No-Reference Quality Assessment for 360-degree Images by Analysis of Multifrequency Information and Local-global Naturalness

Wei Zhou, *Student Member, IEEE*, Jiahua Xu, Qiuping Jiang, *Member, IEEE*, and Zhibo Chen, *Senior Member, IEEE*

***Abstract*—360-degree/omnidirectional images (OIs) have received remarkable attention due to the increasing applications of virtual reality (VR). Compared to conventional 2D images, OIs can provide more immersive experiences to consumers, benefiting from the higher resolution and plentiful field of views (FoVs). Moreover, observing OIs is usually in a head-mounted display (HMD) without references. Therefore, an efficient blind quality assessment method, which is specifically designed for 360-degree images, is urgently desired. In this paper, motivated by the characteristics of the human visual system (HVS) and the viewing process of VR visual content, we propose a novel and effective no-reference omnidirectional image quality assessment (NR OIQA) algorithm by MultiFrequency Information and Local-Global Naturalness (MFILGN). Specifically, inspired by the frequency-dependent property of the visual cortex, we first decompose the projected equirectangular projection (ERP) maps into wavelet subbands by using discrete Haar wavelet transform (DHWT). Then, the entropy intensities of low-frequency and high-frequency subbands are exploited to measure the multifrequency information of OIs. In addition to considering the global naturalness of ERP maps, owing to the browsed FoVs, we extract the natural scene statistics (NSS) features from each viewport image as the measure of local naturalness. With the proposed multifrequency information measurement and local-global naturalness measurement, we utilize support vector regression (SVR) as the final image quality regressor to train the quality evaluation model from visual quality-related features to human ratings. To our knowledge, the proposed model is the first no-reference quality assessment method for 360-degree images that combines multifrequency information and image naturalness. Experimental results on two publicly available OIQA databases demonstrate that our proposed MFILGN outperforms state-of-the-art full-reference (FR) and NR approaches.**

***Index Terms*—Omnidirectional images, no-reference image quality assessment, multifrequency information, local-global naturalness, human visual system.**

I. Introduction

IMMERSIVE multimedia technologies, especially virtual reality (VR), can provide viewers with more realistic and interactive user experiences [1]. As the most common form of VR content, 360-degree/omnidirectional images and videos record visual information that covers the entire $180 \times 360^{\circ}$ viewing sphere, thus attracting considerable attention from both academia and industry in recent years [2]. With a commercial head-mounted display (HMD), users can freely view any direction with the specific content by head movement, which is different from conventional 2D images and videos. Moreover, due to the omnidirectional viewing range, the resolution of 360-degree images/videos is usually ultrahigh, e.g., 4 K, 8 K, or even higher. This creates considerable difficulties in the 360-degree image/video processing chain, such as acquisition, compression, transmission, reconstruction and display [3]. Additionally, the perceptual quality of omnidirectional images can degrade in 360-degree image/video processing systems. Therefore, the study of 360-degree/omnidirectional image quality assessment (OIQA) is more challenging and significant for guiding the development of VR applications.

Recently, there has been ever-increasing interest in the research field of image quality assessment (IQA). Two types of IQA methods are involved: subjective IQA [4]–[8] and objective IQA [9]–[13]. In subjective IQA tests, subjects are asked to give the human ratings for each viewed image. After data processing and outlier elimination, the mean opinion score (MOS) can be obtained by computing the average quality scores of all subjects for each image, which can be regarded as the quantitative ground truth of the perceptual quality [14]. Since humans are the ultimate viewers, subjective IQA is the most reliable quality assessment approach. By conducting such subjective experiments, several subjective OIQA databases have been established. For example, a testbed for the subjective measurement of 360-degree/omnidirectional content was proposed [15], where 6 reference and 54 distorted omnidirectional images (OIs) were included. JPEG compression with various quality parameters and two projection models were considered in this database. Moreover, in [16], 4 high-fidelity uncompressed OIs were used to generate 100 impaired images with different geometric projections and three codecs, i.e., JPEG, JPEG2000, and HEVC intra. Four target bitrates were selected to compress the original OIs. The compressed VR image quality database (CVIQD) was built in [17], which includes 5 reference and 165 compressed OIs by JPEG, AVC, and HEVC codecs. Furthermore, this database was expanded to CVIQD2018 [18] with more visual content, leading to 16 pristine and 528 compressed images. In [19], different resolutions and JPEG compression were considered, where the

This work was supported in part by NSFC under Grants U1908209, 61632001, and 61901236 and the National Key Research and Development Program of China 2018AAA0101400.

W. Zhou, J. Xu and Z. Chen are with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China (e-mail: weichou@mail.ustc.edu.cn; xujiahua@mail.ustc.edu.cn; chenzhibo@ustc.edu.cn).

Q. Jiang is with the School of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jiangqiuping@nbu.edu.cn).

(a) (b)

Fig. 1: Illustration of the projected ERP map and various viewports of a 360-degree image.

database had 12 original and 144 distorted OIs. In addition, an omnidirectional image quality assessment (OIQA) database [20] was developed, consisting of 16 reference and 320 distorted images. In addition to JPEG and JPEG2000 compression artifacts, different levels of white Gaussian noise and Gaussian blur were considered in this database. Since observers usually exploit the HMD to view OIs, absolute category rating with hidden reference (ACR-HR) methods, which are also referred to as single-stimulus (SS) methods, were adopted to build all these subjective quality assessment databases for OIs.

However, subjective tests are generally time consuming and labor intensive. Thus, objective visual quality assessment algorithms used to automatically measure the perceptual quality of OIs are required. As shown in Fig. 1, (a) represents the projected equirectangular projection (ERP) map of a viewed VR scene. The projected ERP map usually has a higher resolution, such as 4 K or 8 K. Hence, it is suitable for multiresolution decomposition, which is also an image decomposition in the frequency channels of constant bandwidth on a logarithmic scale. Moreover, the concept of multiresolution can interpret multifrequency channel decomposition [21]. In this figure, except for the projected ERP map, (b) shows 6 viewports from a variety of viewing directions, each of which is a part of the 360-degree image falling into the field of view (FoV) in the HMD. The naturalness characteristics reflected by statistical regularizations from the global ERP map and local viewports are different. In addition, a no-reference quality assessment model for 3D/stereoscopic omnidirectional images was proposed in [22]. The monocular multiscale features were extracted from left and right views separately, while the binocular perception features were exploited from tensor decomposition and the absolute difference map as well as product images of left and right views. The naturalness features involved in the product image were also validated. Nevertheless, it is different than the 2D omnidirectional images considered in this paper, where we only have one single-view image with multiple viewports.

Based on these observations, in this paper, we propose a novel no-reference (NR) OIQA method by MultiFrequency Information and Local-Global Naturalness (MFILGN). First, according to a series of neuroscience studies on the human visual system (HVS), neuronal responses in the visual cortex are frequency-dependent [23]. In other words, each neuron corresponds to specific spatial and temporal frequency signals. Therefore, the input visual signal can be decomposed into multiple frequency domains, which is more in line with human visual perception. Among various multifrequency channel decomposition methods, wavelet decomposition shows superiority in processing visual signals [24]. Wavelet decomposition has also been demonstrated to have good performance in IQA [25], [26]. Motivated by this mechanism, we decompose the projected ERP maps into wavelet subbands through discrete Haar wavelet transform (DHWT). The decomposed low-frequency and high-frequency subbands represent luminance information and textural details, respectively. We then compute the entropy intensities of low-frequency and high-frequency subbands, which are used to measure the multifrequency information of OIs. Second, due to different viewports during browsing, we propose the local-global naturalness measurement, where the natural scene statistics (NSS) features are extracted from both locally viewed FoVs and global ERP maps. To the best of our knowledge, naturalness has been proven in many IQA studies [27], [28], but it has not been used in OIQA. Finally, the quality predictions of OIs are obtained by the well-known support vector regression (SVR). As demonstrated by extensive experiments, our proposed MFILGN performs better than state-of-the-art FR and NR algorithms on two publicly available OIQA databases.

The main contributions of this work are summarized as follows:

- We propose the first blind OIQA algorithm based on multifrequency information and local-global naturalness measurements.
- According to the frequency-dependent characteristic of the human visual cortex, we derive the decomposed low- and high-frequency subbands and utilize the entropy intensities of these subband images to reflect the multifrequency information in omnidirectional images.
- Considering the viewing process of omnidirectional images, in addition to the global naturalness from the projected ERP maps, we extract the local naturalness features from various viewports together with global NSS to composite the local-global naturalness measurement.

The remaining sections of this paper are organized as follows. In Section II, we present the related works of objective quality assessment for both traditional 2D images and OIs. Section III introduces the proposed MFILGN model for NR OIQA in detail. In Section IV, the experimental results and analysis are presented. Section V concludes the paper with possible future research directions.

II. Related Works

Whether for traditional objective IQA or objective OIQA, when the originally pristine/reference image is available, full-reference (FR) objective visual quality assessment models are developed. For conventional FR IQA, signal fidelity metrics such as the mean square error (MSE) and peak signal-to-noise ratio (PSNR) measure the visual quality by computing the pixel differences between original and distorted images. Due to the simple calculation and optimization processes, they are widely used in image processing. Nevertheless, their performance is relatively unsatisfactory and cannot precisely predict the human perceived visual quality. Therefore, the characteristics of the HVS are employed to construct perception-based

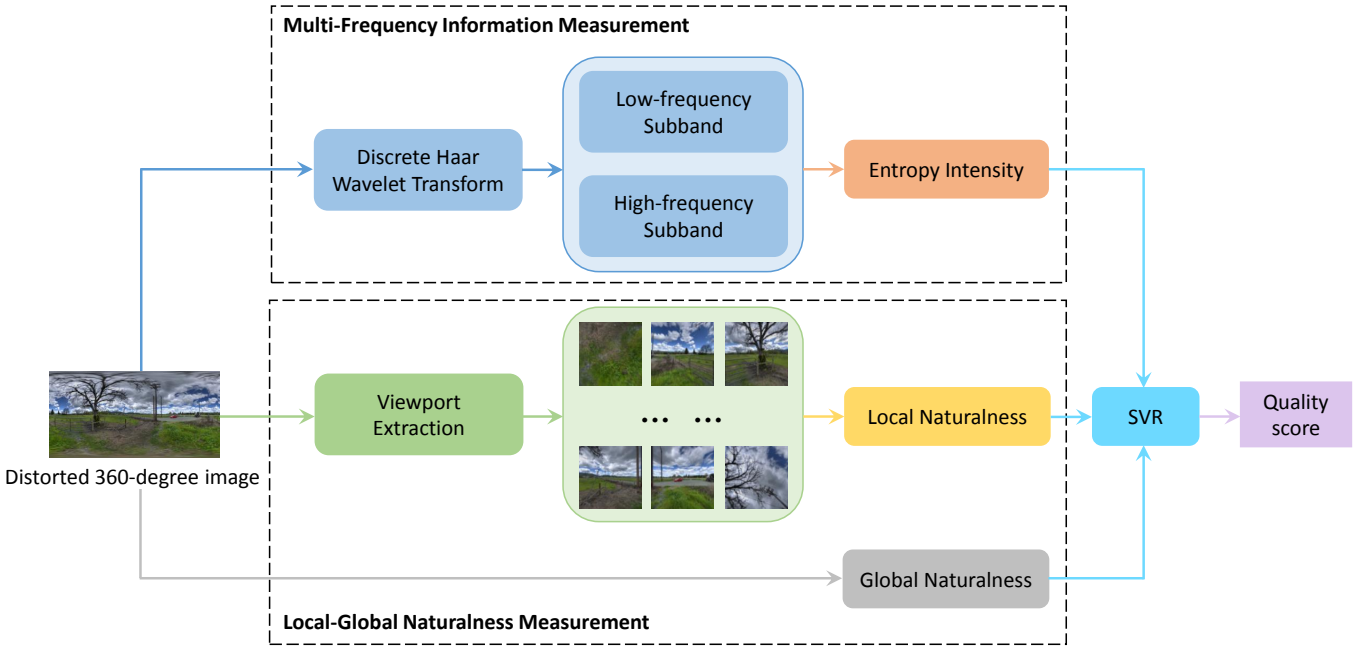

Fig. 2: The overall framework of our proposed MFILGN. It consists of multifrequency information measurement and local-global naturalness measurement.

IQA models, for instance, the structural similarity (SSIM) index [29] and its different variants, including the multiscale SSIM (MS-SSIM) index [30] and the feature similarity (FSIM) index [31].

In practical scenarios, perfect-quality original images are usually difficult to obtain, and thus, no-reference objective visual quality assessment approaches are urgently needed. For traditional NR IQA, many methods extract handcrafted distortion-discriminative features for predicting the perceptual image quality, such as the blind/referenceless image spatial quality evaluator (BRISQUE) [32] and the blind multiple pseudoreference images-based (BMPRI) measure [33]. In recent decades, due to the powerful feature representation learning capacity, deep learning has shown unprecedented success in many image processing and computer vision tasks [34], [35]. This also presents an opportunity for evaluating the image visual quality. Typical methods include deep image quality assessment (DeepQA) [36] and the deep bilinear convolutional neural network (DB-CNN) [37] for FR and NR IQA, respectively. Furthermore, several perception-based preprocessing strategies have been presented in existing works. For example, before training CNN models, saliency maps can be used to assign importance to distorted image patches [38]. The distorted image stream and gradient image stream are both considered in CNNs to predict perceptual image quality scores [39].

Although there exist many classical objective quality assessment algorithms for FR and NR IQA, they are designed for regular flat 2D images and are unsuitable for assessing the perceptual quality of OIs. However, limited research works related to objective OIQA methods have been proposed in the literature. In general, existing objective OIQA models can be classified into two categories. The first is to extend conventional FR IQA approaches to FR OIQA [40]–[45]. For example, several PSNR-based FR OIQA models have been presented. Yu *et al.* [40] proposed the spherical PSNR (S-PSNR), which chose specific points on the spherical surface rather than the projected panoramic image. Sun *et al.* [41] developed the weighted-to-spherically uniform PSNR (WS-PSNR) by combining the error map with the weighted map, which was determined by stretched regions. Zakharchenko *et al.* [42] presented the Craster parabolic projection PSNR (CPP-PSNR), which computed the PSNR on the Craster parabolic projection domain. Xu *et al.* [43] proposed the noncontent-based PSNR (NCP-PSNR) and content-based PSNR (CP-PSNR) by weighting pixels with position information and predicting viewing direction, respectively. Likewise, the performance of PSNR-based FR OIQA methods is insufficient because they do not consider the HVS characteristics. Afterward, objective FR OIQA methods based on SSIM were proposed in succession. Similar to WS-PSNR, Zhou *et al.* [44] designed the weighted-to-spherically uniform SSIM (WS-SSIM), where the position weighted map was used to multiply the SSIM. To reduce the influence of geometric distortions on the projection, Chen *et al.* [45] proposed the spherical SSIM (S-SSIM), which calculated the similarity of each image pixel on the sphere. The second category is the deep learning-based NR IQA methods [46]–[50]. Considering the spherical representation of 360-degree content, Kim *et al.* [46] proposed a deep learning framework for VR image quality assessment (DeepVR-IQA) based on adversarial learning, in which the quality scores of sampled patches were predicted and then weighted according

to the patch positions on the sphere. Moreover, in [47], head movement (HM) and eye movement (EM) were exploited to weight the quality scores in deep learning models. However, the image patches sampled from the projected plane reveal nonnegligible geometric deformation, which could not reflect the actual viewing contents. Thus, viewports were utilized to predict the perceptual quality of OIs. Specifically, Li *et al.* [48] proposed a viewport-based CNN (V-CNN), which predicted the quality scores of viewports instead of image patches sampled from projected planes. Sun *et al.* [49] designed a multichannel CNN for blind 360-degree image quality assessment (MC360IQA), including 6 parallel hyper-ResNet34 networks to process viewport images and an image quality regression module to aggregate learned features for obtaining the final image quality. Xu *et al.* [50] proposed the viewport oriented graph convolution network (VGCN) to address the perceptual quality assessment for OIs, which was guided by different viewports.

From the above reviewed objective quality assessment models for OIs, it can be concluded that existing objective OIQA methods have achieved success to a certain extent. However, we also notice that they all ignore the important multifrequency information as well as the statistical regularizations of OIs from both global projection maps and local viewports. To fill this gap, in this work, we present a no-reference OIQA method by considering multifrequency information and local-global naturalness simultaneously.

III. The Proposed Quality Assessment Method

In this section, we introduce the proposed NR OIQA method that can blindly predict the perceptual quality of OIs in technical details. The overall framework of our proposed MFILGN is shown in Fig. 2, which consists of two separate measurements: the multifrequency information measurement and local-global naturalness measurement. For the multifrequency information measurement, inspired by the HVS characteristics, the Haar wavelet transform is first applied to decompose the projected ERP maps into multiple subbands. Then, the entropy intensities are calculated to measure the multifrequency information. For the local-global naturalness measurement, considering the changeable FoVs during the viewing process, both local naturalness from different viewports and global naturalness from ERP maps are extracted. The final quality index is obtained by the regression of these distortion-related features.

A. Image Decomposition

According to HVS studies, different neurons respond to different visual signal frequencies [23]. The input visual stimulus should be decomposed into various subband images for subsequent processing. Moreover, wavelet transform is one of the multifrequency channel decomposition methods. Here, we choose the wavelet transform to conduct the distorted image decomposition. Generally, there exist many wavelets, including the Haar wavelet, Morse wavelet, Gabor wavelet, and Bump wavelet. In all of these wavelets, the Haar wavelet is symmetrical and a special case of the Daubechies wavelet, which shows great success in perceptual quality assessment

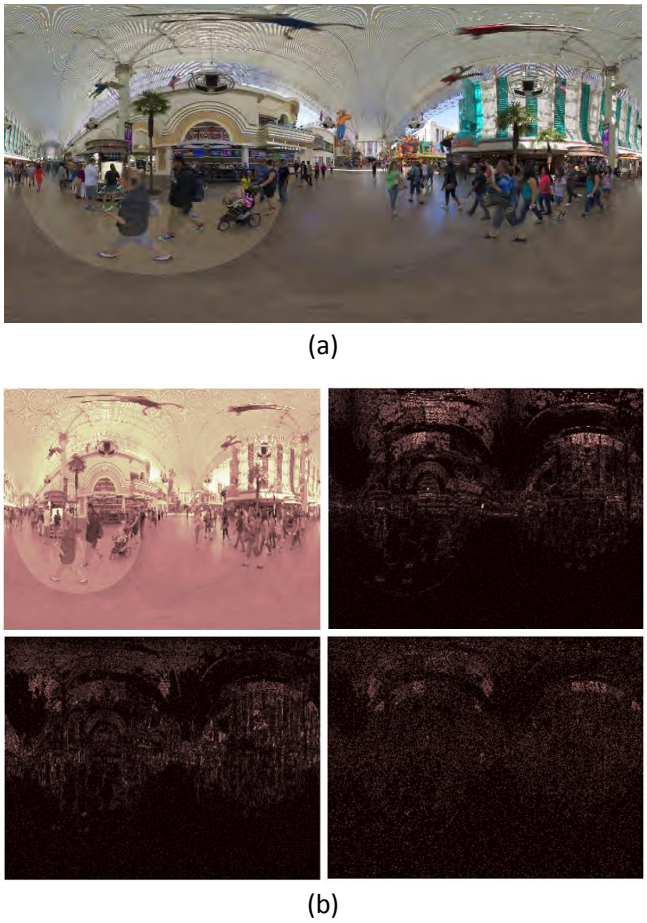

Fig. 3: Multifrequency channel decomposition by DHWT. (a) An example of a distorted 360-degree image; (b) the decomposed four subband images of (a).

[51]–[53]. Therefore, we adopt DHWT to decompose the distorted OIs into multifrequency subbands. Specifically, the Haar wavelet can be formulated as:

$$\begin{aligned}\frac{1}{\sqrt{2}}\psi(t) &= \frac{1}{\sqrt{2}}(\phi(t-1)-\phi(t)) \\ &= \sum_{u=-\infty}^{+\infty}(-1)^{1-u}h[1-u]\phi(t-u),\end{aligned} \tag{1}$$

where the mother wavelet ψ is defined by:

$$\psi(t)=\begin{cases}1, \ for\ t\in[0,\ \frac{1}{2}),\\ -1, \ for\ t\in[\frac{1}{2},\ 1),\\ 0, \ otherwise.\end{cases} \tag{2}$$

Moreover, the father wavelet or scaling function ϕ and its corresponding filter h are computed as:

$$\phi(t)=\begin{cases}1, \ for\ t\in[0,\ 1),\\ 0, \ otherwise.\end{cases} \tag{3}$$

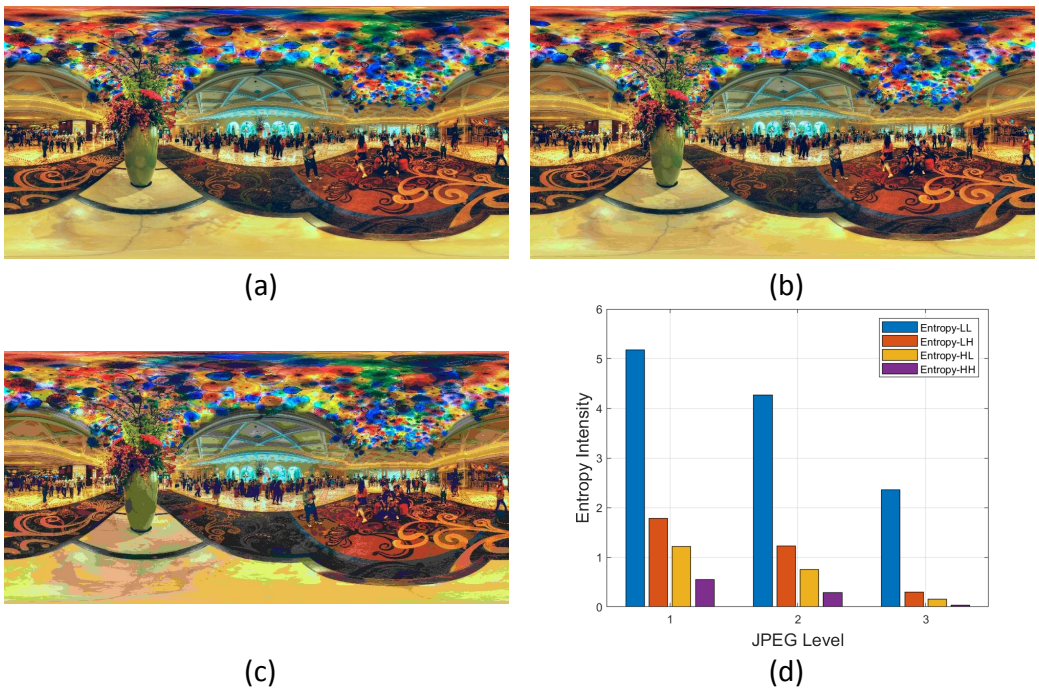

Fig. 4: Entropy intensity changes for different distorted 360 images. (a) Low JPEG distortion denoted by JPEG-Level1; (b) middle JPEG distortion denoted by JPEG-Level2; (c) high JPEG distortion denoted by JPEG-Level3; (d) the corresponding entropy intensity changes of low- and high-frequency subbands for (a), (b) and (c).

$$h[u] = \begin{cases} \frac{1}{\sqrt{2}}, \ for \ u = 0, 1, \\ 0, \ otherwise. \end{cases} \tag{4}$$

With the Haar wavelet, we can obtain the Haar wavelet transform by cross multiplying different shifts and stretches. Here, let H denote the DHWT matrix. Assume a distorted 360-degree image D with resolution $I \times J$. Then, the input 360-degree image can be decomposed into four $\frac{I}{2} \times \frac{J}{2}$ subband images by DHWT as follows:

$$HDH^T = \begin{bmatrix} D_{LL} & D_{HL} \\ D_{LH} & D_{HH} \end{bmatrix}, \tag{5}$$

where D_{LL}, D_{HL}, D_{LH}, and D_{HH} represent the four decomposed subbands with low or high frequency in the horizontal or vertical direction. H^T is the transpose matrix of H.

B. Multifrequency Information

Fig. 3 illustrates the multifrequency channel decomposition by DHWT. We show a distorted 360-degree image in (a), and the four images in (b) are the decomposed results of (a). In this figure, we can see that the decomposed subband images are different from each other, which can reflect various frequency characteristics of the distorted 360-degree image. Since the entropy intensities can reflect the average amount of image information, images with different distortions create various entropy intensities. In particular, after the multifrequency channel decompositions of distorted 360-degree images, the entropy intensities of decomposed low-frequency and high-frequency subbands differ from each other. We then compute the entropy intensities of the four decomposed subband images as:

$$E_s = -\sum_{k=0}^{K} p_k^s \log_2 p_k^s, \tag{6}$$

where $s \in \{LL, \ HL, \ LH, \ HH\}$ denotes the collection of frequency decomposition components. p_k^s and K indicate the probability of the pixel equaling k and the maximum pixel value in the corresponding decomposed subband image. The probability can be calculated by:

$$p_k^s = \frac{N_k^s}{N}, \tag{7}$$

where N_k^s is the number of pixel values equal to k in the corresponding decomposed subband image. N represents the total number of pixels.

After computing the entropy intensities of the four decomposed subband images, we use their joint component features to measure multifrequency information as:

$$F_{MFI} = [E_{LL}, \ E_{HL}, \ E_{LH}, \ E_{HH}], \tag{8}$$

where F_{MFI} denotes the multifrequency information measurement, which reflects the discriminative entropy information from both low- and high-frequency subbands.

In Fig. 4, we show three 360-degree images with different JPEG distortion levels and their corresponding entropy intensities of low-frequency and high-frequency subbands. The image samples are from OIQA database [20]. We find that the entropy intensities for the four frequency subbands change

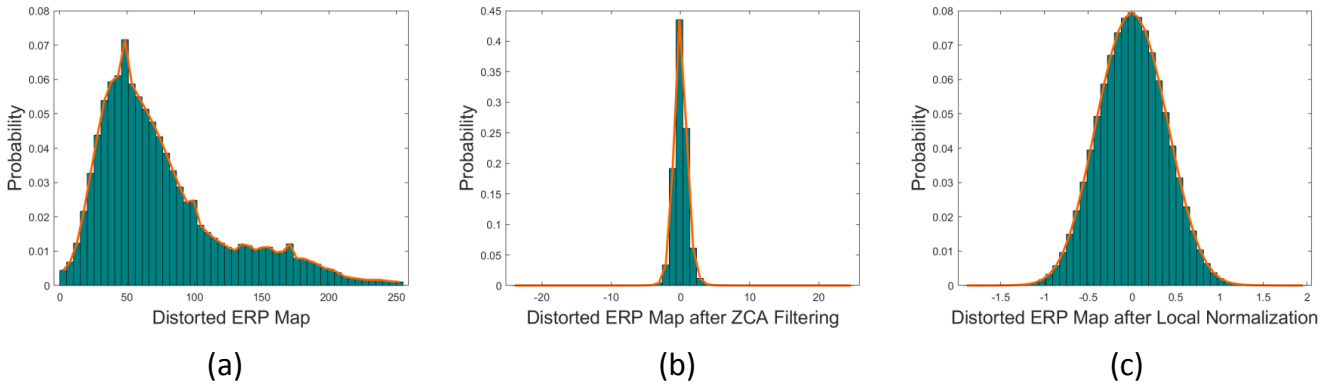

Fig. 5: Effects of the ZCA whitening filter and MSCN operation on the statistical distribution of the distorted ERP map.

significantly with regard to various distortions. That is, more JPEG distortions lead to a decrease in the entropy intensities of low- and high-frequency subbands, which demonstrates the effectiveness of the proposed multifrequency information measurement.

C. Global Naturalness

In addition to the multifrequency information in the projected ERP maps, the image naturalness, which is reflected by statistical regularizations, is crucial to the perceptual quality of OIs. Although NSS features have been applied to traditional multimedia formats, such as 2D [32] and 3D [9], to the best of our knowledge, they have not yet been used to evaluate the perceptual quality of VR images. Therefore, here we explore the naturalness in OIQA. Intuitively, we can extract the NSS features from global ERP maps. Given the distorted 360-degree image D with resolution $I \times J$, to reduce the spatial redundancy of adjacent image pixels, we first adopt the zero-phase component analysis (ZCA) whitening filter as follows:

$$D^z = Z(D), \tag{9}$$

where Z indicates the ZCA whitening filter. D^z is the distorted ERP map after ZCA filtering.

Afterward, the local mean subtracted and contrast normalized (MSCN) coefficients are computed to measure the image naturalness, which can model contrast gain masking in the early human visual cortex [54]. For each distorted ERP map after ZCA filtering, the MSCN coefficients are calculated by:

$$\hat{D}^z(i,j) = \frac{D^z(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \tag{10}$$

where $\hat{D}^z(i,j)$ and $D^z(i,j)$ are the distorted ERP maps after local normalization (i.e., the MSCN coefficients) and after ZCA filtering at spatial position (i,j), respectively. $\mu(i,j)$ and $\sigma(i,j)$ represent the local mean and standard deviation of the distorted ERP map, which are computed as:

$$\mu(i,j) = \sum_{s=-S}^{S}\sum_{t=-T}^{T} w_{s,t} D^z(i,j), \tag{11}$$

$$\sigma(i,j) = \sqrt{\sum_{s=-S}^{S}\sum_{t=-T}^{T} w_{s,t}(D^z(i,j) - \mu(i,j))^2}, \tag{12}$$

where $w = \{w_{s,t} | s = -S..., S, t = -T...T\}$ denotes the 2D circularly symmetric Gaussian weighted function.

In Fig. 5, we show the statistical distributions of the input distorted ERP map and the distorted ERP map after ZCA filtering as well as the local normalization process. In figures (a-c), we can observe that the ZCA filtering and the MSCN operation both make the probability distribution of the distorted ERP map more Gaussian-like. Moreover, the statistical distribution after local normalization is the closest to the Gaussian distribution. Thus, we exploit the MSCN coefficients for the subsequent feature processing. Additionally, Fig. 6 presents the statistical distributions of MSCN coefficients for different distortion types and levels. As seen in this figure, the probability distributions of the distorted ERP map after local normalization are influenced by different distortion types as well as the distortion levels, which demonstrates that the statistical distributions of MSCN coefficients are discriminative for the perceptual quality assessment of 360-degree images.

With the probability distributions of the distorted ERP map after local normalization, we utilize the zero-mean generalized Gaussian distribution (GGD) and asymmetric generalized Gaussian distribution (AGGD) models to quantify the MSCN coefficient distribution. The zero-mean AGGD used to fit the distribution is:

$$f(x;\tau,\sigma_l^2,\sigma_r^2) = \begin{cases} \dfrac{\tau}{(\upsilon_l+\upsilon_r)\Gamma(\frac{1}{\tau})} e^{-(\frac{-x}{\upsilon_l})^\tau}, & x < 0, \\ \dfrac{\tau}{(\upsilon_l+\upsilon_r)\Gamma(\frac{1}{\tau})} e^{-(\frac{-x}{\upsilon_r})^\tau}, & x \geq 0, \end{cases} \tag{13}$$

where

$$\upsilon_l = \sigma_l \sqrt{\frac{\Gamma(\frac{1}{\tau})}{\Gamma(\frac{3}{\tau})}}, \tag{14}$$

$$\upsilon_r = \sigma_r \sqrt{\frac{\Gamma(\frac{1}{\tau})}{\Gamma(\frac{3}{\tau})}}, \tag{15}$$

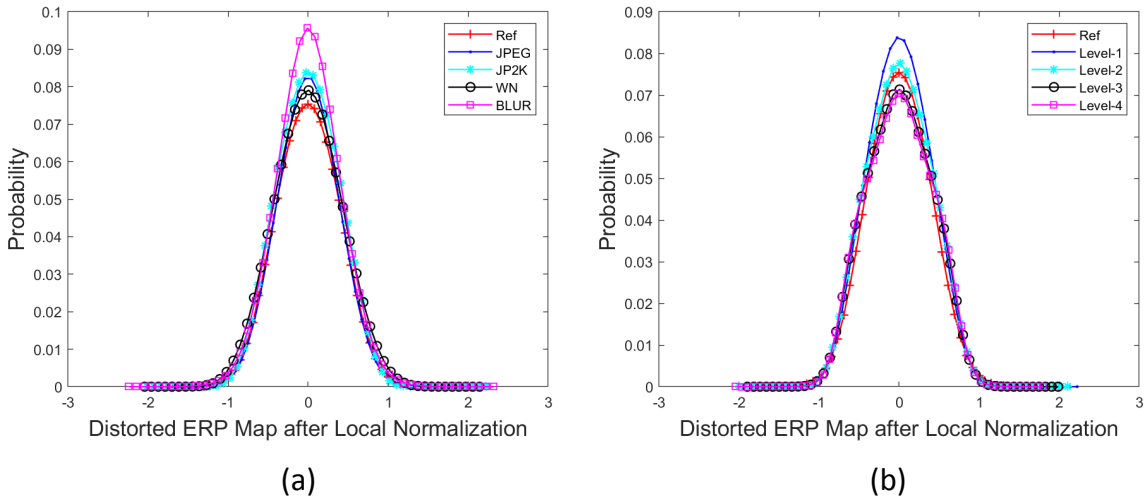

Fig. 6: Statistical distributions of MSCN coefficients for different distortion types and levels from the OIQA database [20]. (a) MSCN coefficients vary with various distortion types; (b) MSCN coefficients under four JPEG2000 compression levels.

and τ denotes the shape parameter, which can control the shape of the statistical distribution. σ_l and σ_r are the scales of the left and right sides for the statistical distribution, respectively. $\Gamma(\cdot)$ is the gamma function that is defined as:

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx, \ a > 0. \tag{16}$$

The best AGGD fitting with parameters $(\eta, \tau, \sigma_l^2, \sigma_r^2)$ is then calculated, and η is given by:

$$\eta = (\upsilon_l - \upsilon_r) \frac{\Gamma(\frac{2}{\tau})}{\Gamma(\frac{1}{\tau})}. \tag{17}$$

Furthermore, if $\sigma_l = \sigma_r$, the AGGD becomes the GGD model as follows:

$$f(x; \tau, \sigma^2) = \frac{\tau}{2\upsilon\Gamma(\frac{1}{\tau})} e^{-(\frac{|x|}{\upsilon})^{\tau}}, \tag{18}$$

where

$$\upsilon = \sigma \sqrt{\frac{\Gamma(\frac{1}{\tau})}{\Gamma(\frac{3}{\tau})}}. \tag{19}$$

In addition, the distorted 360-degree images are down-sampled by a factor of 2. Finally, the two scales, including the original image scale and a reduced resolution scale, are exploited to extract the global NSS features F_{GNSS} from the projected ERP maps.

D. Local Naturalness

When observers browse VR images in the HMD, they can freely change the viewports. Moreover, different FoVs have various contents, and global 360-degree scenery can be reconstructed based on what humans see from multiple viewports. Therefore, apart from the global naturalness from the projected ERP maps, it is also important to explore the local naturalness according to a variety of FoVs. As illustrated in Fig. 7 (a), we show an example of a global ERP map and local images from different spatial positions (i.e., the four FoVs). (b) shows the corresponding statistical distributions of the MSCN coefficients. The probability distributions of the global ERP map and four FoVs are different from each other. In addition, the distributions of FoV-2 and FoV-3 near the equator are close. Thus, we adopt the local NSS from these FoVs located at various spatial positions as the complementary features to global NSS from the projected ERP maps.

Specifically, assume that there exist M FoV images, and we extract the NSS features from them by the same steps as global naturalness. We then can obtain $(F_{LNSS}^1, F_{LNSS}^2 ..., F_{LNSS}^M)$ representing all these viewports. Finally, the average of the NSS features is computed by:

$$F_{LNSS} = \frac{1}{M} \sum_{m=1}^{M} F_{LNSS}^m, \tag{20}$$

where F_{LNSS} indicates the local NSS features from viewed FoVs. We combine the local and global NSS features to constitute the local-global naturalness measurement as follows:

$$F_{LGN} = [F_{LNSS}, \ F_{GNSS}]. \tag{21}$$

To achieve the local-global naturalness measurement, we need to sample M viewports from each distorted 360-degree image. Motivated by the fact that VR images are viewed on a sphere and polar regions usually stretch, resulting in geometric deformation for projected ERP maps, we employ the nonuniform sampling strategy [55], [56]. As shown in Fig. 8, for the equator, we first sample M_0 viewports equidistantly. Then, the remaining viewpoints are selected by:

$$\theta = \frac{360^\circ}{M_0}, \tag{22}$$

$$M_1 = \lfloor M_0 \cos\theta \rfloor, \tag{23}$$

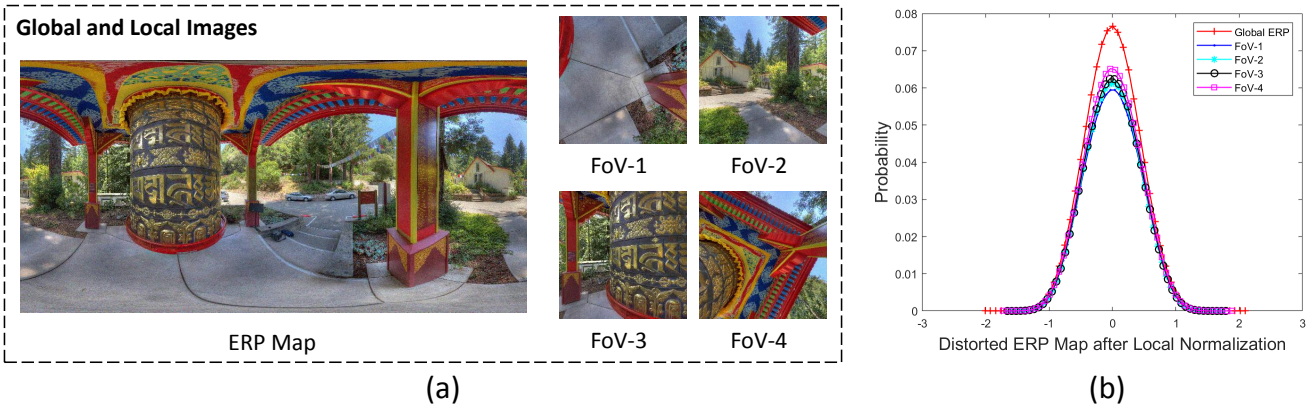

Fig. 7: An example of global and local images from different spatial positions with their corresponding statistical distributions of MSCN coefficients.

$$M_2 = \lfloor M_0 \cos 2\theta \rfloor, \tag{24}$$

$$M_{end} = \left\lfloor M_0 \cos \frac{90^\circ}{\theta} \right\rfloor, \tag{25}$$

where M_1 and M_2 are the numbers of viewports sampled on θ and 2θ degrees north or south latitude, respectively. The sampling process ends when the maximum latitude reaches 90° with M_{end} sampled viewports. After the computation of viewport sampling, we can obtain a total of M viewports for each distorted OI as:

$$M = M_0 + \sum_{m=1}^{end} 2 * M_m. \tag{26}$$

E. Quality Regression

With the multifrequency information measurement and local-global naturalness measurement, the ultimate quality score of 360-degree images is obtained by the well-known SVR [57]. Specifically, we randomly divide all distorted OIs into a training set and a testing set, which are denoted by χ_{train} and χ_{test}, respectively. By adopting SVR, our proposed MFILGN model is achieved by training the distortion-related features of OIs in the training set χ_{train} and their corresponding MOS values. Given a distorted 360-degree image $D_{train} \in \chi_{train}$ and its extracted features $[F_{MFI}, F_{LGN}]$ from the multifrequency information measurement and local-global naturalness measurement, the MFILGN model is defined by:

$$MFILGN = SVR\_TRAIN([F_{MFI}, F_{LGN}], [Q]), \tag{27}$$

where Q is the subjective quality rating (i.e., MOS) of the input 360-degree image.

After training the proposed MFILGN model, we verify the model performance on the testing set χ_{test}. For example, the predicted quality score of a tested 360-degree image $D_{test} \in \chi_{test}$ that does not appear in the training set is computed as follows:

$$q = SVR\_PREDICT([\widehat{F}_{MFI}, \widehat{F}_{LGN}], MFILGN), \tag{28}$$

where $[\widehat{F}_{MFI}, \widehat{F}_{LGN}]$ represent the extracted features from the multifrequency information measurement and local-global naturalness measurement for the tested 360-degree image. Finally, the correlation or error between the predicted scores and the corresponding ground truth MOS values for the testing set is measured as the performance of MFILGN.

IV. Experimental Results and Analysis

In this section, we first introduce the experimental protocol, including the OIQA databases and measurement criteria used in our experiments. Then, we evaluate the proposed MFILGN for overall performance and performance for individual distortion types on the OIQA [20] and CVIQD [17], [18] databases. After that, various weighting methods of viewports, as well as different parameters containing the adopted viewport numbers and training percentages, are analyzed. Finally, an ablation study is conducted to prove the effectiveness of each component in our MFILGN model.

A. Experimental Protocol

1) Databases: Two benchmark OIQA databases are utilized in the experiments, which consist of the OIQA [20] and CVIQD [17], [18] databases.

- **OIQA** comprises 16 pristine images and 320 distorted OIs degraded by 4 distortion types and 5 distortion levels. Among the distortion types, two kinds of compression artifacts are involved, namely, JPEG and JPEG2000 compression. The remaining distortion types are Gaussian blur and Gaussian noise. The subjective quality ratings in the form of MOS are provided in the range [1, 10]. A higher MOS indicates better perceptual image quality.
- **CVIQD** contains 528 compressed images derived from 16 original images. It adopts three popular image/video coding technologies, i.e., JPEG, H.264/AVC, and H.265/HEVC. Moreover, the MOS values in this

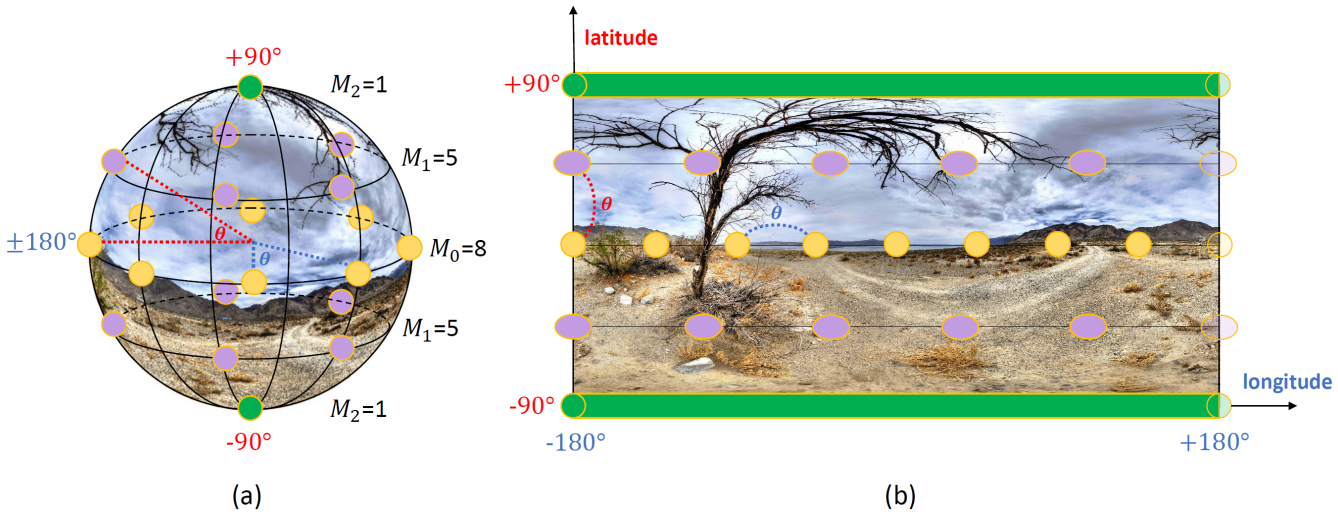

Fig. 8: Demonstration of sampling viewports when $M_0 = 8$ and $\theta = 45°$ for the sphere and plane, respectively.

TABLE I: OVERALL PERFORMANCE COMPARISONS ON THE OIQA AND CVIQD DATABASES. THE BEST RESULTS ARE DENOTED IN BOLD.

	Database	OIQA			CVIQD		
Type	Methods	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
FR	PSNR	0.5226	0.5812	1.7005	0.6239	0.7008	9.9599
	S-PSNR [40]	0.5399	0.5997	1.6721	0.6449	0.7083	9.8564
	WS-PSNR [41]	0.5263	0.5819	1.6994	0.6107	0.6729	10.3283
	CPP-PSNR [42]	0.5149	0.5683	1.7193	0.6265	0.6871	10.1448
	SSIM [29]	0.8588	0.8718	1.0238	0.8842	0.9002	6.0793
	MS-SSIM [30]	0.7379	0.7710	1.3308	0.8222	0.8521	7.3072
	FSIM [31]	0.8938	0.9014	0.9047	0.9152	0.9340	4.9864
	DeepQA [36]	0.8973	0.9044	0.8914	0.9292	0.9375	4.8574
NR	BRISQUE [32]	0.8331	0.8424	1.1261	0.8180	0.8376	7.6271
	BMPRI [33]	0.6238	0.6503	1.5874	0.7470	0.7919	8.5258
	DB-CNN [37]	0.8653	0.8852	0.9717	0.9308	0.9356	4.9311
	MC360IQA [49]	0.9139	0.9267	0.7854	0.9428	0.9429	4.6506
	VGCN [50]	0.9515	0.9584	0.5967	0.9639	0.9651	3.6573
	Proposed MFILGN	**0.9614**	**0.9695**	**0.5146**	**0.9670**	**0.9751**	**3.1036**

database are normalized and rescaled to the range [0, 100].

2) Measure Criteria: To validate the effectiveness of our proposed MFILGN and compare with other state-of-the-art methods, three commonly used measurement criteria [58] are employed, as described below.

- **Spearman rank-order correlation coefficient (SROCC)** is computed by:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} {d_i}^2}{N(N^2-1)}, \tag{29}$$

where N is the number of image samples. d_i indicates the rank difference between the subjective and objective evaluations for the $i-th$ image.

- **Pearson linear correlation coefficient (PLCC)** is calculated as:

$$PLCC = \frac{\sum_{i=1}^{N}(s_i - \mu_{s_i})(o_i - \mu_{o_i})}{\sqrt{\sum_{i=1}^{N}(s_i - \mu_{s_i}) * \sum_{i=1}^{N}(o_i - \mu_{o_i})}}, \tag{30}$$

where s_i and o_i denote the $i-th$ subjective and mapped objective quality values. μ_{s_i} and μ_{o_i} represent the corresponding mean values of s_i and o_i, respectively.

- **Root mean squared error (RMSE)** is defined by:

TABLE II: PERFORMANCE COMPARISONS FOR INDIVIDUAL DISTORTION TYPE ON THE CVIQD DATABASE. THE BEST RESULTS ARE DENOTED IN BOLD.

	Database	JPEG			AVC			HEVC		
Type	Methods	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
FR	PSNR	0.6982	0.8682	8.0429	0.5802	0.6141	10.552	0.5762	0.5982	9.4697
	S-PSNR [40]	0.7172	0.8661	8.1008	0.6039	0.6307	10.3760	0.6150	0.6514	8.9585
	WS-PSNR [41]	0.6848	0.8572	8.3465	0.5521	0.5702	10.9841	0.5642	0.5884	9.5473
	CPP-PSNR [42]	0.7059	0.8585	8.3109	0.5872	0.6137	10.5615	0.5689	0.6160	9.3009
	SSIM [29]	0.9582	0.9822	3.0468	0.9174	0.9303	4.9029	0.9452	0.9436	3.9097
	MS-SSIM [30]	0.9047	0.9636	4.3355	0.7650	0.7960	8.0924	0.8011	0.8072	6.9693
	FSIM [31]	0.9639	0.9839	2.8928	0.9439	0.9534	4.0327	0.9532	0.9617	3.2385
	DeepQA [36]	0.9001	0.9526	4.9290	0.9375	0.9477	4.2683	0.9288	0.9221	4.5694
NR	BRISQUE [32]	0.9031	0.9464	5.2442	0.7714	0.7745	8.4573	0.7644	0.7548	7.7455
	BMPRI [33]	0.9562	0.9874	2.5597	0.6731	0.7161	9.3318	0.6715	0.6154	9.3071
	DB-CNN [37]	0.9576	0.9779	3.3862	0.9545	0.9564	3.9063	0.8693	0.8646	5.9335
	MC360IQA [49]	0.9693	0.9698	3.9517	0.9569	0.9487	4.2281	0.9104	0.8976	5.2557
	VGCN [50]	**0.9759**	**0.9894**	**2.3590**	0.9659	0.9719	3.1490	0.9432	0.9401	4.0257
	Proposed MFILGN	0.9591	0.9862	2.7904	**0.9683**	**0.9785**	**2.4998**	**0.9485**	**0.9581**	**3.3950**

TABLE III: PERFORMANCE RESULTS FOR VARIOUS WEIGHTING STRATEGIES ON THE OIQA AND CVIQD DATABASES.

Database	OIQA			CVIQD		
Methods	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
Average Weighting	0.9614	0.9695	0.5146	0.9670	0.9751	3.1036
Location Weighting	0.9607	0.9688	0.5213	0.9665	0.9748	3.1212
Content Weighting	0.9598	0.9681	0.5252	0.9667	0.9749	3.1073

TABLE IV: PERFORMANCE RESULTS FOR DIFFERENT VIEWPORT NUMBERS ON THE OIQA AND CVIQD DATABASES.

Database	OIQA			CVIQD		
Number	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
6	0.9608	0.9691	0.5155	0.9665	0.9746	3.1160
20	0.9614	0.9695	0.5146	0.9670	0.9751	3.1036
80	0.9616	0.9696	0.5134	0.9678	0.9758	3.0530

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (s_i - o_i)^2}{N}}. \tag{31}$$

In addition, each OIQA database is randomly divided into 80% for training and the remaining 20% for testing. We perform 1,000 cross-validation iterations on each database. The median SROCC, PLCC and RMSE performance values are then taken as the final measurement. Before calculating the PLCC and RMSE for different objective quality assessment approaches, a five-parameter logistic nonlinear fitting function is used to map the predicted quality scores to a common scale as follows:

$$g(x) = \beta_1 (\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}}) + \beta_4 x + \beta_5, \tag{32}$$

where $\{\beta_i | i = 1, 2, ..., 5\}$ are five parameters to be fitted. x and $g(x)$ denote the raw objective quality score and the regressed quality score after the nonlinear mapping process.

In addition, the abovementioned three measurement criteria can reflect different aspects of the performance for various IQA algorithms. Specifically, SROCC is generally used to measure prediction monotonicity, while PLCC and RMSE indicate prediction accuracy. Note that higher correlation coefficients and lower error mean better performance.

B. Performance Comparison with Existing Objective Models

To demonstrate the effectiveness of our proposed MFILGN model, we conduct extensive experiments to compare with existing FR and NR objective image quality assessment algorithms. The FR models include conventional FR IQA and OIQA approaches (i.e., PSNR, SSIM [29], MS-SSIM [30], FSIM [31], S-PSNR [40], WS-PSNR [41] and CPP-PSNR [42]) and a deep learning-based FR IQA method (i.e., DeepQA [36]). Moreover, the NR models consist of traditional NR IQA approaches (i.e., BRISQUE [32] and BMPRI [33]) and three deep learning-based NR IQA and OIQA methods (i.e., DB-CNN [37], MC360IQA [49] and VGCN [50]). Among these existing state-of-the-art objective FR image quality assessment models, PSNR-related metrics, including PSNR, S-PSNR, WS-PSNR and CPP-PSNR, are signal fidelity measurements that compute the pixel differences between the reference and distorted images. Considering the characteristics of the HVS,

TABLE V: PERFORMANCE RESULTS FOR DIFFERENT IMAGE DECOMPOSITION TIMES BY DHWT ON THE OIQA AND CVIQD DATABASES.

Database	OIQA			CVIQD		
Time	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
1	0.9614	0.9695	0.5146	0.9670	0.9751	3.1036
2	0.9632	0.9716	0.4962	0.9671	0.9751	3.0992
3	0.9640	0.9715	0.4986	0.9675	0.9754	3.0945

the SSIM and its variants (i.e., MS-SSIM and FSIM) extract structural information from original and distorted images for perceptual image quality assessment. In addition, DeepQA considers human visual sensitivity in the deep learning framework. For NR methods, the BRISQUE is based on NSS features in the spatial domain and designed for conventional 2D IQA, and the BMPRI generates multiple pseudoreference images and exploits local binary pattern features for quality estimation. Additionally, the DB-CNN uses bilinear pooling for predicting the perceptual image quality in the architecture of CNN. It is worth noting that MC360IQA and VGCN are two deep learning-based methods specifically designed for 360-degree images. The MC360IQA utilizes 6 parallel subnetworks for viewport images, while the VGCN builds the graph convolution network based on different viewports.

Table I shows the overall performance comparisons on the OIQA [20] and CVIQD [17], [18] databases. The best experimental results are highlighted in bold. The compared performance values are from [50]. For a fair comparison, the performance values of traditional image quality assessment models are tested on the used testing data. Moreover, the learning-based models are trained on each specific 360-degree image quality database (i.e., OIQA database or CVIQD database) with 80% randomly selected training data and then tested on the remaining 20% testing data. We can see that the PSNR-based metrics are inferior to other objective models considering the HVS properties. This is a common phenomenon because only signal errors are involved in the framework of PSNR-related models, which is far from human perception. The 2D IQA models show unsatisfactory performance because they do not consider the specific characteristics of 360-degree images, such as the multiple viewports that are important for quality perception when browsing VR visual contents. Moreover, deep learning-based models have advantages over traditional objective image quality assessment approaches for both FR and NR categories, especially MC360IQA and VGCN, which are two deep learning-based methods specifically developed for VR images. In addition, our proposed MFILGN achieves the best performance among the existing state-of-the-art methods, including the deep learning-based OIQA and NSS-based IQA algorithms.

C. Performance Validity of Individual Distortion Types

Since a variety of distortion types exist in current OIQA databases, we validate the performance regarding each distortion type. As shown in Table II, the performance comparisons for individual distortion types are illustrated, and the best results are denoted in bold. We can observe in this table that our proposed MFILGN outperforms other state-of-the-art models in terms of AVC and HEVC artifacts. For JPEG compression distortion, we can see that VGCN delivers the best performance. One possible explanation may be that the selected viewports of VGCN adopt graph modeling, which can better capture block effects caused by JPEG compression. In addition, even compared to deep learning-based methods specifically designed for 360-degree images (i.e., MC360IQA and VGCN), the proposed MFILGN can still achieve promising performance in the case of JPEG artifacts.

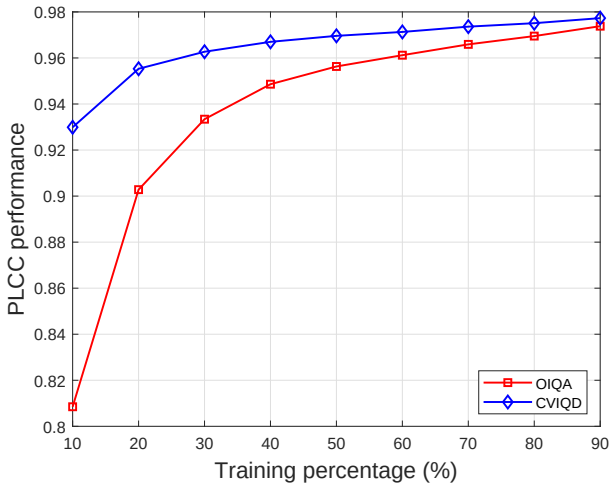

Fig. 9: The change in PLCC performance for our proposed MFILGN regarding different training percentages on the OIQA and CVIQD databases.

D. Performance of Various Weighting Methods

In the proposed MFILGN framework, we first extract local NSS features from multiple viewports. After feature extraction, we then compute the average of these local NSS features to obtain the final local naturalness representations for each 360-degree image. In addition to the average weighting, there exist some other feature aggregation strategies, such as location weighting and content weighting [56]. To be more specific, the location weighting strategy considers the statistics of eye-tracking data and uses the viewing probability to serve as the location weights for different viewport images. The content weighting method is based on the spatial information that reflects the spatial details of viewed regions.

We present the performance results for various weighting strategies on the OIQA [20] and CVIQD [17], [18] databases, as illustrated in Table III. We find that our proposed MFILGN algorithm is insensitive to different weighting methods, which demonstrates the robustness of the proposed model. Therefore, for the sake of simplification, we choose the average weighting in the proposed MFILGN.

E. Effects of Different Parameters

Since viewport extraction is involved in our model, it is interesting to explore how the performance will be affected by

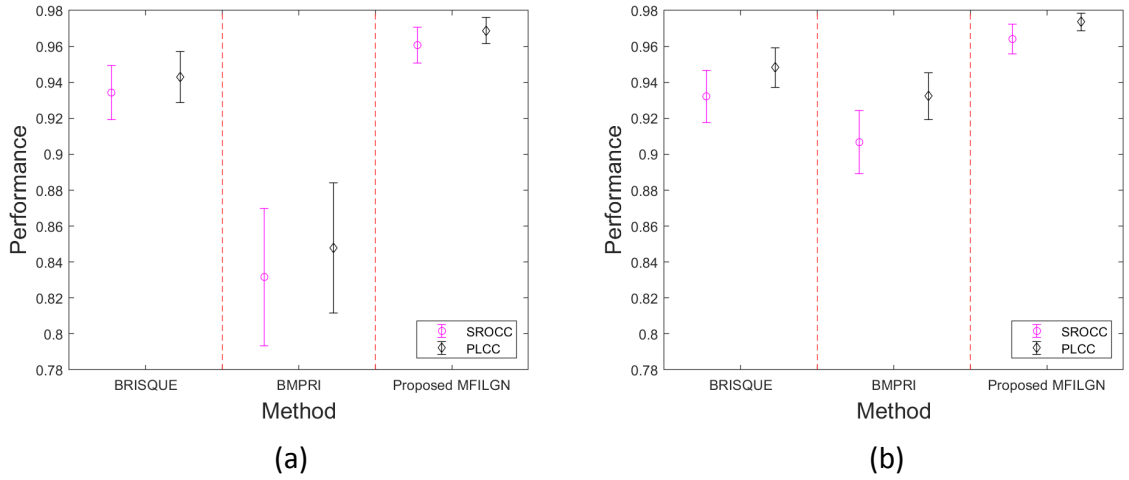

Fig. 10: Mean performance values and standard error bars for machine learning-based algorithms across 1,000 train-test trials. (a) Run on OIQA database; (b) run on CVIQD database.

different viewport numbers to be extracted. We test the performance results with respect to various viewport numbers on the OIQA [20] and CVIQD [17], [18] databases, as shown in Table IV. Three cases are considered, which include numbers equal to 6, 20, and 80. As we can observe from this table, the performance of our proposed MFILGN improves as the number of extracted viewports increases. However, the increasing viewport numbers inevitably create more computational complexity. Thus, to find the balance between performance and computation, 20 viewports for each 360-degree image are utilized in our model, which is demonstrated in Fig. 8.

Furthermore, we change the DHWT decomposition times to see if using more image decompositions will lead to better performance. From the results listed in Table V, we can observe that by adding the number of DHWT layers, i.e., using more DHWT decomposition times, our model has a small performance improvement. However, the increased image decomposition layers may need to extract more features, which results in more computational time. To reduce the computational complexity, we choose one-time DHWT for measuring multifrequency information, which can obtain the tradeoff between performance values and computational complexity.

In addition, we validate the change in the PLCC performance of our proposed MFILGN regarding different training percentages. As presented in Fig. 9, in general, a large quantity of training data achieves a performance increase on both the OIQA [20] and CVIQD [17], [18] databases. In the proposed method, we choose 80%-20% for the training-testing split because this is a common practice for perceptual quality assessment in the literature [9], [27], [28]. In addition, since the data distributions of the OIQA and CVIQD databases are very different, the correlation performance for these two databases may be disparate. Even when using 10% training data, the MFILGN model can still deliver quite competitive performance results, especially for the CVIQD database, which further demonstrates the effectiveness of our proposed MFILGN method. One possible explanation for the performance differences between the two databases may be that the OIQA database seems more challenging than the CVIQD database.

F. Statistical Significance Analysis

Since the compared BRISQUE [32], BMPRI [33], and our proposed MFILGN are all based on the machine learning model called support vector regression, we repeat the process of database splitting 1,000 times to compare the mean and standard deviation of performance values. We show the performance results in Fig. 10, where the mean and standard deviations (std) of the SROCC and PLCC values across the 1,000 trials for three algorithms are illustrated. As seen in this figure, the proposed MFILGN method can achieve a higher mean value and smaller std compared to the others, which further suggests that MFILGN performs more precisely and consistently.

G. Validity of Individual Proposed Quality Measure

We explore the effectiveness of each proposed component in the MFILGN framework, namely, multifrequency information, local naturalness, global naturalness, and local-global naturalness. The performance values are provided in Table VI. We can see that the local naturalness achieves the best performance, demonstrating the importance of viewports in evaluating the perceptual quality of 360-degree images. In addition, the multifrequency information measurement can be used as a supplement to local and global naturalness features to further improve the performance of our proposed model. Especially for the OIQA database, by adding the multifrequency information measurement, the SROCC performance improves from 0.9495 to 0.9614.

TABLE VI: ABLATION STUDY ON THE OIQA AND CVIQD DATABASES. THE BEST RESULTS ARE DENOTED IN BOLD.

Database	OIQA			CVIQD		
Methods	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
Multifrequency information	0.7734	0.7961	1.2666	0.7879	0.8285	7.8328
Global naturalness	0.9260	0.9410	0.7106	0.9520	0.9599	3.9300
Local naturalness	0.9460	0.9549	0.6195	0.9626	0.9723	3.2676
Local-global naturalness	0.9495	0.9593	0.5915	0.9657	0.9739	3.1753
Proposed MFILGN	**0.9614**	**0.9695**	**0.5146**	**0.9670**	**0.9751**	**3.1036**

TABLE VII: PERFORMANCE COMPARISONS FOR CROSS-DATABASE TESTING BY TRAINING ON THE CVIQD DATABASE AND TESTING ON THE OIQA DATABASE. THE BEST RESULTS ARE DENOTED IN BOLD.

Database	JPEG			JPEG2000			ALL		
Methods	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
MC360IQA [49]	0.8412	0.8898	4.3950	0.6221	0.6211	5.1294	0.6981	0.7443	5.9184
Proposed MFILGN	**0.8889**	**0.9027**	**0.9883**	**0.6781**	**0.7107**	**1.5545**	**0.7589**	**0.7885**	**1.3864**

H. Cross-Database Test

We validate the generalization capability of our proposed MFILGN model by a cross-database test, which is widely used to verify the model generalization ability. The CVIQD database has more compression distortion types than the OIQA database. Except for compression artifacts, the OIQA database contains Gaussian blur and Gaussian noise. Following [49], we train objective quality assessment models on the CVIQD database and then test JPEG and JP2000 compression for the OIQA database. The comparison results are shown in Table VII. From this table, we can find that our MFILGN outperforms the state-of-the-art MC360IQA model [49] for both JPEG and JPEG2000 compression as well as the overall performance. It is also interesting to observe that testing on the JPEG compression distortion achieves better performance compared to the JPEG2000 compression distortion. This is mainly because JPEG compression is the only common distortion t in both databases. In summary, we can conclude that the proposed MFILGN method can achieve good generalization capability.

V. CONCLUSIONS

In this paper, we present the MultiFrequency Information and Local-Global Naturalness (MFILGN) scheme for no-reference quality assessment of omnidirectional images. The proposed MFILGN method is composed of two new measurements, including multifrequency information measurements and local-global naturalness measurements. We designed this model by considering the HVS and the viewing process of 360-degree images. Specifically, based on the frequency-dependent property of the visual cortex, we first exploit multifrequency channel decomposition to obtain both low-frequency and high-frequency subbands for 360-degree images. The entropy intensities of these subbands are then used to measure the multifrequency information. Additionally, according to the viewing process, we adopt both local and global naturalness features from projected ERP maps and different viewports. The extracted features from our proposed two measures are fused by regression learning, which can predict the perceptual quality of omnidirectional images. We compare our proposed MFILGN with many state-of-the-art image quality assessment approaches on two publicly available 360-degree image quality databases. The experimental results demonstrate the superiority of our model.

We plan to develop a parametric model based on the proposed features and extend our method to omnidirectional video quality assessment. Furthermore, the optimization of VR processing systems based on our proposed blind quality assessment model is also promising in future work.

REFERENCES

[1] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger, "The impact of perception and presence on emotional reactions: a review of research in virtual reality," *Frontiers in psychology*, vol. 6, p. 26, 2015.

[2] M. Xu, C. Li, S. Zhang, and P. Le Callet, "State-of-the-art in 360 video/image processing: Perception, assessment and compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 5–26, 2020.

[3] W. Zhou, Z. Chen, and W. Li, "Stereoscopic video quality prediction based on end-to-end dual stream deep neural networks," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 482–492.

[4] W. Zhou, N. Liao, Z. Chen, and W. Li, "3D-HEVC visual quality assessment: Database and bitstream model," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2016, pp. 1–6.

[5] Y. Zhou, W. Zhou, P. An, and Z. Chen, "Visual comfort assessment for stereoscopic image retargeting," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.

[6] J. Xu, C. Lin, W. Zhou, and Z. Chen, "Subjective quality assessment of stereoscopic omnidirectional image," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 589–599.

[7] L. Shi, S. Zhao, W. Zhou, and Z. Chen, "Perceptual evaluation of light field image," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 41–45.

[8] S. Zhao, J. Xu, Y. Hu, W. Zhou, S. Liu, and Z. Chen, "How do you perceive differently from an AI-a database for semantic distortion measurement," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2019, pp. 1–5.

[9] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 721–734, 2017.

[10] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2017.

[11] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2018.

[12] J. Xu, W. Zhou, Z. Chen, S. Ling, and P. Le Callet, "Binocular rivalry oriented predictive auto-encoding network for blind stereoscopic image quality measurement," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[13] 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 Transactions on Instrumentation and Measurement*, 2020.

[14] W. Zhou, Q. Jiang, Y. Wang, Z. Chen, and W. Li, "Blind quality assessment for image superresolution using deep two-stream convolutional networks," *Information Sciences*, 2020.

[15] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *2016 Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[16] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017, pp. 1–6.

[17] W. Sun, K. Gu, G. Zhai, S. Ma, W. Lin, and P. Le Calle, "CVIQD: Subjective quality evaluation of compressed virtual reality images," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3450–3454.

[18] W. Sun, K. Gu, S. Ma, W. Zhu, N. Liu, and G. Zhai, "A large-scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison," in *2018 IEEE 20th international workshop on multimedia signal processing (MMSP)*. IEEE, 2018, pp. 1–6.

[19] M. Huang, Q. Shen, Z. Ma, A. C. Bovik, P. Gupta, R. Zhou, and X. Cao, "Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6039–6050, 2018.

[20] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual quality assessment of omnidirectional images," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.

[21] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2091–2110, 1989.

[22] Y. Yang, G. Jiang, M. Yu, and Y. Qi, "Latitude and binocular perception based blind stereoscopic omnidirectional image quality assessment for VR system," *Signal Processing*, vol. 173, p. 107586, 2020.

[23] F. Heitger, L. Rosenthaler, R. Von Der Heydt, E. Peterhans, and O. Kübler, "Simulation of neural contour mechanisms: from simple to end-stopped cells," *Vision Research*, vol. 32, no. 5, pp. 963–981, 1992.

[24] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[25] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Processing Letters*, vol. 19, no. 7, pp. 423–426, 2012.

[26] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, "Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain," *IEEE Transactions on Image Processing*, vol. 29, pp. 1802–1814, 2019.

[27] L. Shi, W. Zhou, Z. Chen, and J. Zhang, "No-reference light field image quality assessment based on spatial-angular measurement," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[28] W. Zhou, L. Shi, Z. Chen, and J. Zhang, "Tensor oriented no-reference light field image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4070–4084, 2020.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

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

[32] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[33] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508–517, 2018.

[34] X. Jin, Z. Chen, J. Lin, J. Chen, W. Zhou, and C. Shan, "A decomposed dual-cross generative adversarial network for image rain removal." in *BMVC*, 2018, p. 119.

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

[36] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1676–1684.

[37] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[38] S. Jia and Y. Zhang, "Saliency-based deep convolutional neural network for no-reference image quality assessment," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 14 859–14 872, 2018.

[39] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2200–2211, 2018.

[40] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *2015 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2015, pp. 31–36.

[41] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, 2017.

[42] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *Optics and Photonics for Information Processing X*, vol. 9970. International Society for Optics and Photonics, 2016, p. 99700C.

[43] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 12, pp. 3516–3530, 2018.

[44] Y. Zhou, M. Yu, H. Ma, H. Shao, and G. Jiang, "Weighted-to-spherically-uniform ssim objective quality evaluation for panoramic video," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*. IEEE, 2018, pp. 54–57.

[45] S. Chen, Y. Zhang, Y. Li, Z. Chen, and Z. Wang, "Spherical structural similarity index for objective omnidirectional video quality assessment," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018, pp. 1–6.

[46] H. G. Kim, H.-T. Lim, and Y. M. Ro, "Deep virtual reality image quality assessment with human perception guider for omnidirectional image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 917–928, 2019.

[47] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between vqa and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," in *Proceedings of the 26th ACM International Conference on Multimedia*, 2018, pp. 932–940.

[48] C. Li, M. Xu, L. Jiang, S. Zhang, and X. Tao, "Viewport proposal CNN for 360 video quality assessment," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 10 169–10 178.

[49] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan, and S. Ma, "MC360IQA: A multi-channel CNN for blind 360-degree image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 64–77, 2019.

[50] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[51] Y.-K. Lai and C.-C. J. Kuo, "A haar wavelet approach to compressed image quality measurement," *Journal of Visual Communication and Image Representation*, vol. 11, no. 1, pp. 17–40, 2000.

[52] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A haar wavelet-based perceptual similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 61, pp. 33–43, 2018.

[53] W. Zhu, G. Zhai, X. Min, M. Hu, J. Liu, G. Guo, and X. Yang, "Multi-channel decomposition in tandem with free-energy principle for reduced-reference image quality assessment," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2334–2346, 2019.

[54] M. Carandini, D. J. Heeger, and J. A. Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *Journal of Neuroscience*, vol. 17, no. 21, pp. 8621–8644, 1997.
[55] J. Xu, Z. Luo, W. Zhou, W. Zhang, and Z. Chen, "Quality assessment of stereoscopic 360-degree images from multi-viewports," in *2019 Picture Coding Symposium (PCS)*. IEEE, 2019, pp. 1–5.
[56] Z. Chen, J. Xu, C. Lin, and W. Zhou, "Stereoscopic omnidirectional image quality assessment based on predictive coding theory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 103–117, 2020.
[57] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207–1245, 2000.
[58] V. Q. E. Group *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," *2003 VQEG*, 2003.