IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 19, 2017, accepted December 11, 2017, date of publication December 19, 2017,
date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2785282

Reduced-Reference Stereoscopic Image Quality Assessment Using Natural Scene Statistics and Structural Degradation

JIAN MA, (Student Member, IEEE), PING AN, (Member, IEEE), LIQUAN SHEN, (Member, IEEE), AND KAI LI, (Member, IEEE)

Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
Key Laboratory of Advance Displays and System Application, Ministry of Education, Shanghai University, Shanghai 200444, China
School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

Corresponding author: Ping An (anping@shu.edu.cn)

This work was supported by the National Natural Science Foundation of China, under Grant U1301257, Grant 61172096, Grant 61571285, and Grant 61601278.

ABSTRACT Perceptual stereo image quality assessment (SIQA) aims to design computational models to measure the stereo image quality in accordance with human opinions. In this paper, a novel reduced-reference (RR) SIQA is proposed by characterizing the statistical and perceptual properties of the stereo image in both the spatial and gradient domains. To be specific, in the spatial domain, we extract the parameters of the generalized Gaussian distribution fits of luminance wavelet coefficients to form the underlying features. In the gradient domain, the modified gradient magnitudes maps are generated by jointly considering human visual system's contrast sensitivity and neighborhood gradient information to weight the gradient magnitudes in a locally adaptive manner. Afterward, perceptual features are extracted based on the entropy of discrete wavelet transform coefficients of modified gradient magnitudes. Furthermore, we consolidate the left and right features into a single set of features per stereo image pair. Finally, the qualities of both the spatial and gradient domains are combined to obtain the overall quality of stereo image. Extensive experiments performed on popular data sets demonstrate that the proposed RR-SIQA method achieves highly competitive performance as compared with the state-of-the-art RR-SIQA models as well as full-reference ones for both symmetric and asymmetric distortions.

INDEX TERMS Reduced reference, stereo image quality assessment (SIQA), natural scene statistics, discrete wavelet transform, gradient magnitudes.

I. INTRODUCTION

With the rapid development of the internet and modern information technologies, people are in the information explosion era. There is an old saying that one picture means a thousand words. Thusly, the image plays a big role on the information transmitting, and influences the people's daily life deeply and thoroughly. Meanwhile, as the ultimate receivers of any visual information, people gradually incline to pursue multimedia contents with higher quality of experiences (QoE) in the traditional entertainment areas such as television pictures and video games, as well as some more specialized applications, including medical care, business and education. Therefore, image quality assessment (IQA) has recently become a hot research topic in the field of image processing [1]. In general, subjective assessment is regarded as the most reliable and accurate way to evaluate the image quality, but subjective assessment is often costly, slow, and difficult to integrate into real-time image processing systems [2]. For this reason, it is necessary to design objective IQA methods to automatically evaluate and control the perceptual quality of the image.

During the last decade, research on the traditional 2D IQA methods have been extensively studied and many representative works have been proposed, such as structural similarity (SSIM) [3], gradient similarity (GSIM) [4], feature similarity (FSIM) [5], video integrity oracle (VIIEDO) [6], RRED indices (RRED) [7], RDCT [8] and so on. Recently, stereo image for three-dimensional (3D) viewing has become a significant contributor to entertainment industry and consumer electronic market, and the demand for stereo image quality assessment (SIQA) has subsequently drawn a high

level of attention from not only researchers but also the industries as well [9]. However, SIQA is still a more complicated and challenging issue than its 2D counterparts with consideration of both the 2D image quality and depth perceptual factors (i.e., additional depth cues [10], crosstalk [11], binocular rivalry [12], and visual discomfort [13]).

Similar to objective 2D IQA methods, objective SIQA methods can be divided into three categories depending on the amount of reference information provided to the computational model. Specifically, full-reference (FR) SIQA metrics operate on a distorted stereo image with an original stereo image available for comparison, while no-reference (NR) SIQA metrics predict the quality of distorted stereo images without any knowledge of reference information. As a compromise between the FR and NR SIQA methods, reduced-reference (RR) SIQA metrics make use of partial information or features extracted from the reference stereo image to estimate the quality of distorted stereo images [9]. In this paper, the discussion is focused on RR SIQA model, which is widely used to act as a guide to optimize 3D content production.

Most of FR SIQA methods usually return the desired results and perform well in prediction the quality scores of human opinions in comparison with RR and NR SIQA methods. However, in most present and emerging visual communication environments, the reference stereo images are not fully accessible. Without any information about the reference stereo image, the SIQA task becomes very difficult. NR SIQA can mainly be grouped into two classes: distortion-specific (such as blocking, blurring, and compression) [14], [15] and general-purpose metrics (distortion-agnostic) [16]–[19]. Distortion-specific metrics are effective for specific distortion when the distortion information is known, while general-purpose metrics work for different distortions. Since the type of distortion is usually unknown and the performance of general-purpose NR SIQA metrics are very limited, they may not be mature enough to many practical 3D applications.

RR SIQA methods provide a solution to address the reference stereo image inaccessible restrictions. These methods generally include a feature extraction process at the sender sider for the reference stereo image and a feature extraction at the receiver side for the distorted stereo image. The key point of RR SIQA methods is to find a balance between the data rate of RR features and the accuracy of stereo image quality prediction. That is, the extracted RR features usually have a much lower data rate than the stereo image data and are typically transmitted to the receiver side through an ancillary channel.

In designing RR SIQA algorithms it is important to select the RR features in a way to be sensitive to different distortions and be closely related to the visual perception of stereo image quality. There are two main directions for RR SIQA metrics. One is stereo image feature description based on natural scene statistics (NSS). The other direction is stereo image feature description based on the visual properties of human visual system (HVS). Several studies have been conducted for RR SIQA in recent years [20]–[25]. For NSS-based RR SIQA, Wang *et al*. [20] proposed a novel 3D RRIQA metric based on 3D NSS in contourlet domain. Ma *et al.* [21] proposed a RR SIQA model by characterizing the statistical properties of stereo image in the reorganized discrete cosine transform (RDCT) domain. For HVS-based RR SIQA, Qi *et al*. [22] proposed a RR SIQA model based on binocular perceptual information (BPI). BPI is represented by the distribution statistics of visual primitives in left and right view's image, which are extracted by sparse coding and representation. Xu *et al*. [23] proposed a RR SIQA model through measuring structural degradation and saliency based parallax compensation model (SSPM). Zheng *et al*. [24] proposed a RR SIQA model based on binocular perceptual properties of HVS. Hewage and Martini [25] predicted the quality of color plus depth 3D video by using the extracted edge information of depth maps and extracted information from the corresponding color image in the areas approaching edges. Even though the above-mentioned algorithms can provide a useful and effective way to evaluate the quality of distorted stereo image, the performance is still limited due to only single strategy adopted. According to [26], Liu *et al*. have indicated that the single method does not provide the best performance in all situations, and proposed a multimethod fusion approach to overcome the deficiency when only one method is adopted.

Based on the above analysis, in this paper, we propose a training-free general-purpose RR SIQA metric which combines NSS-based features with HVS-based features. Our motivation stems from the HVS is highly adapted for extracting structural information from the input scene [3], and the marginal distributions of luminance wavelet coefficients are modelled well by the generalized Gaussian density (GGD) [27]; The main contributions of this work are threefold. (1) The selected features are comprehensive and not limited to NSS-based or HVS-based; (2) According to the perceptual properties of HVS, we consolidate the left and right HVS-based features into a single set of features per stereo image pair by uncertainty weighting; (3) The new RR SIQA can be used in general purpose without training and tuning. Experimental results on LIVE 3D phase I [28] and LIVE 3D phase II [29] demonstrate that the proposed metric is efficient and promising compared with the state-of-the-art methods.

This paper is organized as follows. An overview of related work is given in section II. Section III presents the proposed RR SIQA method. Experimental results are drawn in section IV, followed by a conclusion in section V.

II. RELATED WORK

Owing to the thriving market of stereo image based various of applications, efficient and effective SIQA techniques become colossally required these days. Existing perceptual quality metrics for stereo images can be grouped into three categories, named quality model extended from 2D IQA methods, the quality model designed by considering the properties of

HVS, and the quality model proposed by using the regularities of NSS.

In earlier research, since a 3D image consists of two view's 2D images, many SIQA metrics is simply extended from the existing 2D IQA methods. These models handle each view of stereo image independently. The overall quality index of the stereo image is determined by integrating the obtained quality index of each view image. For example, Campisi *et al.* [30], Gorley and Holliman [31], Benoit *et al.* [32] and You *et al.* [33] utilized a straightforward way to apply the existing 2D IQA methods in their SIQA metrics. Hewage *et al.* [34] studies the correlation between subjective opinions and three quality metrics, including PSNR, Video Quality Model (VQM) [35] and SSIM [3] for stereo video content. Such models handle each view image independently without consider the binocular vision properties. Obviously, the simple combination of two views' image quality is not consistent with human opinions.

Aiming to further optimize the performance of SIQA metrics, the specific properties of HVS, such as the binocular vision and depth perception properties are taken into account for SIQA design. Bensalma and Larabi [36] proposed a SIQA metric based on binocular energy perception. Chen *et al.* [29] proposed a ''cyclopean image'' model to account for binocular rivalry. Shao *et al.* [37] proposed a new FR-SIQA metric by learning binocular receptive field properties. In [38], a FR-SIQA metric based on monocular and binocular visual perception was presented. Shao *et al.* [39] proposed a perceptual quality assessment approach for stereoscopic images by modeling visual properties of the primary visual cortex. Likewise, Shao *et al.* [40] developed a new NR-SIQA by using joint sparse representation. Lin and Wu [41] proposed a quality index for stereo images by using the binocular combination and the binocular frequency integration as the bases. Zhou *et al.* [42] proposed an efficient NR-SIQA metric that utilizes binocular vision-based dictionary learning (DL) and k-nearest-neighbors (KNN)-based machine learning (ML) to more accurately align with human opinions. However, the aforementioned metrics depend on the entire original stereo image or depth/disparity information to evaluate the quality of the distorted stereo image, they are regarded as very complex, and not suitable for real-time 3D applications. Besides, such models only consider the properties of HVS for SIQA design, which are incompetent to accurately reflect the change caused by different distortions.

Recently, the statistical properties of stereo images have been extensively studied, several SIQA models have been proposed. For example, Chen *et al.* [16] extracted 2D and 3D features from the ''cyclopean image'' and the ''uncertainty image'' based on NSS for SIQA. Su *et al.* [17] proposed a NR SIQA metric by extracting bivariate and correlation NSS features from distorted stereoparis, which dubbed stereoscopic/3D blind image naturalness quality index (S3D-BLINQ Index). Md *et al.* [18] utilized the parameters of GDD fits of luminance wavelet coefficients along with correlation values form excellent features for SIQA. Hachicha *et al.* [19] proposed a NR SIQA based on joint wavelet decomposition and GGD models. However, these models need to perform the disparity map estimation (the computational complexity is high), which is not suitable for real-time image processing system. What's more, HVS perception properties of stereo images have not been adequately considered, which significantly affect the quality analysis of stereo image. To devise an efficient RR SIQA metric, we have explored the fusion features which contained NSS-based features and HVS-based features to build the new quality assessment model for stereo image.

III. PROPOSED RR-SIQA INDEX

The framework of our proposed RR SIQA metric is illustrated in Fig. 1. We consider not only the statistical properties of the stereo image in the spatial domain but also the perceptual properties of HVS in gradient domain. At the sender side, in space domain, each view image of stereo image is first converted to grayscale images $I^V, V \in \{L, R\}$ (L refers to left image, R refers to right image). Subsequently, the original grayscale images are decomposed to subbands with L-level by discrete wavelet transform (DWT), and the parameters of GGD fits of luminance wavelet coefficients are extracted to form the NSS-based features. Moreover, in gradient domain, each view image of stereo image is first processed by applying a contrast sensitivity function (CSF) based filter resulting in the filtered images $I^V_{CF}, V = \{L, R\}$. Then, the gradient magnitude maps, $I^V_{GM}, V = \{L, R\}$ are computed and locally weighted to generate modified gradient magnitude maps $\hat{I}^V_{GM}, V \in \{L, R\}$. An L-level DWT of the $\hat{I}^V_{GM}$ are performed, and the entropy of DWT coefficients of the $\hat{I}^V_{GM}$ are computed as the HVS-based features. Note that, details about the parameters of GGD fits, CSF-based filtering, local adaptive weighting of gradient magnitudes, DWT and entropy computations are given below. At the receiver side, likewise, the distorted stereo images are processed by the same procedure at the sender side for the statistical and perceptual features. Besides, we consolidate the left and right features into a single set of features per stereo image pair. Finally, the stereo image qualities of both space and gradient domains are combined to yield the overall quality of stereo image. We describe our approach in the following.

A. NATURAL SCENE STATISTICS-BASED FEATURES

In the previous studies, there has been much attention recently to the statistics of natural images. The statistical properties of natural scenes have inspired scientific and engineering studies to understand and simulate visual perception by designing artificial visual systems (i.e., image quality assessment [16]–[19], image compression [43], and image denoising [44], [45]). Our work is inspired by the statistical properties of marginal distributions of luminance wavelet coefficients which can be modelled well by the GGD model [27]. According to [27], one popular approach is to model the wavelet coefficients distributions using a

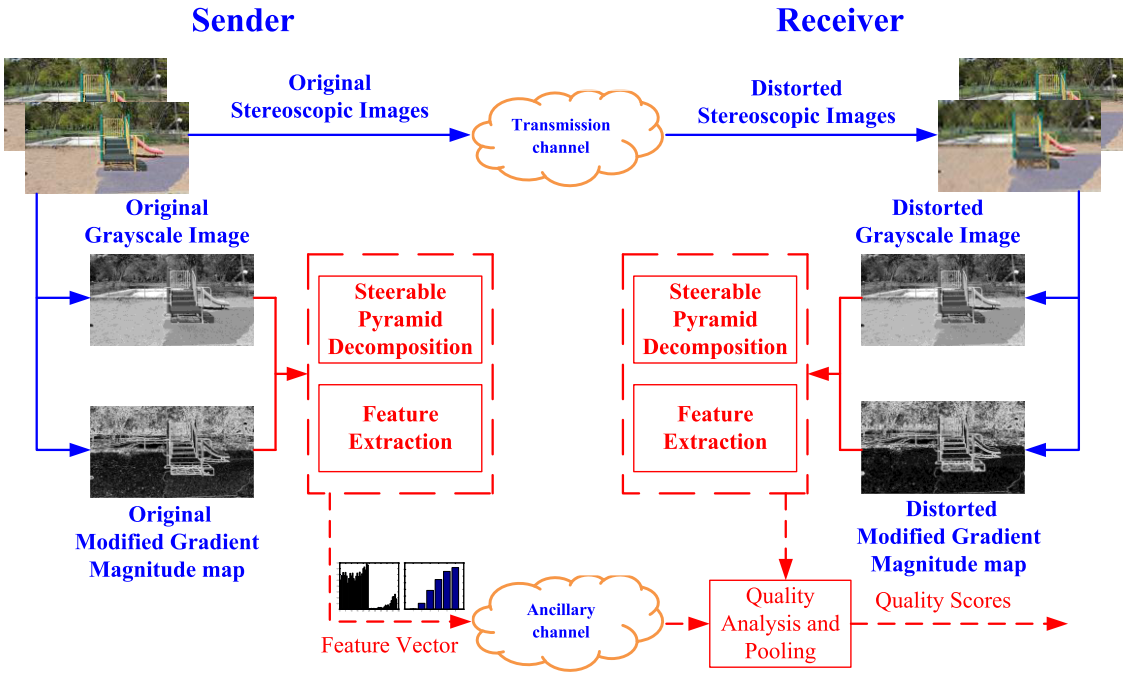

FIGURE 1. The framework of the proposed RR-SIQA metric.

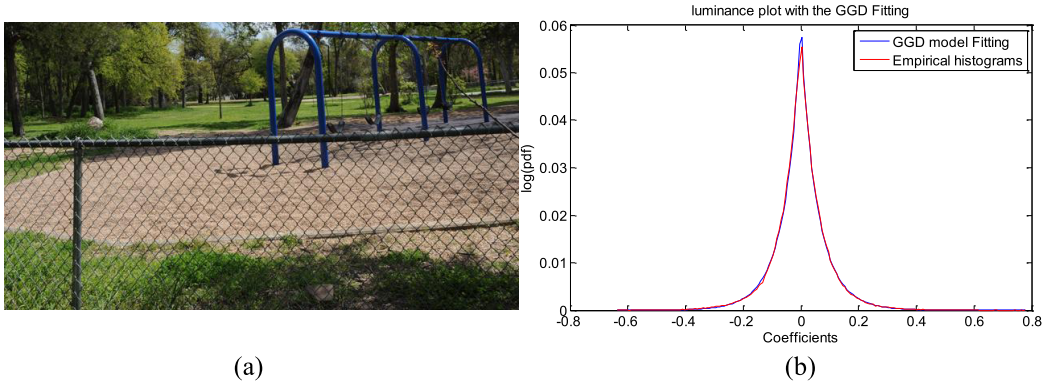

FIGURE 2. The coefficient distribution of a luminance wavelet subband (first scale, 0°) of natural image and its GGD modeling curve. (a) natural image. (b) luminance plot with the GGD fitting.

generalized Gaussian distribution:

$$f(x) = \frac{e^{-|x/s|^g}}{Z(x, g)} \tag{1}$$

where $Z(x, g)$ is a normalizing constant to force the integral of $f(x)$ to be 1, g controls the geometric shape and s controls the spread. Note that, $f(x)$ is Gaussian when $g = 2$, and Laplacian when $g = 1$. As shown in Fig. 2, we can observe that the wavelet coefficients of natural images yield very good fits between empirical histograms and the GGD model fitting. In this work, we illustrate how distortion affect the luminance wavelet coefficients distributions using the GGD model. From Fig. 3, it can be seen clearly from these plots that the log-histograms of original image and its five distorted images have shape peaks and heavy tails and therefore corroborate GGD modelling. In addition, the distortions cause the histograms (and hence the GGD model parameters (g, s)) to change significantly with respect to original image statistics. Therefore, we can conclude that the GGD model parameters serve as good discriminatory features for SIQA task.

To capture above-mentioned NSS-based features of stereo images, we have applied a steerable pyramid decomposition [46], [47] whose space-scale-orientation decomposition models the band pass filtering that occurs in area V1 of primary visual cortex [48]. This decomposition is performed at four scales (fine to coarse) and six orientations (0°, 30°, 60°, 90°, 120°, 150°) to analyze stereo

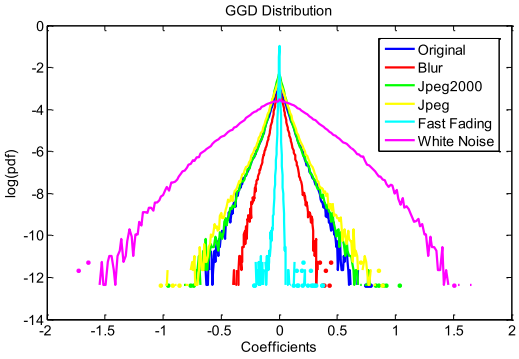

FIGURE 3. The log-histograms of original image and its distorted image subbands.

images. Afterwards, GGD model fitting is performed on each of the luminance subband histograms for both the original and distorted stereo images. We utilize the method in [49] for estimating the GGD model parameters (g, s). Following the formula in (1), a pair of (g, s) values are computed for each of the 24 luminance subband pairs and denoted by (g_{l_j}, s_{l_j}), (g_{r_j}, s_{r_j}). Note that, the subscript l and r denote the left and right subband respectively and j is the subband index ($1 \leq j \leq 24$). With the left and right features, we consolidate them into a single set of features per stereo image pair. This consolidate is motivated by the fact that binocular strength is a convex sum of monocular strengths [50]. That is, the convex weights of left and right views rely on relative eye dominance that is particularly visible in the case of asymmetric distortion. As we have 24 subbands of each image, the stimulus strength is directly bound up with band strength of each subband. According to [18], we adopt root mean square (RMS) subband value as its band strength. Here, band strenghs of the j^{th} subband of left and right images are denoted by p_{l_j} and p_{r_j} respectively. Therefore, the convex weights of left and right views are calculated as follows:

$$w_{l_j} = \frac{p_{l_j}}{p_{l_j} + p_{r_j}} \tag{2}$$

$$w_{r_j} = \frac{p_{r_j}}{p_{l_j} + p_{r_j}} \tag{3}$$

The normalized weights are then used for feature vector consolidation as follows:

$$g_j = w_{l_j} g_{l_j} + w_{r_j} g_{r_j} \tag{4}$$

$$s_j = w_{l_j} s_{l_j} + w_{r_j} s_{r_j} \tag{5}$$

As a result, the NSS-based features consist of one g_j and s_j value of j^{th} subband. As there are 24 subbands, we have obtained 24-length vectors $\mathbf{g} = [\mathbf{g_1}, \ldots \mathbf{g_{24}}]^{\mathbf{T}}$ and $\mathbf{s} = [\mathbf{s_1}, \ldots \mathbf{s_{24}}]^{\mathbf{T}}$ per stereo pair.

B. HUMAN VISUAL SYSTEM-BASED FEATURES

Apart from the statistical properties of natural images for IQA, there are many HVS oriented theories about IQA metrics in the past few years. The one most widely accepted today is based on the assumption that the HVS is highly adapted for extracting structural information from the input scene [3]. Moreover, the gradient of an image describes its geometric features which conveys important visual information for understanding. The gradient-relevant methods have gotten big success in IQA [5], [51], [52]. In [53], it has been discovered that there are some inherent limitations with respect to the visibility of stimuli, the HVS is not equally sensitive to all stimuli. Inspired by these previous works, we jointly consider human visual system's contrast sensitivity and neighborhood gradient information to weight the gradient magnitudes in a locally adaptive manner. Then, perceptual features are extracted based on the entropy of DWT coefficients of modified gradient magnitudes. The computational process of the HVS-based features is shown in Fig. 4. Detailed implementation will be given in the following subsection.

1) CSF BASED FILTERING

According to [53], the binocular visual sensitivity to stimulus at different spatial frequencies is different which could be modeled by an empirical CSF. In this paper, we consider variations in sensitivity to spatial frequency by applying the adjusted CSF filtering to both the reference and distorted stereopairs. Here this adjusted CSF model [52] is given by:

$$H_{CSF}(f, \theta) = H_a(f, \theta) H_b(f, \theta) \tag{6}$$

where f denotes the radial spatial frequency in cycles per degree of visual angle (c/deg), $\theta \in [-\pi, \pi]$ denotes the orientation. In equation (6), $H_a(f, \theta)$ is the frequency response of a CSF model originally introduced by Mannos and Sakrison [54] with adjustments specified by Daly [55] and $H_b(f, \theta)$ is the frequency response of a circularly symmetric Gaussian filter. Therefore, $H_a(f, \theta)$, is denoted by:

$$H_a(f, \theta) = \begin{cases} 2.6(0.0192 + \lambda f_\theta) \exp[-(\lambda f_\theta)^{1.1}], & \text{if } f \geq f_{peak}\, c/\deg \\ 0.981 & \text{otherwise} \end{cases} \tag{7}$$

where $f_\theta = f/[0.15\cos(4\theta + 0.85)]$ accounts for the oblique effect. Note that, we set $\lambda = 0.114$, resulting in a peak at a frequency of $f_{peak} \approx 8c/\deg$. The frequency response of the Gaussian filter, $H_b(f, \theta)$, is denoted by:

$$H_b(f, \theta) = \exp\left(-2\pi^2 \sigma^2 f^2\right) \tag{8}$$

where σ is a parameter that controls the cutoff of the filter. In our experiments, in order to filter out components of very high frequency that are perceptually insignificant without significantly affecting the perceptually relevant edge components, we set a 5×5 filter window size and a small $\sigma = 0.5$, resulting in a lowpass filter with a relatively very high cutoff frequency.

Suppose we apply the adjusted CSF filtering to the left or right image of stereo images $I^V, V \in \{L, R\}$, This

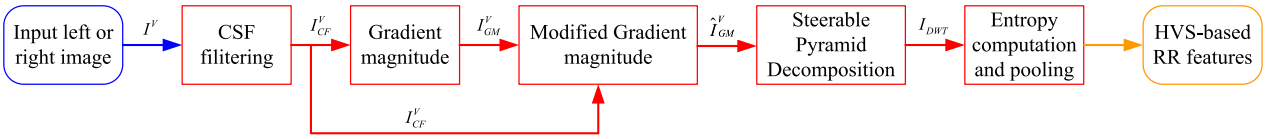

FIGURE 4. Block diagram illustrating the computation of the HVS-based features for the proposed RR SIQA metric.

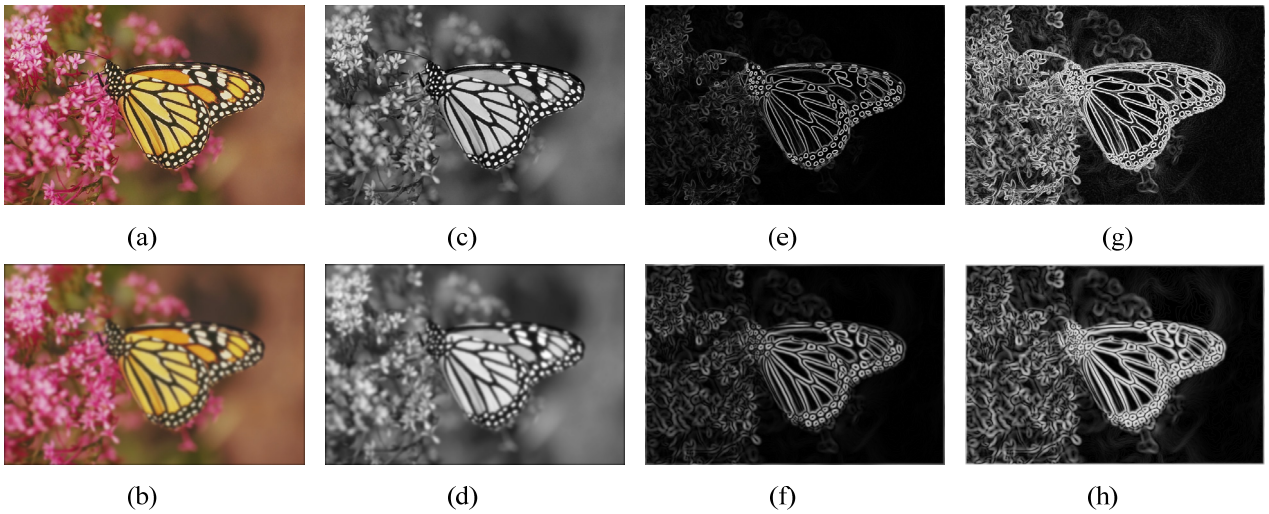

FIGURE 5. Illustration of steps of the modified gradient images. (a) Original input images; (b) Distorted input images; (c) and (d) CSF filtered images; (e) and (f) Gradient magnitude images; (g) and (h) Modified gradient magnitude images.

adjusted CSF filtering is performed in the frequency domain via:

$$I^V_{CF} = F^{-1}\left(\hat{H}_{CSF}(u,v) \times F\left(I^V\right)\right) \tag{9}$$

where u, v are the DFT indices, and $F[\cdot]$ and $F^{-1}[\cdot]$ denote the DFT and inverse DFT, respectively. The quantity $\hat{H}_{CSF}(u,v)$ denotes a DFT based version of $H_{CSF}(u,v)$.

2) MODIFIED GRADIENT MAGNITUDES

From a computational point of view, the responses of classical cortical receptive fields in the simple cells at different scales and orientations can be modeled by Gaussian derivative functions [56], and can be represented using gradient magnitudes. Moreover, it has been discovered that edges of an image are perceptually significant to the HVS, and the gradient of an image corresponding to the edge is crucial importance to observers. Therefore, at this stage, we calculate the gradient magnitude I^V_{GM} of the stereo image I^V_{CF} resulted from the previous CSF-based filtering stage. For simplicity, I^V_{GM} is computed by using the Sobel operator [57]. However, since the effect of luminance masking and contrast/texture masking directly influence the visibility of distortions, it is not sufficient only by using gradient information to estimate the image quality accurately. In order to enhance the local image structure while removing the contrast variations, we compute a locally weighted gradient image that utilize the information in I^V_{CF} and I^V_{GM} jointly to consider the effect of local background luminance (given by information in I^V_{CF}) as well as local image structure (given by information in I^V_{GM}) on distortion visibility [52]. To be specific, the modified gradient magnitude image is calculated as follows:

$$\hat{I}^V_{GM} = \frac{I^V_{GM}(i,j)}{\beta(i,j) + \varepsilon} \tag{10}$$

$$\beta(i,j) = \sqrt{\sum\sum_{(i',j')\in\Omega_{i,j}} I^V_{NF}(i',j')\, w(i',j')} \tag{11}$$

$$I^V_{NF}(i,j) = \frac{I^{V2}_{GM} + I^{V2}_{CF}}{2} \tag{12}$$

$$w(i,j) = \frac{K(i,j)}{\sum_{i,j} K(i,j)} \tag{13}$$

where ε is a small positive constant to avoid numerical instability when $\beta(i,j)$ has a small value. $\Omega_{i,j}$ is a local window centered at (i,j), and $w\left(i',j'\right)$are positive symmetric weights satisfying $\sum_{i',j'} w\left(i',j'\right) = 1$ via equation (13). $K(i,j)$ denotes a Gaussian kernel coefficient at (i,j) with window size 5×5 and $\sigma = 0.5$. Examples of modified gradient magnitude images are shown in Fig. 5 (g) and (h). Fig. 6 shows that the gradient magnitude map without the local weighting operator fails in effectively capturing certain types and levels of distortions. Moreover, from Fig. 6 (b), it is quite clear that the gradient magnitude maps of the input images have not changed significantly under different types and levels of distortions (Fig.6 (b) and close-ups illustrated in Fig. 6 (d), (f), (h), (j), (l), (n)), while the modified

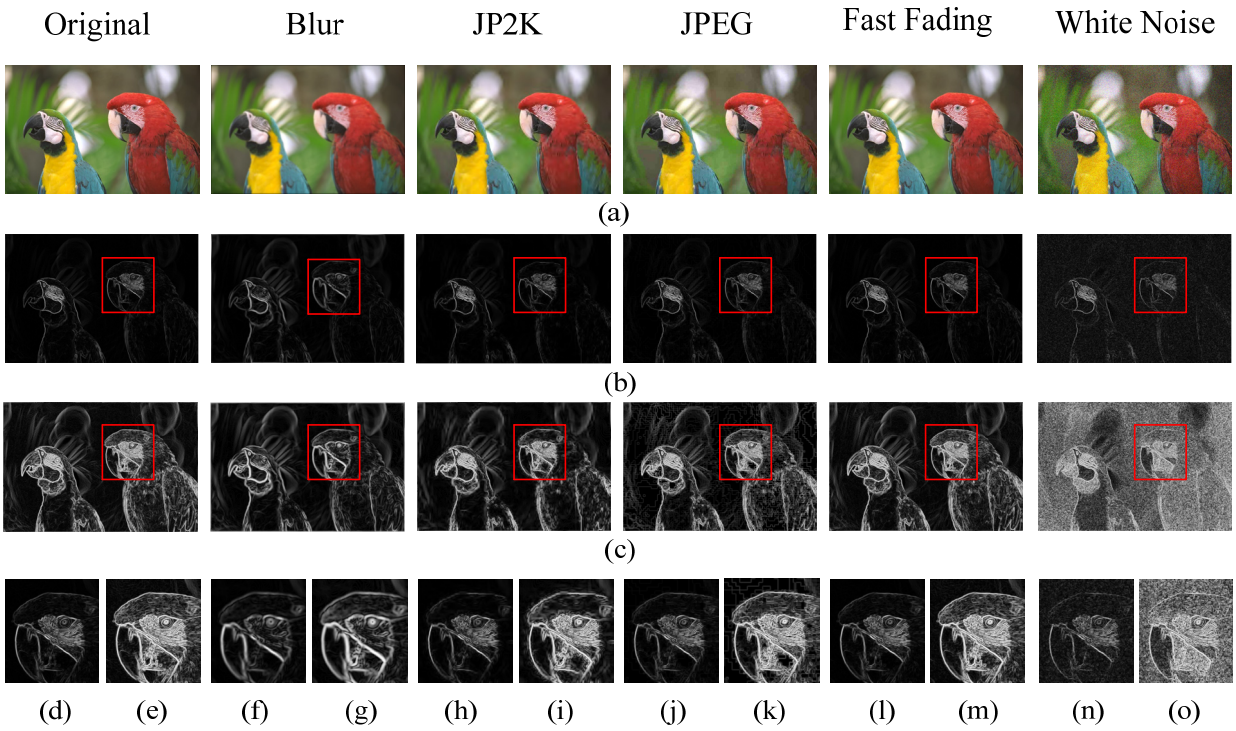

FIGURE 6. Comparison between gradient magnitude and modified gradient magnitude maps. (a) First row represents the input image (original image and its five distorted versions). (b) Second row represents the corresponding gradient images of the first row input images. (c) Third row represents the corresponding modified gradient magnitude images. Close-up images of gradient images and modified gradient images are given in the fourth row, where (d), (f), (h), (j), (l), and (n) are the corresponding gradient images in the second row, and (e), (g), (i), (k), (m), and (o) are the corresponding modified gradient magnitude images in the third row.

gradient magnitude maps (Fig.6 (c) and close-ups illustrated in Fig. 6 (e), (g), (i), (k), (m), (o)) can more clearly reflect the local changes in the image structures caused by different distortion types and levels.

3) DWT COEFFICIENTS COMPUTATION

In [58], it has been discovered that wavelet transform provides a convenient way for localized representation of signals simultaneously in space and frequency. They have been widely used to model the processing in the early stages of HVS and have also become the popular form of representations for IQA metrics [18], [19], [52]. Therefore, in this paper, a steerable pyramid wavelet decomposition is used to decompose the modified gradient magnitude maps ($\hat{I}^V_{GM}$) into components at six scales (fine to coarse) with four orientations $(0°, 45°, 90°, 135°)$ at each scale in order to extract perceptual features that are sensitive to different image distortions.

4) ENTROPY COMPUTATION AND POOLING

Information entropy, which is a fundamental measure of the uncertainty associated with a probability distribution, is usually used to compute the uncertainty of shape and structure information of an image [59]. In general, images with single orientation and simple shapes have a low entropy, while images with varying edge orientations and complex shapes have a relatively high entropy. In this work, we calculate the entropy of the wavelet coefficients of $\hat{I}^V_{GM}$ at each scale and orientation in order to efficiently capture the image structure information. The pooled entropies of the left and right images across orientations at each scale are utilized as HVS-based features and to evaluate the perceptual quality of stereo image.

Let $H^V_{S,O}, V \in \{L, R\}$ (L and R denote the left and right images of stereo image, respectively) denote the computed entropy value of the DWT coefficients of each view of stereo image, after removing the mean value and converting to unsigned 8-bit integers, for the S^{th} scale and O^{th} orientation ($I^{V_{S,O}}_{DWT}$). Meanwhile, let $I^{V_{S,O,K}}_{DWT}, K = 1, \ldots, M_{S,O}$, denote all the values of $I^{V_{S,O}}_{DWT}$. The entropy of $I^{V_{S,O}}_{DWT}$ is defined by:

$$H^V_{S,O} = -\sum_{K=1}^{M_{S,O}} P\left(I^{V_{S,O,K}}_{DWT}\right) \log P\left(I^{V_{S,O,K}}_{DWT}\right) \tag{14}$$

where P denotes the probability density function associated with $I^{V_{S,O}}_{DWT}$. Subsequently, the entropies $H^V_{S,O}$ are pooled at each scale $S, S = 1, \ldots N_S$ along the orientations as follows:

$$H^V_S = \sum_{O=1}^{N_O} \log\left(1 + H^V_{S,O}\right) \tag{15}$$

where H_S^V is the pooled value of entropies over all the orientation of the S_{th} scale.

After computing the entropies of left image H_S^L and the entropies of right image H_S^R as Eqs. (14) and (15), we consolidate them into a single set of features per stereo image pair by uncertainty weighting as follows:

$$H_S = W_L H_S^L + W_R H_S^R \tag{16}$$

$$W_L = \frac{\hat{H}_S^L}{\hat{H}_S^L + \hat{H}_S^R} \tag{17}$$

$$W_R = \frac{\hat{H}_S^L}{\hat{H}_S^L + \hat{H}_S^R} \tag{18}$$

where $\hat{H}_S^L$ and $\hat{H}_S^R$ are pooled at each scale S, $S = 1, \ldots N_S$ along the orientations of left and right images of stereo image, respectively. They are computed as follows:

$$\hat{H}_S^V = \sum_{O=1}^{N_O} H_{S,O}^V \tag{19}$$

where $V \in \{L, R\}$ (L and R denote the left and right images of stereo image, respectively). The total number of RR HVS-based features is equal to the number of scales N_S; that is, $\mathbf{H_S} = \left[\mathbf{H_{S_1}}, \ldots \mathbf{H_{S_{N_S}}}\right]^{\mathbf{T}}$. Since we decompose the modified gradient magnitude maps ($\hat{I}_{GM}^V$) into components at six scales (fine to coarse) with four orientations at each scale, resulting in only six RR HVS-based features per stereo pair.

C. QUALITY INDEX COMPUTATION AND POOLING

In the receiver side, the extracted features of original and distorted stereo image are compared to evaluate the stereo image perceptual quality. First, the following two quality indexes will be determined to measure the perceptual quality degradation caused by the introduced distortion.

1) QUALITY INDEX IN SPACE DOMAIN

Based on the extracted NSS features, the quality index in space domain is computed by using the standard Wave Hedges distance [60], [61] metric as follows:

$$Q_s\left(\mathbf{D^r_{NSS}}, \mathbf{D^d_{NSS}}\right) = \sum_{NSS_i=1}^{24} \frac{\left|\mathbf{D^r_{NSS_i}} - \mathbf{D^d_{NSS_i}}\right|}{\max\left(\mathbf{D^r_{NSS_i}}, \mathbf{D^d_{NSS_i}}\right)} \tag{20}$$

where $\mathbf{D^r_{NSS}} = \left[\mathbf{g^r}; \mathbf{s^r}\right]^{\mathbf{T}}$, $\mathbf{D^d_{NSS}} = \left[\mathbf{g^d}; \mathbf{s^d}\right]^{\mathbf{T}}$ are the NSS feature vectors of stereo images. Note that, the superscript r and d denote the reference and distorted stereo pairs respectively.

2) QUALITY INDEX IN GRADIENT DOMAIN

Let $\mathbf{H^r_S}$ and $\mathbf{H^d_S}$ represent pooled entropies of the reference and distorted image at the S^{th} scale, respectively. We denote the quality index Q_g in gradient domain, which is computed as follows:

$$Q_g = \log\left(\sum_{S=1}^{N_S}\left(\mathbf{H^r_S} - \mathbf{H^d_S}\right)^2\left(T^d + 1\right) + 1\right) \tag{21}$$

$$T^d = \sum_{S=1}^{N_S}\sum_{O=1}^{N_O} T_{S,O}^d \tag{22}$$

$$T_{S,O}^d = W_L T_{S,O}^{d_L} + W_R T_{S,O}^{d_R} \tag{23}$$

where N_S and N_O denote the number of scales and orientation, respectively. $T_{S,O}^{d_L}$ and $T_{S,O}^{d_R}$ are the mean absolute value of DWT coefficients at the S^{th} scale and O^{th} orientation for the left and right images of the distorted stereo image. Note that, $T_{S,O}^d$ is computed based on Eqs. (17), (18) and (23).

Given the quality indexes Q_s and Q_g, the final perceptual quality of the distorted stereo image Q is determined as

$$Q = \log\left(\left(Q_s \times Q_g\right) + 1\right) \tag{24}$$

IV. EXPERIMENTAL RESULTS

In this section, the performance of our proposed RR SIQA index is analyzed in terms of its ability to evaluate subjective ratings of image quality for both symmetric and asymmetric distortions of stereo images. We evaluate the performance in terms of prediction accuracy, prediction monotonicity, and prediction consistency and provide comparisons with existing state-of-the-art FR-SIQA and RR-SIQA methods. Detailed description will be given in the following subsection.

A. DATABASES AND PERFORMANCE MEASURES

We conducted experiments on two best-known 3D IQA databases, i.e., Laboratory for Image and Video Engineering (LIVE) 3D Phase I Database [28] and LIVE 3D Phase II Database [29]. LIVE 3D phase I contains five different distortion categories including JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blur (Gblur), and fast fading (FF). There is a total of 365 symmetrically distorted stereopairs generated from 12 reference stereo images. Although the LIVE 3D phase I database is an excellent tool to develop 3D IQA methods, but it only considers symmetrically distorted 3D images. However, in certain 3D image processing applications, each of view of 3D images may be corrupted with different amounts or types of distortions, which is also called asymmetrically distorted stimuli. To fill this gap, the LIVE 3D phase II database contains 8 reference images and 360 distorted images (72 each for JP2K, JPEG, WN, FF, and Gblur) with co-registered human scores in the form of difference mean opinion score (DMOS). For each distortion type, every reference stereo pair is processed to create three symmetric and six asymmetric distorted stereo pairs.

For objective evaluation, three popular criteria are used to benchmark the performance of our proposed method: the Pearson linear Correlation Coefficient (PLCC), the Spearman Rank-order Correlation Coefficient (SROCC), and the Root Mean Square Error (RMSE). The PLCC, SROCC, and RMSE metrics measure prediction linearity, prediction monotonicity, and prediction accuracy, respectively. The higher values of PLCC and SROCC, the lower value of RMSE, represent the better performance. Before computing the linear correlation between the quality scores of the algorithm and the

TABLE 1. Performance comparison of the proposed SIQA model and ten competing indices on two benchmark datasets.

Databases	Criteria	FR					RR					
		Chen[29]	Benoit[32]	Bensalma[36]	Shao[38]	Lin[41]	RRED[7]	RDCT[8]	Hewage[25]	Ma[21]	Wang[20]	Proposed
LIVE 3D Phase I	PLCC	0.8791	0.8899	0.8887	0.9262	0.8765	0.8818	0.9042	0.8300	0.9056	0.8998	**0.9301**
	SROCC	0.8603	0.8901	0.8767	0.9278	0.8721	0.8423	0.9042	0.8140	0.9052	0.8922	**0.9292**
	RMSE	7.8164	7.4786	7.5194	6.1812	8.0236	7.7332	7.0050	9.1390	6.9542	7.1557	**6.0241**
LIVE 3D Phase II	PLCC	0.8837	0.7475	0.7699	0.9116	0.6795	0.6917	0.8004	0.558	0.8179	0.5216	**0.9213**
	SROCC	0.8775	0.7642	0.7513	0.9076	0.6844	0.7258	0.7698	0.501	0.7938	0.6123	**0.9175**
	RMSE	5.2841	7.2806	7.2035	4.3562	8.2295	8.1511	6.7659	------	6.4939	9.6303	**4.3903**

subjective scores available with the two databases, a five parameters logistic function is used for nonlinear regression [62].

$$f(x) = \alpha_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(a_2 (x - \alpha_3))} \right) + \alpha_4 x + \alpha_5 \tag{25}$$

where x denotes the predicted quality of image and $f(x)$ denotes the quality score after nonlinear fitting, and $a_i, i = 1, 2, \ldots, 5$, are the regression model parameters to be fitted.

B. OVERALL PERFORMANCE COMPARISION

To investigate the effectiveness and robustness of our proposed method for all distorted types, we choose several existing state-of-art FR-SIQA and RR-SIQA metrics for comparison, i.e., Chen *et al.* [29], Benoit *et al.* [32], Bensalma and Larabi [36], Shao *et al.* [38], Lin and Wu [41], RRED [7], RDCT [8], Ma *et al.* [21], Wang *et al.* [20], and Hewage and Martini [25]. The overall performance comparisons are tabulated in Table 1. Note that, in Table 1, the best performing metrics for each database are highlighted in boldface. In our experiments, Chen *et al.* [29], Benoit *et al.* [32], Bensalma and Larabi [36], Shao *et al.* [38], and Lin and Wu [41] are five representative state-of-art FR-SIQA metrics. For the RR-SIQA metrics, we compare our proposed metric with RRED [7], RDCT [8], Ma *et al.* [21], Wang *et al.* [20], and Hewage and Martini [25]. Note that, RRED [7] and DNT [8], which are extended from the conventional 2D RR metrics. The two RR metrics will perform the quality analysis for the left and right view image independently. Then the average quality index is regarded as the perceptual quality of the stereo image. From Table 1, it can be observed that the proposed method correlates highly with human opinion ratings, indicating that the statistical properties of natural images and the properties of HVS are sufficiently considered for RR-SIQA. To be specific, it has been discovered that Chen's metric [29], Shao's metric [38], and Ma's metric [21] achieve reasonably well performance for both symmetric and asymmetric distortions of stereo images. One possible explanation is that the cyclopean map model (Chen's metric [29]), monocular and binocular visual perception model (Shao's metric [38]), and NSS features in reorganized DCT domain (Ma's metric [21]) are highly in line with human subjective perception. However, Chen's metric [29] only takes into account binocular rivalry properties of HVS. Shao's metric [38] only simulates the response of the receptive fields without consider the statistical properties of natural images. Ma's metric [21] only extracts the NSS features without consider the perceptual properties of HVS. All of which are not adequately for SIQA design. Moreover, observe that RRED [7] and RDCT [8] extended from the RR 2D IQA models generally perform worse on both symmetric and asymmetric distortions of stereo images. Since our proposed metric more comprehensive considers the statistical properties of natural images and the properties of HVS, the performance is highly consistent with human perception, which is markedly superior to all the other metrics used for comparison on the two databases. Especially, for the asymmetric distortion of stereo images, the proposed metric achieves promising performance in terms of three general criteria. Besides, as shown in Fig. 7, we also show the scatter plots of objective scores predicted by the proposed metric against subjective DMOS scores on the two databases. In Fig. 7 (c) and (f), it clearly illustrates that our proposed algorithm achieves significantly high consistent with human opinions. Based on these observations, we can conclude that the proposed metric is stably calibrated for quantifying and assessing the perceptual distortions of stereo images.

C. PERFORMANCE ON INDIVIDUAL DISTORTION TYPES

In order to more comprehensively evaluate the prediction performance of the proposed scheme, we compare the ten schemes on each type of distortion. The PLCC and SROCC results are tabulated in Table 2 and Table 3, where the best metrics have been highlighted in boldface. One can see that the performance of the proposed metric on some specific types of distortions are among the top 7 times in terms of PLCC and SROCC, followed by Shao's metric [38] 5 times, Wang's metric [20] 4 times, Chen's metric [29] 2 times, and RDCT [8] 1 times. Particularly, the proposed metric achieves very promising results for JP2K and FF distortions on the LIVE 3D Phase I Database, and most of distortions on the LIVE 3D Phase II Database. One likely reason is that the proposed metric is more sensitive to these two kinds of distortions. However, the proposed metric is not very prominent for JPEG distortion. A plausible explanation is that the luminance and gradient information cannot comprehensive reflect depth quality degradation of this distortion. To be sure, the proposed metric achieves high consistent with the subjective opinions, and can be competitive with the most effective metric for individual types across the two databases.

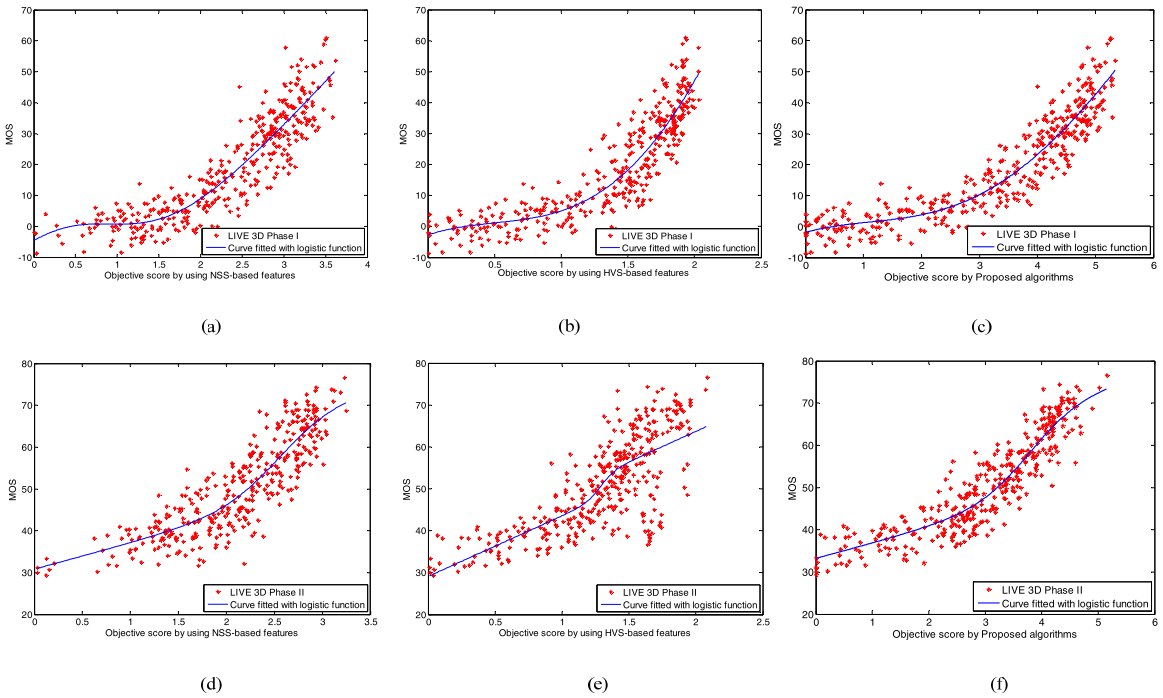

FIGURE 7. Scatter plots of objective scores versus subjective data (DMOS). (a-c) LIVE 3D Phase I. (d-f) LIVE 3D Phase II.

TABLE 2. Performance comparison of the other ten metrics on each individual distortion types in terms of PLCC.

Databases	Types	FR					RR					
		Chen[29]	Benoit[32]	Bensalma[36]	Shao[38]	Lin[41]	RRED[7]	RDCT[8]	Hewage[25]	Ma[21]	Wang[20]	Proposed
LIVE 3D Phase I	JPEG	0.2914	0.5537	0.3868	0.6856	0.4773	0.3367	**0.7228**	0.5300	0.7222	0.5574	0.7200
	JP2K	0.8615	0.8789	0.8410	0.9389	0.8650	0.8051	0.9157	0.9040	0.9182	0.9252	**0.9399**
	WN	**0.9502**	0.9357	0.9157	0.9368	0.9374	0.9425	0.9118	0.8950	0.9131	0.9196	0.9347
	Gblur	0.9086	0.9181	0.9374	0.9503	0.9173	0.9272	0.9204	0.7980	0.9247	**0.9596**	0.9356
	FF	0.6783	0.7377	0.7385	0.8251	0.7203	0.6959	0.8073	0.6690	0.8068	0.8339	**0.8427**
LIVE 3D Phase II	JPEG	0.8218	0.5328	0.8577	0.8528	0.6417	0.8077	0.7547	0.734	0.7544	**0.8971**	0.7645
	JP2K	**0.9612**	0.6467	0.6667	0.8496	0.7224	0.8323	0.7468	0.664	0.8094	0.9524	0.8795
	WN	0.7467	0.8610	0.9436	**0.9557**	0.9271	0.9431	0.8243	0.891	0.8220	0.7738	0.9325
	Gblur	0.9558	0.8814	0.9077	**0.9758**	0.8417	0.9490	0.9694	0.450	0.9721	0.9433	0.9131
	FF	0.8887	0.8472	0.9097	**0.9259**	0.8561	0.9167	0.8925	0.746	0.9016	0.9030	0.9063

D. CONTRIBUTIONS OF THE NSS-BASED AND HVS-BASED QUALITY-PREDICTIVE FEATURES IN THE PROPOSED METRIC

To obtain additional and further insight into how the prediction performance of the proposed scheme is enhanced by considering NSS-based and HVS-based together, we design two experiments for performance comparison, referred to as metric-NSS and metric-HVS. For metric-NSS, we only apply NSS-based features to yield the perceptual quality of stereo image. For metric-HVS, only the HVS-based features are adapted to measure the perceptual quality of stereo image. The results of the metric-NSS, metric-HVS, and the proposed model in terms of PLCC, SROCC, and RMSE are shown in Table 4. One can see that the prediction performance of the proposed scheme can be further improved by appropriately jointing the NSS-based and HVS-based features together. Interestingly, the metric-NSS even achieves better performance than metric-HVS on the two databases. The most likely reason is that the statistical properties of HVS take precedence over the perceptual properties of HVS. Furthermore, the scatter plots of objective scores predicted by using the NSS-based, HVS-based features and the proposed scheme against subjective DMOS scores are illustrated in Fig. 7 on the two databases. From Fig. 7 (c) and (f), it is quite obvious that the overall performance can be gradually enhanced by considering the NSS-based and HVS-based features simultaneously. Based on above observations, we can conclude that the NSS-based and HVS-based features in the proposed scheme play a complementary role for RR-SIQA.

E. IMPACT OF VARYING THE NUMBER OF SCALES AND ORIENTATIONS

The steerable pyramid wavelet decomposition in the proposed scheme may greatly affect the number of extracted

TABLE 3. Performance comparison of the other ten metrics on each individual distortion types in terms of SROCC.

Databases	Types	FR					RR					
		Chen[29]	Benoit[32]	Bensalma[36]	Shao[38]	Lin[41]	RRED[7]	RDCT[8]	Hewage[25]	Ma[21]	Wang[20]	Proposed
LIVE 3D Phase I	JPEG	0.2462	0.5561	0.3466	0.6482	0.4361	0.3166	0.6217	0.5000	0.6163	0.5025	**0.6595**
	JP2K	0.8199	0.8667	0.8192	0.9006	0.8584	0.7628	0.8853	0.8560	0.8866	0.8881	**0.9073**
	WN	0.9327	0.9388	0.9082	**0.9450**	0.9379	0.9307	0.9133	0.9400	0.9124	0.9149	0.9283
	Gblur	0.9004	0.8726	0.9179	0.9273	0.8793	0.9116	0.8692	0.6900	0.8791	**0.9290**	0.9207
	FF	0.6446	0.6106	0.6533	**0.8056**	0.5861	0.6610	0.6943	0.5450	0.6964	0.7858	0.7919
LIVE 3D Phase II	JPEG	0.8159	0.5078	0.8461	0.8421	0.6789	0.8076	0.7291	0.736	0.7288	**0.8897**	0.7912
	JP2K	0.9427	0.6325	0.8038	0.8846	0.7027	0.8562	0.7419	0.598	0.7939	**0.9496**	0.8680
	WN	0.7583	0.8569	0.9386	0.9517	0.9206	0.9421	0.7567	0.880	0.7492	0.7709	**0.9541**
	Gblur	0.8901	0.8545	0.8838	0.9159	0.8358	0.8962	0.8822	0.028	0.8868	0.8929	**0.9229**
	FF	0.8497	0.8319	0.8743	0.9013	0.8348	0.8911	0.8845	0.684	0.8886	0.8524	**0.9438**

TABLE 4. Contributions of the NSS-based and HVS-based quality-prediction features in the proposed scheme in terms of PLCC, SROCC, and RMSE.

Databases	Metrics	PLCC	SROCC	RMSE
LIVE 3D Phase I	Metric-NSS	0.9093	0.9068	6.8226
	Metric-HVS	0.9060	0.9127	7.0508
	Proposed	0.9301	0.9292	6.0241
LIVE 3D Phase II	Metric-NSS	0.8773	0.8741	5.4173
	Metric-HVS	0.7760	0.7619	7.1293
	Proposed	0.9213	0.9175	4.3903

features, i.e., the GGD model parameters (g, s) and the entropy of DWT coefficients of modified gradient magnitudes will be affected by the steerable pyramid wavelet decomposition. Based on such consider-ation, we design two groups of experiments, referred to as $M_{NSS^{S,O}}$ and $M_{HVS^{S,O}}$. Note that, $M_{NSS^{S,O}}$ denotes evaluating perceptual quality of stereo image by using NSS-based features with S scales and O orientations. Table 5 provides the results of $M_{NSS^{S,O}}$ metric with different scales and orientations. Similarly, $M_{HVS^{S,O}}$ denotes evaluating perceptual quality of stereo image by using HVS-based features with S scales and O orientations. Table 6 shows the obtained performance of $M_{HVS^{S,O}}$ metric with different scales and orientations. From Table 5, we can observe that the performance of the $M_{NSS^{S,O}}$ metric can be gradually enhanced with increasing the number of scales from 2 to 4, and orientations from 2 to 6 on both symmetric and asymmetric distortions. Furthermore, we also find that the performance of the $M_{NSS^{S,O}}$ metric can be improved with the same scales and the larger number of the orientations on the two databases. Meanwhile, in Table 6, one can see that the performance of the $M_{HVS^{S,O}}$ metric can be gradually enhanced with increasing the number of scales from 3 to 6, and orientations from 2 to 4 on asymmetric distortions. For symmetric distortions (LIVE 3D Phase I), observe that the performance of the $M_{HVS^{S,O}}$ metric has a significant drop in accuracy of the estimated quality with the same scales and the larger orientations. One possible explanation for this situation is that the changes of orientation play an important role in binocular rivalry of HVS, while there is little or no binocular rivalry in symmetrically distorted stereo images. In general, our proposed multi-scale feature representation can efficiently capture changes of the statistical properties and the perceptual properties of stereo images using very few parameters.

TABLE 5. Effect of changing the number of scales and orientations in $M_{NSS^{S,O}}$ metric in terms of PLCC, and SROCC for the LIVE 3D Phase I and Phase II databases.

	LIVE 3D Phase I		LIVE 3D Phase II		
	PLCC	SROCC	PLCC	SROCC	# Features
$M_{NSS^{2,2}}$	0.8738	0.8724	0.8169	0.8203	4
$M_{NSS^{2,4}}$	0.8800	0.8779	0.8317	0.8319	8
$M_{NSS^{2,6}}$	0.8804	0.8788	0.8333	0.8343	12
$M_{NSS^{3,2}}$	0.8926	0.8904	0.8497	0.8485	6
$M_{NSS^{3,4}}$	0.9024	0.8997	0.8638	0.8608	12
$M_{NSS^{3,6}}$	0.9025	0.9008	0.8657	0.8627	18
$M_{NSS^{4,2}}$	0.8965	0.8944	0.8544	0.8613	8
$M_{NSS^{4,4}}$	0.9093	0.9062	0.8761	0.8725	16
$M_{NSS^{4,6}}$	0.9093	0.9068	0.8773	0.8741	24

TABLE 6. Effect of changing the number of scales and orientations in $M_{HVS^{S,O}}$ metric in terms of PLCC, and SROCC for the LIVE 3D Phase I and Phase II databases.

	LIVE 3D Phase I		LIVE 3D Phase II		
	PLCC	SROCC	PLCC	SROCC	# Features
$M_{HVS^{3,2}}$	0.8783	0.9044	0.5656	0.7021	3
$M_{HVS^{3,4}}$	0.8656	0.8738	0.7519	0.7176	3
$M_{HVS^{4,2}}$	0.9062	0.9150	0.7540	0.7278	4
$M_{HVS^{4,4}}$	0.9019	0.9055	0.7295	0.7382	4
$M_{HVS^{5,2}}$	0.9101	0.9155	0.7571	0.7356	5
$M_{HVS^{5,4}}$	0.9071	0.9135	0.7693	0.7500	5
$M_{HVS^{6,2}}$	0.9115	0.9172	0.7512	0.7381	6
$M_{HVS^{6,4}}$	0.9060	0.9127	0.7760	0.7619	6

V. CONCLUSION

In this paper, we have presented a new RR-SIQA metric by using natural scene statistics and structural degradation.

The novelty of our research resides in the application of the statistical properties of stereo images and perceptual properties of HVS together to predict the perceived quality of stereo images. In general, most of the existing RR-SIQA metrics need to either train or tune their methods to predict the quality of the stereo images accurately. Unlike such frameworks, our proposed method performs well without training and tuning. Furthermore, finding a balance between the number of RR features and the predicted stereo image quality plays a key role in RR-SIQA design. Interestingly, our proposed method not only needs a very small number of RR features (30 RR features per stereo pair), but also yields a high accuracy estimation on both symmetric and asymmetric stereo images. In conclusion, the proposed scheme achieves significantly better consistency with subjective opinions. In the future work, we expect to extend our proposed method to RR stereo video quality assessment (SVQA).

REFERENCES

[1] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, no. 8, pp. 1–53, Nov. 2012.

[2] *Methodology for the Subjective Assessment of the Quality of Television Picture*, IEEE Standard ITU-R BT. 500-11, 2002.

[3] 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.

[4] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[5] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[6] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan. 2016.

[7] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 26–517, Feb. 2012.

[8] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference image quality assessment in reorganized dct domain," *Signal Process. Image Commun.*, vol. 28, no. 8, pp. 884–902, 2013.

[9] C. C. Su, A. K. Moorthy, and A. C. Bovik, "Visual quality assessment of stereoscopic image and video: Challenges, advances, and future trends," in *Visual Signal Quality Assessment*. Cham, Switzerland: Springer, 2015.

[10] X. Cao, Z. Li, and Q. Dai, "Semi-automatic 2D-to-3D conversion using disparity propagation," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 491–499, Jun. 2011.

[11] J. Kim, T. Kim, S. Lee, and A. C. Bovik, "Quality assessment of perceptual crosstalk on two-view auto-stereoscopic displays," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4885–4899, Oct. 2017.

[12] 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.

[13] J. Park, H. Oh, S. Lee, and A. C. Bovik, "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101–1114, Mar. 2015.

[14] K. Gu, G. Zhai, X. Yang, and W. Zhang, "No-reference stereoscopic IQA approach: From nonlinear effect to parallax compensation," *J. Elect. Comput. Eng.*, vol. 2012, no. 5, pp. 436031-11–436031-12, 2012.

[15] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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, 2015.

[21] 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.

[22] F. Qi, D. B. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Trans. Mutimedia*, vol. 17, no. 12, pp. 2338–2343, Dec. 2015.

[23] Q. Xu, G. Zhai, M. Liu, and K. Gu, "Using structural degradation and parallax for reduced-reference quality assessment of 3D images," in *Proc. IEEE Int. Symp. Broadband Mutltimedia Syst. Broadcast.*, Jun. 2014, pp. 2338–2343.

[24] K. Zheng, M. Yu, X. Jin, G. Jiang, Z. J. Peng, and F. Shao, "New reduced-reference objective stereo image quality assessment model based on human visual system," in *Proc. IEEE 3DTV-Conf.*, Jul. 2014, pp. 1–4.

[25] C. T. E. R. Hewage and M. G. Martini, "Reduced-reference quality assessment for 3D video compression and transmission," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1185–1193, Aug. 2011.

[26] T. J. Liu, W. Lin, and C. C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1806, May 2013.

[27] 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.

[28] 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.

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

[30] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. Eur. Signal Process. Conf.*, 2007, pp. 2110–2113.

[31] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, pp. 680305-1–680305-12, Feb. 2008.

[32] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 1–13, 2008.

[33] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual images quality for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, vol. 9. 2010, pp. 1–6.

[34] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-D video," *Electron. Lett.*, vol. 44, no. 16, pp. 963–965, Jul. 2008.

[35] 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.

[36] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 2, no. 3, pp. 281–316, 2013.

[37] 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. 2971–2983, Oct. 2015.

[38] F. Shao, W. Lin, G. Jiang, and Q. Dai, "Models of monocular and binocular visual perception in quality assessment of stereoscopic," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 1940–1953, Jun. 2016.

[39] F. Shao, W. Chen, G. Jiang, and Y. S. Ho, "Modeling the perceptual quality of stereoscopic images in the primary visual cortex," *IEEE Access*, vol. 5, pp. 15706–15716, Aug. 2017.

[40] 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.
[41] 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.
[42] W. Zhou, W. Qiu, and M. W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3D images," *IEEE Trans. Broadcast.*, vol. 63, no. 2, pp. 404–415, Jun. 2017.
[43] R. W. Buccigrossi and E. P. Simoncelli, "Imag compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.
[44] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising suing scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
[45] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Oct. 1998, pp. 633–636.
[46] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. Int. Conf. Image Process.*, vol. 3. 1995, pp. 444–447.
[47] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.
[48] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2379–2394, 1987.
[49] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
[50] W. J. Levelt, *On Binocular Rivalry*, vol. 2. The Hague, The Netherlands: Mouton, 1968.
[51] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.
[52] S. A. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5293–5303, Nov. 2016.
[53] T. Mitsa and K. L. Varkur, "Evaluation of contrast sensitivity function for the formulation of quality measures incorporated in halftoning algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1993, pp. 301–304.
[54] J. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 525–536, Jul. 1974.
[55] S. Daly, "Subroutine for the generation of a two dimensional human visual contrast sensitivity function," Eastman Kodak, Rochester, NY, USA, Tech. Rep. Y 233203, 1987.
[56] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," *Vis. Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.
[57] I. Sobel and G. Feldman, "A 3×3 isotropic gradient operator for image processing," presented at the Stanford Artif. Intell. Project (SAIT), 1968.
[58] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.
[59] L. W. Renninger, P. Verghese, and J. Coughlan, "Where to look next? Eye movements reduce local uncertainty," *J. Vis.*, vol. 7, no. 3, p. 6, 2007.
[60] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *Int. J. Math. Models Methods Appl. Sci.*, vol. 1, no. 4, pp. 300–307, 2007.
[61] M. Hatzigiorgaki and A. N. Skodras, "Compressed domain image retrieval: A comparative study of similarity metrics," in *Visual Communications and Image Processing*. Bellingham, WA, USA: SPIE, 2003, pp. 439–448.
[62] 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.

JIAN MA received the B.S. degree in electronic information engineering from Anhui Normal University in 2010 and the M.S. degree in computer software and theory from Bohai University in 2013. He is currently pursuing the Ph.D. degree in communication and information engineering with Shanghai University. His research interests include 3-D image/video quality assessment, deep learning, and computer vision.

PING AN received the B.S. and M.S. degrees from the Hefei University of Technology in 1900 and 1993, and the Ph.D. degree from Shanghai University in 2002. She joined as a Professor with the Video Processing Group, School of Communication and Information Engineering, Shanghai University, China, in 1993. From 2011 to 2012, she joined the Communication Systems Group, Technische University at Berlin, Germany, as a Visiting Professor. She has finished over ten projects supported by the National Natural Science Foundation of China, the National Science and Technology Ministry, and the Science and Technology Commission of Shanghai Municipality. Her research interest is image and video processing, especially focuses on 3-D video processing recent years. She received the second prize of Shanghai Municipal Science and Technology Progress Award in 2011 and the second prize in the Natural Sciences of the Ministry of Education in 2016.

LIQUAN SHEN received the B.S. degree in automation control from Henan Polytechnic University, Jiaozuo, China, in 2001, and the M.S. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. He was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, as a Visiting Professor from 2013 to 2014. He has been a Professor with the Faculty of the School of Communication and Information Engineering, Shanghai University, since 2008. He has authored and co-authored over 80 refereed technical papers in international journals and conferences in the field of video coding and image processing. He holds ten patents in the areas of image/video coding and communications. His research interests include High Efficiency Video Coding, perceptual coding, video codec optimization, 3-DTV, and 3-D image/video quality assessment.

KAI LI received the B.S. degree (Hons.) in communication engineering from Shanghai University, Shanghai, China, in 2010, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2015. He was a Student Intern at the Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA, from 2012 to 2013. Since 2015, he has been an Assistant Professor with the School of Communication and Information Engineering, Shanghai University. His research interests include image and video processing, computer vision, graphics, and 3-D image/video quality assessment.