is defined by

$$p^k = \frac{a^k}{\sum_{k=1}^{n} a^k} \tag{9}$$

With the probability density distribution p_L, p_R and p, the MIGP can be defined by

$$MIGP\left(I_L; I_R\right) = \sum_{k \in \Omega}^{n} p^k \log\left(\frac{p^k}{p_L^k p_R^k}\right) \tag{10}$$

Note that, where $\Omega = \left\{k \,|p_L^k \times p_R^k \neq 0\right\}$.

Finally, the **PVI** of a test stereoscopic image can be described by

$$\mathbf{PVI}\,(\mathbf{i}) = [\mathbf{EGP}\,(\mathbf{I_L})\,, \mathbf{EGP}\,(\mathbf{I_R})\,, \mathbf{MIGP}\,(\mathbf{i})] \tag{11}$$

where **i** refers to the i-th pair of distorted stereopair.

C. STRUCTURAL DEGRADATION DESCRIPTION

In addition to the sparsity properties of HVS, it is believed that the HVS learns through evolution and experience over the lifespan to exploit the statistical structure of natural images when performing visual tasks [66]. Considering the local spatial contrast features of images convey important structural information, and are closely related to the perceived quality of images. In this paper, the local contrast features, namely GM maps and LOG response, are used to measure the structural degradations due to the presence of distortions of stereopairs. The reasons for this operation are twofold. On one hand, in designing RRSIQA model it is critical to choose the quality-aware features in a way that have low computational requirements. To meet that need, we take into account quality-aware features from the spatial domain in order to reduce costly computation introduced by image transform, i. e. the spatial domain image is transformed to frequency domain or wavelet domain to obtain the features. On the other hand, bandpass image responses, in especial Gaussian derivative responses, can be employed for characterizing all kinds of image semantic structures, including lines, blobs, corners, and edges, etc. These semantic structures are closely related to human perception of image quality. Therefore, according to [25], the joint statistics of GM and LOG features are extracted from each view image of stereopair in order to measure changes in image semantic structures.

To be specific, each view image of stereopair is decomposed into just two channels, the GM map channel and the LOG response channel. Then, the GM map is computed based on the formula (1) and (2). Meanwhile, the LOG of each view image of stereopair I is defined by

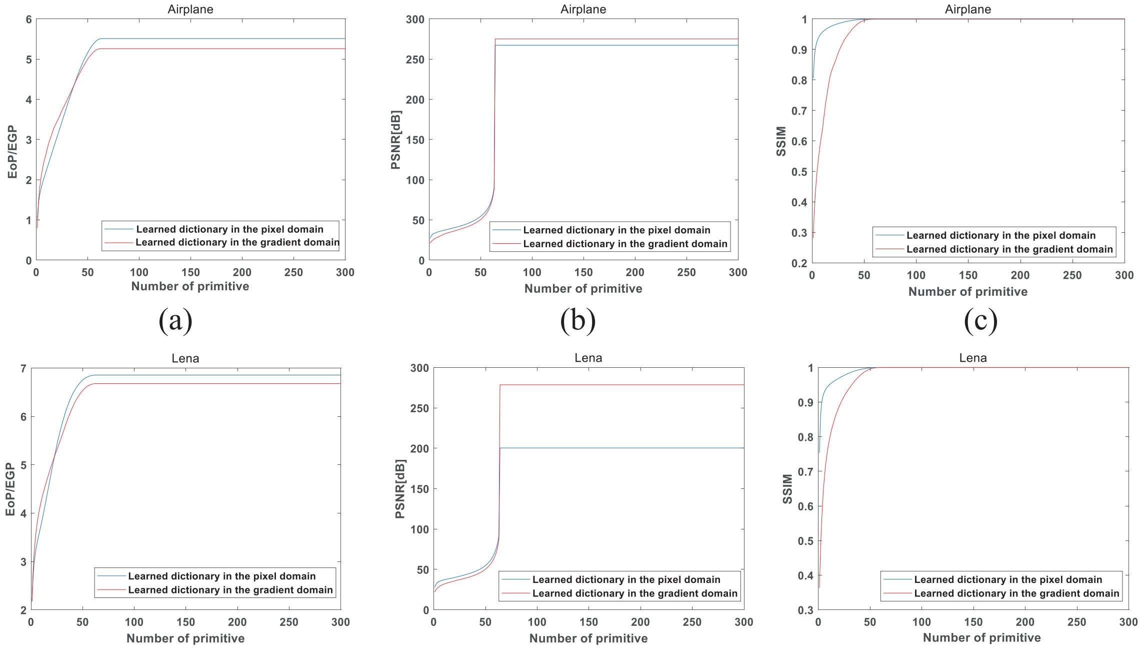
$$L_I = I \otimes h_{LOG} \tag{12}$$

where

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

Then, the coefficients of calculated GM and LOG are normalized to obtain stable statistical image representations:

$$\widetilde{G}_I = \frac{G_I}{\left(N_I + \zeta\right)} \tag{13}$$



(a) (b) (c)

(d) (e) (f)

FIGURE 5. EoP/EGP, PSNR and SSIM comparison curves for test images Airplane and Lena with regarding to image primitives and gradient primitives.

$$\widetilde{L}_I = \frac{L_I}{(N_I + \zeta)} \tag{14}$$

Note that, the locally adaptive normalization factor N_I in the formula (12) and (13) is given at each location (i, j) as follow:

$$N_I(i,j) = \sqrt{\sum\sum\nolimits_{(l,k)\in\Omega_{i,j}} \omega(l,k) T_I^2(l,k)} \tag{15}$$

where $\Omega_{i,j}$ is a local window centered at (i, j), $\omega(l, k)$ are positive weights with satisfying $\sum_{l,k} \omega(l, k) = 1$, and $T_I(i, j) = \sqrt{G_I^2(i,j) + L_I^2(i,j)}$.

The marginal probability functions of $\widetilde{G}_I$ and $\widetilde{L}_I$, denoted by $P_{\widetilde{G}}$ and $P_{\widetilde{L}}$, respectively, which are defined by

$$P_{\widetilde{G}_I}\left(\widetilde{G}_I = g_m\right) = \sum\nolimits_{n=1}^{N} K_{m,n} \tag{16}$$

$$P_{\widetilde{L}_I}\left(\widetilde{L}_I = l_n\right) = \sum\nolimits_{m=1}^{M} K_{m,n} \tag{17}$$

where $K_{m,n} = P\left(\widetilde{G}_I = g_m, \widetilde{L}_I = l_n\right), m = 1, \ldots, M;\ n = 1, \ldots, N$ is the joint empirical probability function of $\widetilde{G}_I$ and $\widetilde{L}_I$, while m and n refer to the quantization levels of $\widetilde{G}_I$ and $\widetilde{L}_I$. Considering the fact that there are dependencies between the GM and LOG, the following two quality-aware features to measure the dependency between GM and LOG can be defined by

$$Q_{\widetilde{G}_I}\left(\widetilde{G}_I = g_m\right) = \frac{1}{N}\sum\nolimits_{n=1}^{N} P\left(\widetilde{G}_I = g_m \middle| \widetilde{L}_I = l_n\right) \tag{18}$$

$$Q_{\widetilde{L}_I}\left(\widetilde{L}_I = l_n\right) = \frac{1}{M}\sum\nolimits_{m=1}^{M} P\left(\widetilde{L}_I = l_n \middle| \widetilde{G}_I = g_m\right) \tag{19}$$

As a result, the feature vectors **SD** can be obtained by concatenating all the above-mentioned four types features to measure the structural degradation of a test stereoscopic image:

$$\mathbf{SD} = \left[\mathbf{P}_{\widetilde{\mathbf{G}}_\mathbf{I}}, \mathbf{P}_{\widetilde{\mathbf{L}}_\mathbf{I}}, \mathbf{Q}_{\widetilde{\mathbf{G}}_\mathbf{I}}, \mathbf{Q}_{\widetilde{\mathbf{L}}_\mathbf{I}}\right] \tag{20}$$

To better illustrate how the distortions of stereoscopic images affect the distribution of the feature vectors **SD**, the joint normalized histograms of **SD** at different DMOS levels with five distortion types are shown in Fig. 6. Intuitively, we can clearly see that the shapes of the joint normalized histograms resemble each other in appearance across the same type of distortion. This means that the joint normalized histogram behaves in a content independent manner, and the feature vector **SD** is a stability and dependable statistical feature for RRSIQA task. Furthermore, as can be seen from Fig. 6 also demonstrates that the histograms are changed with different levels of distortion. Obviously, the more serious the distortion, the greater the change of histogram shape. This reveals that the histogram shape is closely related to the distortion level. Consequently, we can summarize that the feature vectors **SD** serve as good discriminatory features for measuring the structural degradation of distorted stereopair.

D. QUALITY PREDICTION

In the quality prediction stage, we believe that the perceptual visual information loss and the joint statistics of structural degradation can objectively reflect the quality difference between reference and distorted stereopairs. To be specific,

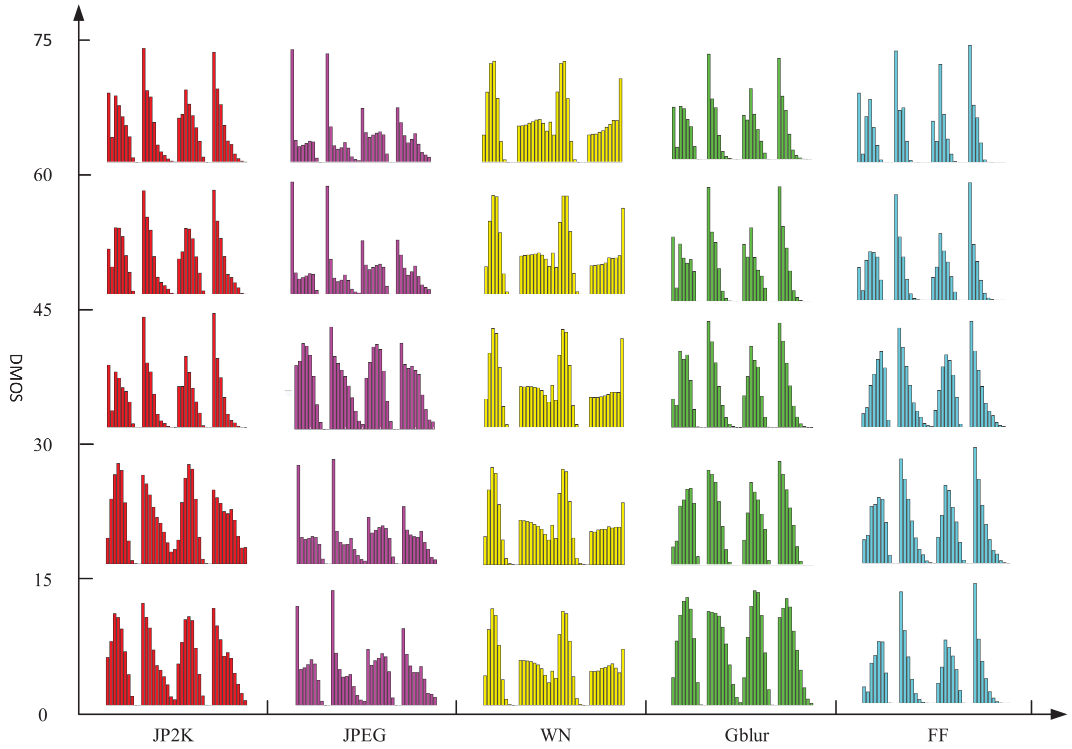


FIGURE 6. Joint normalized histograms of the feature vector $\mathbf{SD_I}$ at different DMOS levels with five types of distortions: JP2K, JPEG, WN, Gblur, and FF. The tested images are selected from LIVE 3D IQA datasets [36], [68].

the EGP and MIGP are represented as the binocular perceptual visual information, which are monocular cue and binocular cue respectively. The joint statistics of GM and LOG features **SD** are represented as structural degradation, which are complementary to the monocular cue EGP. The differences of **PVI** and **SD** between the reference stereopairs and their distorted versions are computed as loss vector **F**, which can be defined by:

$$\mathbf{F} = [\mathbf{LPVI}, \mathbf{LSD\_L}, \mathbf{LSD\_R}] \tag{21}$$

$$\mathbf{LPVI} = \mathbf{PVI^O} - \mathbf{PVI^D} \tag{22}$$

$$\mathbf{LSD_V} = \mathbf{SD_V^O} - \mathbf{SD_V^D} \tag{23}$$

where O and D denote original and distorted stereoscopic images, respectively. Note that, $V \in \{L, R\}$, L and R refer to left image and right image of a stereoscopic image, respectively.

To obtain the quality index of stereoscopic image, the KRR framework is used to build a map from the loss vector **F** to the perceived image quality. More specifically, with regarding to a training set $\{(\mathbf{x_1},\mathbf{y_1}), (\mathbf{x_2},\mathbf{y_2}) \ldots (\mathbf{x_n},\mathbf{y_n})\} \in R^m \times R^1$, The classic approach is to minimize the quadratic cost:

$$C(w) = \arg\min \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{y_i} - w^T\mathbf{x_i}\right)^2 \tag{24}$$

However, in the eigenspace, when we substitute $\mathbf{x_i} \rightarrow \Phi(\mathbf{x_i})$, it may be run into the risk of over-fitting. For avoiding this case, it is necessary to regularize it and set reasonable standards for selecting a mapping $C : R^m \rightarrow R$ to minimize the cost function as follow:

$$C = \arg\min \sum_{i=1}^{n}\left(\mathbf{y_i} - w^T\mathbf{x_i}\right)^2 + \frac{1}{2}\lambda\|w\|^2 \tag{25}$$

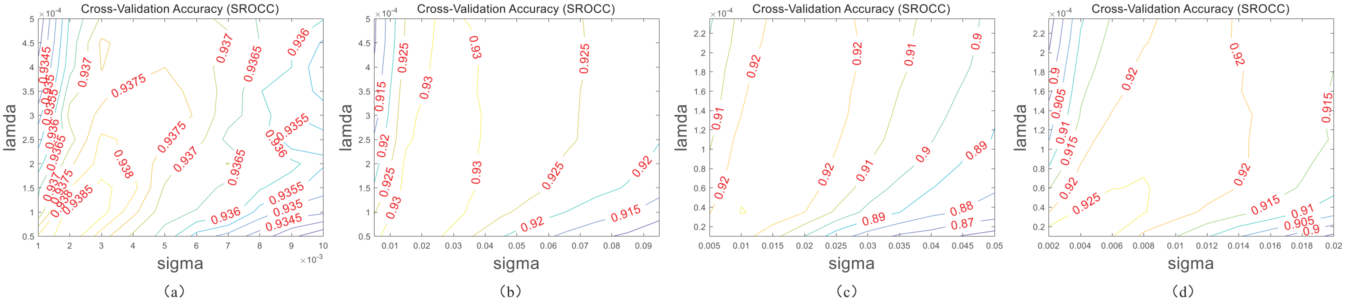
where $\lambda\|w\|^2$ is a regularization term used to stabilize the inverse numerically [65]. In this paper, $\mathbf{x_i}$ denotes the loss vector, and $\mathbf{y_i}$ is the subjective score of the i-th stereoscopic image. In accordance to [67], the solution of Eq. (25) can be formulated as follow:

$$w = \sum_{i=1}^{n} a_i \Phi(\mathbf{x_i}) \tag{26}$$

Substitute Eq. (26) into Eq. (25), the problem is converted to the optimal solution of the coefficient α, we can obtain Eq. (27) as follow:

$$a^* = \arg\min \frac{1}{2}\sum_{i}\left(\mathbf{y_i} - \sum_{j} a_j \Phi^T(\mathbf{x_j})\,\Phi(\mathbf{x_i})\right)^2 + \frac{1}{2}\lambda\sum_{ij} a_i a_j \Phi^T(\mathbf{x_j})\Phi(\mathbf{x_i}) \tag{27}$$

Then, the inner product of the feature space can be expressed as the kernel functions $K(\mathbf{x_i}, \mathbf{x_j}) = \Phi^T(\mathbf{x_i})\,\Phi(\mathbf{x_j})$, which is substituted into Eq. (27), we can



(a)

(b)

(c)

(d)

FIGURE 7. The proposed metric for the KRR parameters (λ, σ) selection process via 2D grid search technique on four databases (a) LIVE 3D IQA database phase I [68], (b) LIVE 3D IQA database phase II [36], (c) Waterloo IVC 3D database phase I [69], (d) Waterloo IVC 3D database phase II [70].

obtain Eq. (28) as follow:

$$\alpha^{*} = \arg\min \frac{1}{2}\sum_{i}\left(\mathbf{y_i} - \sum_{j}\alpha_j K\left(\mathbf{x_i}, \mathbf{x_j}\right)\right)^2 + \frac{1}{2}\lambda\sum_{ij}\alpha_i\alpha_j K\left(\mathbf{x_i}, \mathbf{x_j}\right) \quad (28)$$

For simplicity, we arrange Eq. (28) in matrix form:

$$\alpha^{*} = \arg\min \frac{1}{2}(\mathbf{Y} - K\alpha)^T\left(\mathbf{Y} - K\alpha\right) + \frac{1}{2}\lambda\alpha^T K\alpha \quad (29)$$

where K is named as reproducing kernel, and $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}]^T$. It should be noted that the Gaussian kernel is adopted in this paper, which can be defined by:

$$K\left(\mathbf{x_i}, \mathbf{x_j}\right) = \exp\left(-\frac{\left\|\mathbf{x_i} - \mathbf{x_j}\right\|_2^2}{\sigma^2}\right) \quad (30)$$

Besides, when the KRR algorithm is used to build the map between the loss vector and subjective scores, we need to set the parameters of the regularization term λ and the Gaussian kernel σ. To serve this purpose, we use a 2D grid search technique with 200 times cross-validation to find out the optimal parameter values of (λ, σ). As illustrated in Fig. 7, the optimal values of (λ, σ) are set to be (1.0e-04, 0.002), (5.0e-05, 0.015), (3.5e-0.5, 0.01), (1.0e-0.5, 0.002) on the LIVE 3D IQA database phase I [68], LIVE 3D IQA database phase II [36], Waterloo IVC 3D database phase I [69] and Waterloo IVC 3D database phase II [70], respectively. We use them in the following experiments.

IV. EXPERIMENTAL RESULTS

To verify the performance of the proposed RRSIQA model, we analyze its ability to evaluate symmetric and asymmetric of distorted stereoscopic images from the aspects of prediction accuracy, monotonicity, and consistency on four popular SIQA databases. A more detailed description is given in the following section.

A. 3D IQA DATABASES AND PREDICTION PROTOCOLS

The LIVE 3D IQA database consists of two phases, in which five distortion types and different distortion levels are provided, including JP2K (JPEG2000 compression), JPEG (JPEG compression), WN (additive white Gaussian noise), Gblur (Gaussian blur), and FF (fast fading). Specifically, the LIVE 3D IQA Phase I [68] contains 20 reference stereopairs and 365 symmetrically distorted versions. The LIVE 3D IQA Phase II [36] contains 8 reference stereopairs and 120 symmetrically distorted stereopairs and 240 asymmetrically distorted stereopairs.

The Waterloo IVC 3D database also contains two phases, in which three distortion types and four distortion levels are provided. Specifically, the Waterloo IVC 3D Phase I [69] contains 6 reference stereoscopic images and 324 distorted versions (72 symmetrically and 252 asymmetrically distorted stereopairs). The Waterloo IVC 3D Phase II [70] contains 10 reference stereopairs and 450 distorted versions (120 symmetrically and 330 asymmetrically distorted stereopairs). Compared with the LIVE 3D IQA database, the Waterloo IVC 3D database contains mixed distortion types and distortion levels in asymmetrically distorted stereoscopic images.

Three popular objective quality metrics, namely, PLCC (Pearson linear Correlation Coefficient), SRCC (Spearman Rank-order Correlation Coefficient), and RMSE (Root Mean Square Error), are utilized to evaluate quality prediction performance. The more PLCC and SRCC values tend to 1, the more RMSE values tend to 0, representing better performance. Before calculating the criteria, the nonlinearity regression of quality scores with subjective opinions is required by using a five parameters logistic function [71] as follows:

$$f\left(t\right) = \alpha_1\left(\frac{1}{2} - \frac{1}{1 + \exp\left(\alpha_2\left(t - \alpha_3\right)\right)}\right) + \alpha_4 t + \alpha_5 \quad (31)$$

where t and $f\left(t\right)$ refer to the predicted objective quality score and the nonlinear fitting score, and a_i, $i = 1, 2, \ldots, 5$, are the regression parameters to be fitted.

In order to assess the correlation performance of the proposed model, each database is divided into two nonoverlapped sets: the training set and the test set. To be specific, we firstly randomly select 80% of the reference stereoscopic images as the training set and the rest 20% as the test set. By this means, we ensure that there is no content overlap between the training set and the test set. Next, the proposed

TABLE 1. The performance of the proposed model and the twelve competitive indicators are compared on two benchmark databases.

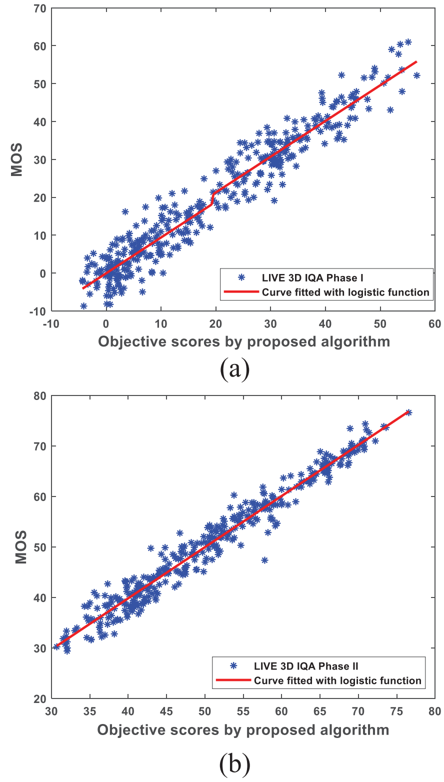
Databases	Criteria	FR				NR		RR						
		Liu[46]	Liu[47]	Jiang[59]	Ma[45]	Sun[61]	Shao[43]	Wang[53]	Ma[54]	Qi[21]	Wan[22]	Ma[72]	Ma[73]	Proposed
LIVE 3D IQA Phase I	PLCC	0.9430	0.9298	**0.960**	0.9469	0.951	0.91	0.8998	0.9056	0.934	0.944	0.9301	0.9321	0.9531
	SRCC	0.9402	0.9321	**0.953**	0.934	0.959	0.872	0.8922	0.9052	0.902	0.928	0.9292	0.9101	0.9375
	RMSE	5.4238	6.044	**4.455**	5.2111	4.573	–	7.1557	6.9542	5.98	5.547	6.0241	5.8854	4.9384
LIVE 3D IQA Phase II	PLCC	0.8417	–	0.932	0.93	0.938	0.887	0.5216	0.8179	0.915	0.945	0.9213	0.9252	**0.9464**
	SRCC	0.8307	–	0.927	0.9218	0.918	0.867	0.6123	0.7938	0.867	0.932	0.9175	0.9106	**0.9333**
	RMSE	6.0946	–	4.041	4.1232	3.809	–	9.6303	6.4939	4.409	3.65	4.3903	4.2360	**3.5774**

metric trained from the training set is examined on the test set, which is repeated 1000 times. Finally, the median PLCC, SRCC, and RMSE results from 1000 train-test interactions represent the final performance prediction.

B. OVERALL PERFORMANCE COMPARISON

To investigate the performance of the proposed RRSIQA metric for all distortion types, twelve representative advanced SIQA metrics are selected for comparison. They are divided into two categories: one consists of four FR SIQA models, including Liu's metric [46], Liu's metric [47], Jiang's metric [59], Ma's metric [45]; the other one consists of five RR SIQA models, including Wang's metric [51], Ma's metric [52], Qi's metric [21], Wan's metric [22], Ma's metric [72] and Ma's metric [73]. Among all the FR and RR SIQA models, Liu's metric [46] is by simulating binocular behaviors of HVS; Liu's metric [47] is by considering the depth and integral color information of stereoscopic image; Jiang's metric [59] and Sun's metric [61] are based on deep learning; Ma's metric [54] is based on NSS models; Wan's metric [22] and Ma's metric [72] are the fusion approaches by jointly considering NSS and HVS models. In [73], a preliminary version of this study is proposed based on entropy of gradient primitives. It is well known that the FR SIQA metric should has better performance than RR/NR SIQA methods due to the whole reference information of stereoscopic image used. However, we still choose some FR SIQA methods for comparison to prove the superior performance of the proposed RRSIQA method. Note that, because Shao's metric [43] focused on the asymmetrically distorted stereopairs based on sparse representation, and Sun's metric [61] used CNN to learn deeper local quality-aware structures for stereoscopic images, we also use them as comparisons.

The overall performance of the proposed RRSIQA metric on the LIVE 3D IQA Phase I and LIVE 3D IQA Phase II databases are tabulated in Table 1, where the best performing metrics for each database are highlighted in boldface. As can be seen from Table 1, the proposed metric achieves highly consistent with human evaluation, especially for asymmetrically distorted stereopairs. To be specific, we can see that most of the SIQA models achieve relatively well performance for the symmetric distortion but fall in the asymmetric distortion. One likely reason for that these models do not fully consider the statistical characteristics of natural scenes and the perceptual properties of HVS. For example, Liu's metric [46] only considers binocular behaviors of HVS; Ma's metric [54] extracts the NSS-based quality-aware features without taking into account the perceptual properties of HVS. Interestingly, the statistical characteristics of natural image and the perceptual properties are considered simultaneously in Wan's metric [22] and Ma's metric [72], resulting in improved performance. However, the general drawbacks of these models are that the sparse representation is restricted to original pixel domain, and some computationally expensive image transformations are adopted. Therefore, all of which achieve very limited performance improvement.



(a)

(b)

FIGURE 8. Scatter plots of objective scores against subjective ratings on the LIVE 3D IQA database. (a) LIVE 3D IQA Phase I. (b) LIVE 3D IQA Phase II.

To overcome the shortcomings of the above-mentioned models, this study more comprehensively considers the sparse properties of HVS and joint statistics of structural degradation of stereoscopic image. Experimental results confirm our hypothesis, and show the prediction performance is more consistent with human opinions. Moreover, the scatter

TABLE 2. Performance comparison of the other eleven metrics for each individual distortion type in terms of PLCC.

Databases	Types	FR				NR	RR						
		Liu[46]	Liu[47]	Jiang[59]	Ma[45]	Sun[61]	Wang[53]	Ma[54]	Qi[21]	Wan[22]	Ma[72]	Ma[73]	Proposed
LIVE 3D IQA Phase I	JPEG	0.7315	0.7792	0.799	0.7746	0.806	0.5574	0.7222	0.740	0.743	0.720	0.7389	**0.8047**
	JP2K	0.9423	0.9288	0.953	**0.961**	0.948	0.9252	0.9182	0.934	0.950	0.9399	0.9177	0.9563
	WN	0.9463	0.9276	0.963	0.9412	0.956	0.9196	0.9131	0.832	0.951	0.9347	0.9607	**0.9644**
	Gblur	0.9530	0.9499	0.952	0.9711	0.960	0.9596	0.9247	0.920	0.953	0.9356	0.9723	**0.9796**
	FF	0.8658	0.7406	0.887	0.8941	0.890	0.8339	0.8086	0.817	**0.926**	0.8427	0.8839	0.8972
LIVE 3D IQA Phase II	JPEG	0.8758	–	0.900	0.935	0.823	0.8971	0.7544	0.871	0.917	0.7645	0.9241	**0.9609**
	JP2K	0.8701	–	0.936	**0.967**	0.900	0.9524	0.8094	0.858	0.885	0.8795	0.9605	0.9616
	WN	0.9325	–	0.971	0.9341	0.956	0.7738	0.822	0.891	**0.980**	0.9325	0.9421	0.9544
	Gblur	0.9430	–	0.984	0.9384	**0.996**	0.9433	0.9721	0.981	0.990	0.9131	0.9435	0.9682
	FF	0.9218	–	0.950	0.9489	0.901	0.9030	0.9016	0.9250	**0.9640**	0.9063	0.955	0.9584

TABLE 3. Performance comparison on waterloo IVC 3D phase I database.

Criteria		PLCC			SRCC			RMSE		
Type	Method	Symmetric	Asymmetric	All	Symmetric	Asymmetric	All	Symmetric	Asymmetric	All
FR	You[4]	0.8681	0.7089	0.7125	0.7517	0.5706	0.5968	9.929	11.3482	11.6526
	Benoit[5]	0.8503	0.6970	0.6797	0.7275	0.5766	0.5852	9.9599	11.0993	11.5208
	Chen[36]	0.9553	0.7324	0.7337	0.9241	0.6428	0.6815	5.8938	10.592	10.7095
RR	Qi[21]	0.8349	0.7713	0.8266	0.7394	0.6972	0.7234	7.5374	7.049	8.8437
	Wan[22]	0.9582	0.8945	0.9177	0.9048	0.8377	0.8647	4.6881	5.454	6.0588
	Proposed	**0.9771**	**0.9361**	**0.9487**	**0.9473**	**0.9148**	**0.929**	**3.8785**	**4.970**	**4.9447**

plots of objective scores predicted against subjective mean opinion scores (MOS) on the two LIVE 3D IQA databases are showed in Fig. 8. As can be seen from Fig. 8 (a) and (b), the proposed RRSIQA model achieves high consistency with subjective scores. Therefore, the sparse representation in gradient domain and the joint statistics of image semantic structural degradation are two complementary components for measuring the degradation of stereo image quality. Based on the above analysis, we can conclude that the proposed RRSIQA model can be utilized to quantify and assess the symmetric and asymmetric distortions of stereoscopic images.

C. PERFORMANCE COMPARISON ON INDIVIDUAL DISTORTION TYPES

Different types of image distortion result in different viewing experiences, it is necessary to show the universality of the proposed model for individual distortion types. Therefore, the eleven typical schemes are selected and compared with the proposed method on each type of individual distortion. To save space, experimental results in terms of PLCC is tabulated in Table 2, and the best metrics have been highlighted in boldface. As can be seen from Table 2, the proposed metric ranks among the top 4 times in terms of PLCC on some specific types of distortion, followed by Wan's metric [22] 3 times, Ma's metric [45] 2 times and Sun's metric 1 times, in especial, the proposed model achieves an impressive performance for JPEG distortion on the symmetric and asymmetric distortions. The principal reason is that the JPEG distortion mainly come from image blurring caused by the high frequency attenuation, which in turn lead to the degradation of image structure. Moreover, we have also found that the performance of the proposed model is close to the best for all individual distortion types. Therefore, we can conclude that the proposed RRSIQA model is comparable to the most efficient model for individual types across both symmetric and asymmetric distortions.

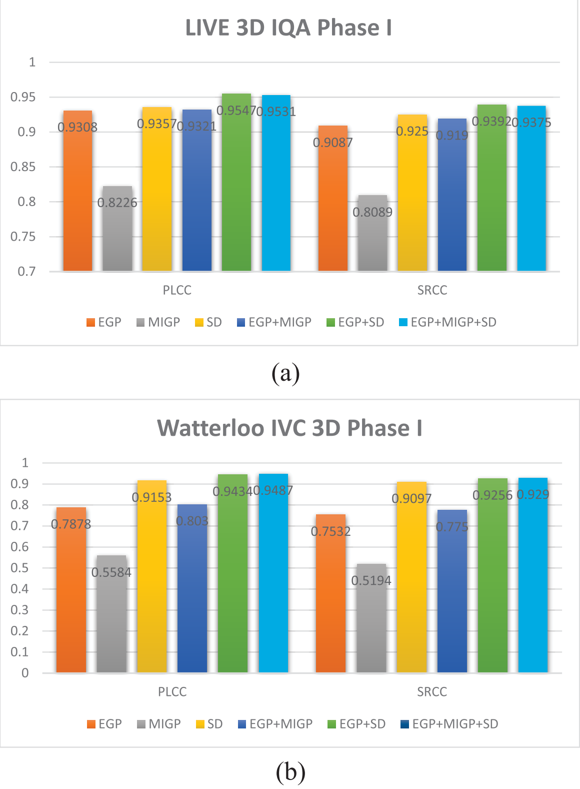
D. PERFORMANCE COMPARISON ON SYMMETRIC AND ASYMMETRIC DISTORTION TYPES

In order to further verify the effectiveness of the proposed method for asymmetric distorted stereoscopic images, we also conduct experiments on the databases of Waterloo IVC 3D phase I and Waterloo IVC 3D phase II. For PLCC, SRCC, and RMSE, the results for symmetric and asymmetric distortions are illustrated in Table 3 and Table 4, respectively, where the best performance metrics for each database are highlighted in bold. From Table 3 and Table 4, it can be observed that the performance of the proposed scheme is better than other schemes in both symmetric and asymmetric distorted stereoscopic images. For instance, the PLCC and SRCC values of the proposed metric are about 0.15 and 0.21 higher than the work of [21] on the Waterloo IVC 3D phase I database. Also, the PLCC and SRCC values of the

TABLE 4. Performance comparison on waterloo IVC 3D phase II database.

Criteria		PLCC			SRCC			RMSE		
Type	Method	Symmetric	Asymmetric	All	Symmetric	Asymmetric	All	Symmetric	Asymmetric	All
FR	You[4]	0.7634	0.6857	0.6817	0.5602	0.5997	0.5873	12.4031	14.8307	14.7335
	Benoit[5]	0.7549	0.5548	0.5507	0.5713	0.4539	0.4595	13.4575	16.152	16.177
	Chen[36]	0.8371	0.633	0.613	0.7581	0.4595	0.5781	11.6165	14.9701	15.1222
RR	Qi[21]	0.6403	0.7037	0.6815	0.5612	0.6425	0.6081	11.2828	12.2424	14.1667
	Wan[22]	0.9105	0.8797	0.8916	0.8571	0.8256	0.8492	7.7036	7.7044	8.7521
	Proposed	**0.9651**	**0.9338**	**0.9420**	**0.9379**	**0.9200**	**0.9292**	**5.1548**	**6.4679**	**6.3361**

proposed scheme are respectively 0.26 and 0.32 higher than the work of [21] on the Waterloo IVC 3D phase II database. The reason is that sparse representation in original pixel domain and a single strategy do not provide the best performance in all situations. Based on the above observations, we can draw conclusion from the proposed method is in significant agreement with subjective judgments on symmetric and asymmetric distorted stereoscopic images.



(a)

(b)

FIGURE 9. Performance comparison of EGP, MIGP, and SD on the two 3D IQA databases. (a) LIVE 3D IQA Phase I. (b) Waterloo IVC 3D Phase I.

E. IMPACT OF EACH COMPONENT IN THE PROPOSED SCHEME

Since the perception of image quality by human eyes is sparse and sensitive to the image structure degradation, we should consider these two visual characteristics simultaneously when designing SIQA model. In order to further understand how to combine sparse representation and structural degradation to improve the prediction performance of the proposed measurement method, some feature analyses and ablation experiments are given. Three sets of quality-aware features are used in the proposed scheme, including one binocular cue MIGP, and two monocular cues EGP and SD. On the LIVE 3D IQA phase I database and Waterloo IVC 3D phase II database, PLCC and SRCC feature groups and their combined performance comparisons are provided respectively, and the results are shown in Fig 9. The MIGP represents the binocular visual information, and the EGP represents the monocular visual information, which are extracted from sparse representation in the gradient domain. Since the sparse representation considers the sparse characteristics of HVS, their respective performance looks good. The SD is the joint statistics of LOG and GM features, which can be used to measure the structural degradation of natural image. The combinations of each group's feature are EGP+MIGP, EGP+SD and EGP+MIGP+SD. As can be seen from Fig. 9, the prediction performance of the proposed scheme can be further improved by properly chaining each set of features together. Interestingly, the feature EGP achieves better performance than feature MIGP. The most likely reason is that monocular cues mainly emphasize the characteristics of visual stimuli, while binocular cues emphasize the role of feedback information generated by the coordinated activities of the two eyes. Furthermore, we can also observe that the features SD achieve better performance than the features EGP, MIGP and EGP+MIGP. There are two main reasons for this. One is that the human eye evaluates the image quality, the statistical properties of HVS may be take precedence over the perceptual properties of HVS. The other is that the HVS is very sensitive to structural degradation in an image. In addition, from Fig. 9.(a), it can be seen that the performance of the features EGP+SD achieve better performance than the features EGP+MIGP+SD. A logical explanation is that the symmetric distorted stereoscopic image does not trigger binocular rivalry. In general, it can be seen from Fig. 9 that the combined effect of the three feature groups is better than

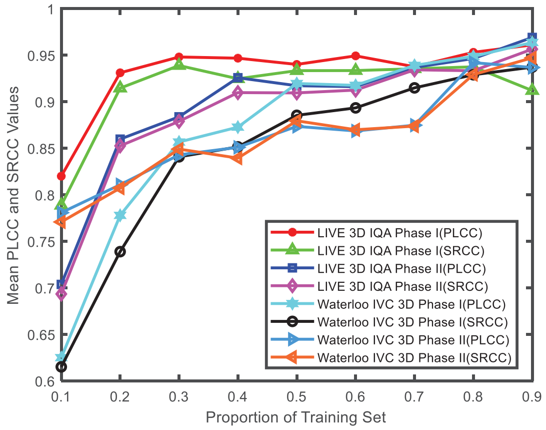


FIGURE 10. PLCC and SRCC performed different cross-validation on LIVE 3D IQA Phase I, LIVE 3D IQA Phase II, Waterloo IVC Phase I and Waterloo IVC Phase II.

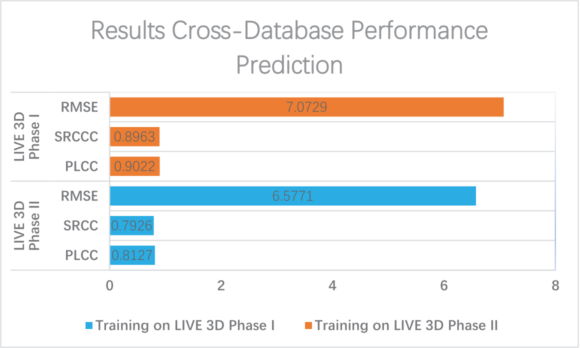


FIGURE 11. Cross-database performance prediction on LIVE 3D IQA Phase I and LIVE 3D IQA Phase II.

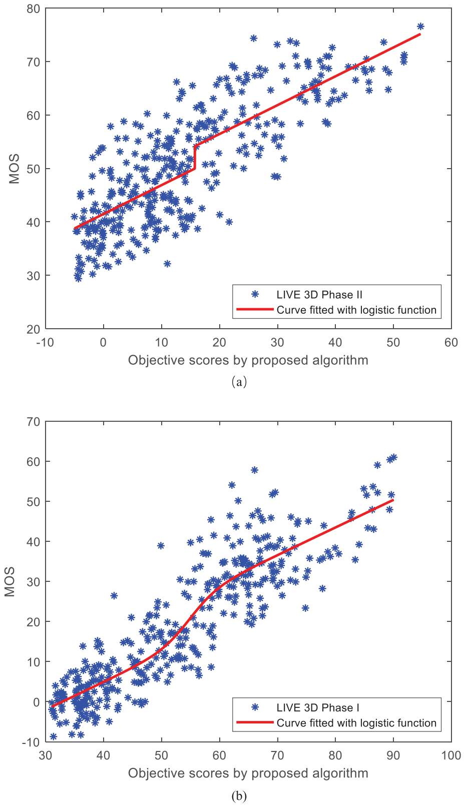
that of each feature group, which proves the complementarity and effectiveness of each feature group.

F. IMPACT OF PROPORTION OF TRAINING SET

In order to show that the proposed scheme is not highly dependent on the size of the training set, 1000 cross-validation experiments are conducted to test the performance of the proposed scheme under different proportion of a training set and a testing set. Note that, the same database uses the same KRR parameters. The results are demonstrated in Fig. 10. From Fig. 10, we can observe that a stable SIQA model can be derived from a small amount set of stereoscopic images. Intuitively, the PLCC and SRCC slightly decrease with a reduction of the proportion of training data, but it is significant above 30% of the LIVE 3D IQA phase I, 40% of the LIVE 3D IQA phase II, and 50% of the Waterloo IVC 3D phase I and phase II databases, respectively. Therefore, the proposed scheme is essentially independent of the size of the training set.

G. CROSS-DATABASE PERFORMANCE PREDICTION

The prediction strategy utilized in Section IV.A-F is inadequate to evaluate the generalization ability and robustness of the proposed metric, because the training subset and testing subset have the same distortions selected from a same database. Therefore, we conduct cross-database validation experiments on the LIVE 3D IQA Phase I [68] and LIVE 3D IQA Phase II [36]. Note that, the KRR parameters are set to be (1.0e-04, 0.002) and (5.0e-05, 0.015), respectively. Specifically, the proposed model is trained on one dataset, then testing it on the other dataset. The results are shown in Fig. 11. From Fig. 11, we can find that, when the proposed model learned on the LIVE 3D Phase II dataset, the cross-database performance prediction has better than the proposed model is trained on the LIVE 3D Phase I dataset. The most possible reason is that the LIVE 3D Phase II dataset includes not only symmetrically distorted stereopairs, but also asymmetrically distorted stereopairs, which raises the generalization ability and robustness to the proposed model learned. Moreover, the scatter plots of objective cross-database prediction scores against MOS on the two LIVE 3D IQA databases are showed in Fig. 12. As can be seen from Fig. 12(a) and (b), in this cross-database validation experiments, the proposed model achieves high consistency with subjective ratings. Therefore, we believe that the proposed model can deliver high generalization ability and robustness.



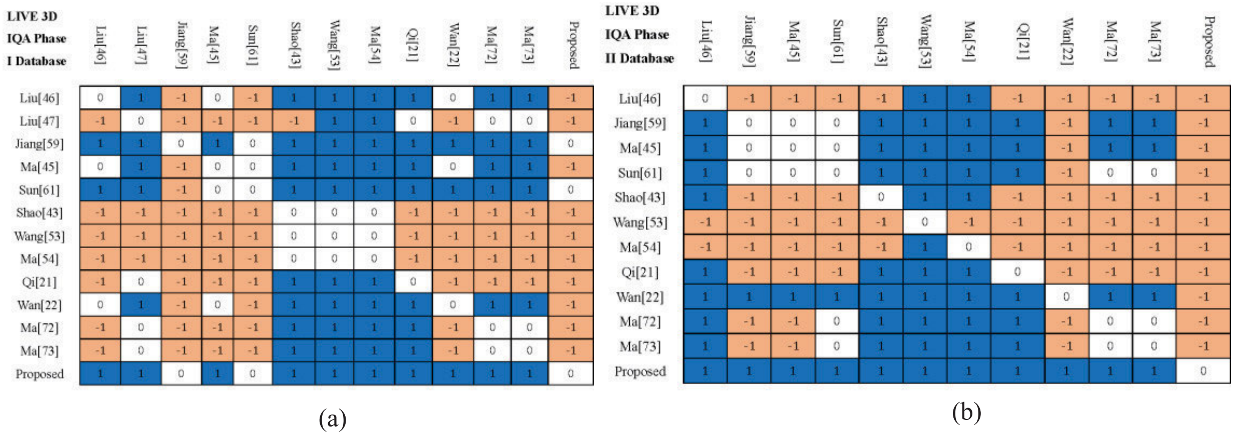
(a)

(b)

FIGURE 12. Scatter plots of objective cross-database prediction scores against subjective ratings on the two LIVE 3D IQA databases. (a) Training on LIVE 3D IQA Phase I. (b) Training on LIVE 3D IQA Phase II.

H. STATISTICAL EVALUATION

To verify whether a metric is statistically superior to another one, we conduct the one-sided t-test between the



(a)

LIVE 3D IQA Phase I Database	Liu[46]	Liu[47]	Jiang[59]	Ma[45]	Sun[61]	Shao[43]	Wang[53]	Ma[54]	Qi[21]	Wan[22]	Ma[72]	Ma[73]	Proposed
Liu[46]	0	1	-1	0	-1	1	1	1	1	0	1	1	-1
Liu[47]	-1	0	-1	-1	-1	-1	1	1	0	-1	0	0	-1
Jiang[59]	1	1	0	1	0	1	1	1	1	1	1	1	0
Ma[45]	0	1	-1	0	0	1	1	1	1	0	1	1	-1
Sun[61]	1	1	-1	0	0	1	1	1	1	1	1	1	0
Shao[43]	-1	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1
Wang[53]	-1	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1
Ma[54]	-1	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1
Qi[21]	-1	0	-1	-1	-1	1	1	1	0	-1	-1	-1	-1
Wan[22]	0	1	-1	0	-1	1	1	1	1	0	1	1	-1
Ma[72]	-1	0	-1	-1	-1	1	1	1	1	-1	0	0	-1
Ma[73]	-1	0	-1	-1	-1	1	1	1	1	-1	0	0	-1
Proposed	1	1	0	1	0	1	1	1	1	1	1	1	0

(b)

LIVE 3D IQA Phase II Database	Liu[46]	Jiang[59]	Ma[45]	Sun[61]	Shao[43]	Wang[53]	Ma[54]	Qi[21]	Wan[22]	Ma[72]	Ma[73]	Proposed
Liu[46]	0	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1
Jiang[59]	1	0	0	0	1	1	1	1	-1	1	1	-1
Ma[45]	1	0	0	0	1	1	1	1	-1	1	1	-1
Sun[61]	1	0	0	0	1	1	1	1	-1	0	0	-1
Shao[43]	1	-1	-1	-1	0	1	1	-1	-1	-1	-1	-1
Wang[53]	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1
Ma[54]	-1	-1	-1	-1	-1	1	0	-1	-1	-1	-1	-1
Qi[21]	1	-1	-1	-1	1	1	1	0	-1	-1	-1	-1
Wan[22]	1	1	1	1	1	1	1	1	0	1	1	-1
Ma[72]	1	-1	-1	0	1	1	1	1	-1	0	0	-1
Ma[73]	1	-1	-1	0	1	1	1	1	-1	0	0	-1
Proposed	1	1	1	1	1	1	1	1	1	1	1	0

FIGURE 13. Results of the one-sided t-test performance between SRCC values on the two LIVE 3D databases (a) LIVE 3D IQA Phase I. (b) LIVE 3D IQA Phase II.

TABLE 5. Computational complexity analysis of two FR and RRSIQA schemes.

Methods	Chen[36]	Ma[72]	Proposed
Run time (s)	36.12	4.31	0.16

correlation scores generated by the algorithms across the 1000 train-test trials. Note that, our analysis here is based on the mean SRCC values across all distortions over 1000 test sets. Fig 13. shows the t-test results conducted between any two SIQA methods on the two LIVE 3D IQA databases. A value of "1" indicates that the algorithm (row) is statistically superior to the algorithm (column). A value of "0" indicates statistical equivalence between the row and column, while a value of "-1" indicates that the algorithm (row) is statistically inferior to the algorithm (column). It is clearly shown that the proposed metric is statistically better than almost all existing SIQA schemes, especially for asymmetric distortion of stereoscopic image.

I. COMPUTATIONAL COMPLEXITY ANALYSIS

Computational complexity is another key factor to assessment the feasibility of the proposed metric. Therefore, we compare the computational complexity (the average running time in testing a pair of stereoscopic image with the resolution of 1920*2780 from the Waterloo IVC 3D Phase I) in the testing stage of all competing methods. All experiments are implemented by MATLAB R2020a and the server of Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz, 32GB RAM, NVIDIA GeForce GTX 1650. The comparison results are shown in Table 5. We can see that the proposed metric proved superior to Chen's metric [36] and Ma's metric [72]. The reason is that for the proposed metric, once the **PVI** and **SD** are calculated, the testing time complexity is very low. Anyway, the proposed metric achieves a low complexity solution to high performance RRSIQA.

V. CONCLUSION

In this study, we propose an RRSIQA method via combining the two features to perform gradient sparse representation and image semantic structure extraction at the initial stage of stereoscopic vision. The gradient sparse representation is used to extract binocular visual information, and the gradient primitive entropy (EGP) of each viewpoint image is used as the monocular cue, and the gradient primitive mutual information (MIGP) between the left and right view's images is used as the binocular cue. The joint statistics of GM and LOG features are taken into account to measure the structural degradation of each view image of distorted stereopair, which is complementary to the monocular cue EGP. The novelty of this study is that we jointly consider the sparsity properties of HVS in gradient domain and the statistical characteristics of image semantic structure. Different from most of the existing SIQA metrics, the proposed metric has two advantages in practical 3D multimedia applications. The one is that the proposed model performs well without the disparity/depth information and traditional NSS-based image transformation, which greatly reduces the complexity of SIQA algorithm. The other is that the proposed model not only requires very few quality-aware features, but also significantly improves consistency with subjective ratings. Thusly, we look forward to further extending this concept to RR 3D video quality measurement in the future work.

ACKNOWLEDGMENT

(Jian Ma and Guoming Xu contributed equally to this work.)

REFERENCES

[1] Z. Wang and A. C. Bovik, "Modern image quality assenment," in *Syntheses Lectures on Image, Video and Multimedia Processing*. San Rafael, CA, USA: Morgan & Claypool, Mar. 2006.

[2] Y. Zhang and D. M. Chandler, "3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3810–3825, Nov. 2015.

[3] Y. Niu, Y. Zhong, W. Guo, Y. Shi, and P. Chen, "2D and 3D image quality assessment: A survey of metrics and challenges," *IEEE Access*, vol. 7, pp. 782–801, 2018.

[4] J. You, L. Xing, A. Perkis, and X. Wang, ‘‘Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis,’’ in *Proc. 5th Int. Workshop Video Process. Quality Metrics Consum. Electron.*, 2010, pp. 1–6.

[5] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, ‘‘Quality assessment of stereoscopic images,’’ *EURASIP J. Image Video Process.*, vol. 2008, Dec. 2009, Art. no. 659024.

[6] B. A. Olshausen and D. J. Field, ‘‘Emergence of simple-cell receptive field properties by learning a sparse code for natural images,’’ *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[7] 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.

[8] E. P. Simoncelli and B. A. Olshausen, ‘‘Natural image statistics and neural representation,’’ *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[9] D. L. Donoho, ‘‘Compressed sensing,’’ *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, ‘‘A survey of sparse representation: Algorithms and applications,’’ *IEEE Access*, vol. 3, pp. 490–530, 2015.

[11] X. Lu and X. Li, ‘‘Group sparse reconstruction for image segmentation,’’ *Neurocomputing*, vol. 136, pp. 41–48, Jul. 2014.

[12] S. Jeong, X. Li, J. Yang, Q. Li, and V. Tarokh, ‘‘Sparse representation-based denoising for high-resolution brain activation and functional connectivity modeling: A task fMRI study,’’ *IEEE Access*, vol. 8, pp. 36728–36740, 2020.

[13] Y. Qi, L. Qin, J. Zhang, S. Zhang, Q. Huang, and M.-H. Yang, ‘‘Structure-aware local sparse coding for visual tracking,’’ *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3857–3869, Aug. 2018.

[14] J. Yang, J. Wright, T. S. Huang, and Y. Ma, ‘‘Image super-resolution via sparse representation,’’ *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[15] J. Yang, B. Jiang, Y. Wang, W. Lu, and Q. Meng, ‘‘Sparse representation based stereoscopic image quality assessment accounting for perceptual cognitive process,’’ *Inf. Sci.*, vol.s. 430–431, pp. 1–16, Mar. 2018.

[16] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, ‘‘Learning sparse representation for no-reference quality assessment of multiply distorted stereoscopic images,’’ *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1821–1836, Aug. 2017.

[17] Y. Zhang, H. Zhang, M. Yu, S. Kwong, and Y. S. Ho, ‘‘Sparse representation-based video quality assessment for synthesized 3D videos,’’ *IEEE Trans. Image Process.*, vol. 29, pp. 509–524, 2020.

[18] M. Aharon, M. Elad, and A. Bruckstein, ‘‘K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,’’ *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[19] W. Shi, F. Jiang, and D. Zhao, ‘‘Image entropy of primitive and visual quality assessment,’’ in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 7674, Sep. 2016, pp. 674–685.

[20] S. Ma, X. Zhang, S. Wang, J. Zhang, H. Sun, and W. Gao, ‘‘Entropy of primitive: From sparse representation to visual information evaluation,’’ *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 249–260, Feb. 2017.

[21] F. Qi, D. B. Zhao, and W. Gao, ‘‘Reduced reference stereoscopic image quality assessment based on binocular perceptual information,’’ *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2338–2343, Oct. 2015.

[22] Z. Wan, K. Gu, and D. Zhao, ‘‘Reduced reference stereoscopic image quality assessment using sparse representation and natural scene statistics,’’ *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 2024–2037, Aug. 2020.

[23] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang, ‘‘Adaptive dictionary learning in sparse gradient domain for image recovery,’’ *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4652–4663, Dec. 2013.

[24] Y. Liu, L. K. Cormack, and A. C. Bovik, ‘‘Statistical modeling of 3-D natural scenes with application to Bayesian stereopsis,’’ *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2515–2530, Sep. 2011.

[25] 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 Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[26] D. Marr and E. Hildreth, ‘‘Theory of edge detection,’’ *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 207, pp. 187–217, Feb. 1980.

[27] D. L. Ruderman, ‘‘The statistics of natural images,’’ *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[28] N. Dalal and B. Triggs, ‘‘Histograms of oriented gradients for human detection,’’ in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[29] C. Harris and M. Stephens, ‘‘A combined corner and edge detector,’’ in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.

[30] H. Bay, T. Tuytelaars, and L. Van Gool, ‘‘SURF: Speeded up robust features,’’ in *Proc. ECCV*, 2006, pp. 886–893.

[31] P. Campisi, P. L. Callet, and E. Marini, ‘‘Stereoscopic images quality assessment,’’ in *Proc. Eur. Signal Process. Conf.*, 2007, pp. 2110–2113.

[32] P. Gorley and N. Holliman, ‘‘Stereoscopic image quality metrics and compression,’’ *Proc. SPIE*, vol. 6830, Feb. 2008, Art. no. 680305.

[33] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz, ‘‘Prediction of stereoscopic video quality using objective quality models of 2D video,’’ *Electron. Lett.*, vol. 44, no. 16, pp. 963–965, Jul. 2008.

[34] M. H. Pinson and S. Wolf, ‘‘A new standardized method for objectively measuring video quality,’’ *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[35] 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.

[36] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, ‘‘Full-reference quality assessment of stereopairs accounting for rivalry,’’ *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1143–1155, Oct. 2013.

[37] 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.

[38] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, ‘‘Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties,’’ *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2917–2983, May 2015.

[39] J. Yang, K. Sim, W. Lu, and B. Jiang, ‘‘Predicting stereoscopic image quality via stacked auto-encoders based on stereopsis formation,’’ *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1750–1761, Jul. 2019.

[40] J. Yang, K. Sim, X. Gao, W. Lu, Q. Meng, and B. Li, ‘‘A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route,’’ *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1314–1328, Mar. 2019.

[41] S. Li, X. Han, and Y. Chang, ‘‘Adaptive cyclopean image-based stereoscopic image-quality assessment using ensemble learning,’’ *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2616–2624, Oct. 2019.

[42] M. Karimi, M. Nejati, S. M. R. Soroushmehr, S. Samavi, N. Karimi, and K. Najarian, ‘‘Blind stereo quality assessment based on learned features from binocular combined images,’’ *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2475–2489, Nov. 2017.

[43] F. Shao, Z. Zhang, Q. Jiang, W. Lin, and G. Jiang, ‘‘Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images,’’ *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 573–585, Mar. 2018.

[44] 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.

[45] J. Ma, P. An, L. Shen, and K. Li, ‘‘Full-reference quality assessment of stereoscopic images by learning binocular visual properties,’’ *Appl. Opt.*, vol. 56, no. 29, pp. 8291–8302, 2017.

[46] Y. Liu, F. Kong, and Z. Zhen, ‘‘Toward a quality predictor for stereoscopic images via analysis of human binocular visual perception,’’ *IEEE Access*, vol. 7, pp. 69283–69291, 2019.

[47] X. Liu, L. Zhang, and K. Lu, ‘‘A 3D image quality assessment method based on vector information and SVD of quaternion matrix under cloud computing environment,’’ *IEEE Trans. Cloud Comput.*, vol. 8, no. 2, pp. 326–337, Apr. 2020.

[48] C. Galkandage, J. Calic, S. Dogan, and J.-Y. Guillemaut, ‘‘Full-reference stereoscopic video quality assessment using a motion sensitive HVS model,’’ *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 452–466, Feb. 2021.

[49] 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.

[50] S. K. Md, B. Appina, and S. S. Channappayya, ‘‘Full-reference stereo image quality assessment using natural stereo scene statistics,’’ *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1985–1989, Nov. 2015.

[51] 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.

[52] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh, "No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models," *Signal Process., Image Commun.*, vol. 54, pp. 107–117, May 2017.

[53] X. Wang, Q. Liu, R. Wang, and Z. Chen, "Natural image statistics based 3D reduced reference image quality assessment in contourlet domain," *Neurocomputing*, vol. 151, no. 2, pp. 683–691, Mar. 2015.

[54] L. Ma, X. Wang, Q. Liu, and K. N. Ngan, "Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment," *Neurocomputing*, vol. 215, pp. 21–31, Nov. 2016.

[55] 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 Trans. Image Process.*, vol. 28, no. 10, pp. 5027–5040, Oct. 2019.

[56] 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.

[57] 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.

[58] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4923–4936, Oct. 2017.

[59] Q. Jiang, W. Zhou, X. Chai, G. Yue, F. Shao, and Z. Chen, "A full-reference stereoscopic image quality measurement via hierarchical deep feature degradation fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9784–9796, Dec. 2020.

[60] W. Zhou, Z. Chen, and W. Li, "Dual-stream interactive networks for no-reference stereoscopic image quality assessment," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3946–3958, Aug. 2019.

[61] G. Sun, B. Shi, X. Chen, A. S. Krylov, and Y. Ding, "Learning local quality-aware structures of salient regions for stereoscopic images via deep neural networks," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2938–2949, Nov. 2020.

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

[63] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[64] W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," *Science*, vol. 287, no. 5456, pp. 1273–1276, 2000.

[65] W. Shi, F. Jiang, and D. Zhao, "Image entropy of primitive and visual quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2087–2091.

[66] A. D. D'Antona, J. S. Perry, and W. S. Geisler, "Humans make efficient use of natural image statistics when performing spatial interpolation," *J. Vis.*, vol. 13, no. 14, pp. 1–13, 2013.

[67] J. Ma and Y. Zhang, "A visual perceptual Bayesian theory for stereoscopic images' quality assessment," *IEEE Photon. Technol. Lett.*, vol. 30, no. 20, pp. 1788–1791, Oct. 15, 2018.

[68] 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.

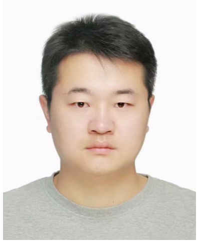
[69] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

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

[71] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[72] J. Ma, P. An, L. Shen, and K. Li, "Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation," *IEEE Access*, vol. 6, pp. 2768–2780, 2018.

[73] J. Ma, X. Zhao, and Y. Xu, "Reduced-reference stereoscopic image quality assessment based on entropy of gradient primitives," in *Proc. IEEE 5th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 206–209.



JIAN MA received the Ph.D. degree in communication and information engineering from Shanghai University, Shanghai, China, in 2018. He was a Lecturer with the Institute of Logistics Science and Engineering, Shanghai Maritime University, from August 2018 to December 2019. Since January 2020, he has been an Assistant Professor with the School of Internet, Anhui University. Since October 2020, he has also been holding a postdoctoral position with the School of Computer Science, Fudan University. His research interests include 3D image/video quality assessment, multimedia computing, and deep learning.

GUOMING XU received the Ph.D. degree in signal and information processing from the Hefei University of Technology, in 2015. He held a postdoctoral position with the Army Artillery and Air Defense Forces Academy of PLA, from 2017 to 2019. He joined as a Professor with the Video Processing Group, School of Internet, Anhui University, China, in 2019. He has authored over 40 technical articles in refereed journals and proceedings in image and video sparse representation, image super resolution, and target detection. His research interests include signal sparse representation, image super resolution, and target detection.

XIYU HAN received the B.E. degree from Chizhou University, Chizhou, China, in 2018. He is currently pursuing the M.S. degree in signal and information processing with the Shool of Internet, Anhui University, Heifei, China. His research interests include signal sparse representation, image super resolution and deep learning.