Perceptual Depth Quality Assessment of Stereoscopic Omnidirectional Images

Wei Zhou, *Senior Member, IEEE*, and Zhou Wang, *Fellow, IEEE*

***Abstract*—Depth perception plays an essential role in the viewer experience for immersive virtual reality (VR) visual environments. However, previous research investigations in the depth quality of 3D/stereoscopic images are rather limited, and in particular, are largely lacking for 3D viewing of 360-degree omnidirectional content. In this work, we make one of the first attempts to develop an objective quality assessment model named depth quality index (DQI) for efficient no-reference (NR) depth quality assessment of stereoscopic omnidirectional images. Motivated by the perceptual characteristics of the human visual system (HVS), the proposed DQI is built upon multi-color-channel, adaptive viewport selection, and interocular discrepancy features. Experimental results demonstrate that the proposed method outperforms state-of-the-art image quality assessment (IQA) and depth quality assessment (DQA) approaches in predicting the perceptual depth quality when tested using both single-viewport and omnidirectional stereoscopic image databases. Furthermore, we demonstrate that combining the proposed depth quality model with existing IQA methods significantly boosts the performance in predicting the overall quality of 3D omnidirectional images.**

***Index Terms*—Depth perception, overall quality, no-reference, 3D omnidirectional images, multi-color-channel, adaptive viewport selection, interocular discrepancy, human visual system.**

I. Introduction

THERE has been a rapid development of virtual reality (VR) technology and a growing popularity of VR devices in recent years. A great amount of VR content has been acquired, stored, transmitted, and displayed in the form of 360-degree omnidirectional images (OIs) [1]. These OIs may be rendered to cover the whole $180 \times 360^{\circ}$ range on a spherical scene surrounding the viewer, and thus provide the viewer with a richer immersive quality-of-experience (QoE) in a 3D environment as compared to conventional 2D images that only occupy a restricted plane [2], [3]. Beyond the traditional OIs that consist of a single view, it is also possible to render the spherical scene with stereoscopic pairs of OIs with both left and right views, creating even richer and more realistic immersive 3D plus 360-degree viewer experiences [4]. Nevertheless, a variety of quality issues may arise in the creation, transmission and display processes of OIs [5]. In order for the device manufacturers and service providers of omnidirectional content to optimize the perceptual QoE of end consumers, there is an urgent need of accurate and easy-to-use omnidirectional image quality assessment (OIQA) and 3D OIQA methods [6]–[9].

W. Zhou is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, United Kingdom (e-mail: zhouw26@cardiff.ac.uk).

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: zhou.wang@uwaterloo.ca).

In general, OIQA approaches are divided into two main categories − subjective and objective quality assessment. Subjective quality assessment is considered the most accurate method [10]. Several subjective quality databases for OIs have been built in recent years [11]–[13], where each OI is associated with a perceptual visual quality (i.e., image quality). For 3D OIs, Chen et al. [14] constructed the LIVE 3D VR database, where six distortion types and five distortion levels were introduced to generate 450 distorted 3D OIs. The NBU-SOID subjective quality database [15] considered three kinds of classic compression artifacts. Head-mounted displays (HMDs) were adopted in a single stimulus quality evaluation setup and the 3D OIs underwent both symmetric and asymmetric distortions. Besides image quality, depth perception is another essential quality factor [16] that affects the overall user experience of 3D OIs [17]. This motivated Xu et al. [18] to create the SOLID subjective quality database, in which each 3D OI was rated by three quality dimensions − image quality, depth perception, and overall QoE.

In practice, subjective quality evaluation is often laborious, expensive, and inconvenient [19]. Objective quality assessment provides an attractive alternative that not only predicts perceptual quality automatically, but may also be integrated into working VR processing systems for optimal performance. In the literature, there have emerged a number of objective OIQA models [20]–[22]. Most of these methods have targeted at image quality, but none of them assesses depth perception of 3D OIs, which is an essential quality dimension in 3D visual QoE. In this work, we focus on designing an effective depth quality measure of 3D OIs, and explore the impact of depth quality assessment to the overall QoE of 3D OIs. To the best of our knowledge, this is the first attempt to address this challenging problem.

From the perspective of computational models, unlike IQA that is often based on the characteristics of viewed images such as structure, texture and content [23], depth quality assessment (DQA) is harder to deal with, mainly because of the complex 3D vision mechanisms for the binocular human visual system (HVS), especially in the more immersive case of 3D OIs. More importantly, an efficient DQA model should be beneficial to overall QoE prediction. According to the study of psychophysics and neuroscience, binocular depth depends on the ability of the HVS to precisely match corresponding feature representations in the left and right eyes [24]–[27]. Therefore, we employ the interocular discrepancy of left and right views. According to the proposed multi-color-channel and adaptive viewport selection scheme, a blind/no-reference depth quality index (DQI) is designed for assessing the depth quality of both

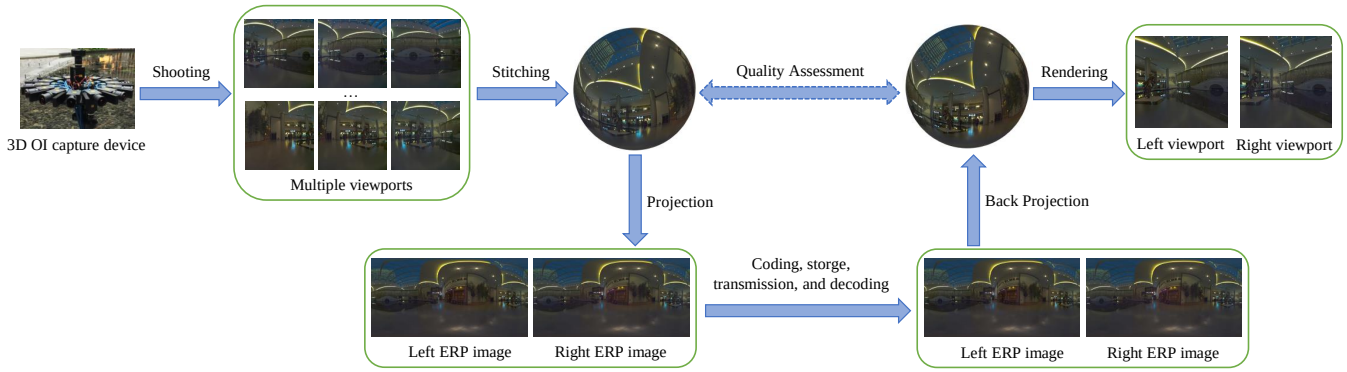

Fig. 1: An end-to-end 3D omnidirectional content processing pipeline.

single-viewport and omnidirectional stereoscopic images. In addition to the depth quality evaluation, we further extend the depth quality measure to the modeling of the overall QoE of 3D OIs.

The main contributions of this work are summarized as follows:

- The interocular discrepancy is proved to be a good depth quality indicator, which can also save computational complexity. The adaptive viewport and region selection approaches are proposed based on the discriminative depth information of interocular discrepancy, for 3D OIs and common 3D images, respectively.
- Motivated by the perceptual color peculiarity of the HVS, we resort to using color decomposition for interocular discrepancy maps, which is more consistent with the human perception. Based on the depth perception reflecting global properties of input signals and frequency independent mechanism, the interocular discrepancy statistics of decomposed frequency subbands are extracted for the final depth quality regression.
- We extend the proposed DQI to overall quality assessment of 3D OIs, leading to depth-guided overall QoE measure. We find that by integrating our DQI into classical IQA features, the performance of overall quality prediction for 3D OIs can be significantly boosted, which further demonstrates the effectiveness of our proposed depth quality measure.

The rest of this paper is organized as: In Section II, we provide the review of related works, including objective IQA as well as DQA methods for both traditional and omnidirectional images, and then give the motivations of our work. The proposed depth quality measure is described in Section III. Extensive experimental results are presented in Section IV, and we conclude the paper in Section V.

II. Related Works and Motivations

In this section, the related objective quality assessment methods are first reviewed, including 2D IQA, OIQA, 3D IQA, 3D OIQA, and DQA models. Then, we present the motivations of our proposed DQI and the depth-guided overall QoE measure.

A. Related Objective Quality Assessment Models

Since humans are the ultimate receivers of most visual signals, the goal of objective quality assessment is to predict the human-perceived quality. Therefore, the average quality ratings from subjective tests are usually adopted for creating the ground-truth labels for image quality, typically in the form of the mean opinion score (MOS) or difference mean opinion score (DMOS). Based on the availability of original reference images, the objective quality assessment methods generally have three types that consist of full-reference (FR), reduced-reference (RR), and no-reference (NR) quality assessment which is also known as blind quality assessment. The NR models are the most applicable and challenging approaches in real-world, but due to the fully accessible pristine information, FR models could deliver better performance compared with the others [28]. The earliest PSNR calculates the signal fidelity by pixel-to-pixel error. However, this way is inconsistent with human perception. Thus, according to the characteristics of the HVS, the structure similarity (SSIM) [29] and several variants have been proposed, such as multiscale SSIM (MS-SSIM) [30] and feature similarity (FSIM) [31], etc.

Intuitively, these 2D IQA metrics can be directly applied to the perceptual quality assessment of OIs by performing the computation process on equi-rectangular projection (ERP) format [32]. But the ERP images often contain inevitable geometric deformation. To bridge the gap between conventional 2D IQA approaches and omnidirectional characterizations, the spherical PSNR (S-PSNR) [33] was proposed to compute PSNR by sampling points on the sphere instead of ERP images. Besides, the weighted-to-spherically uniform PSNR (WS-PSNR) [34] was developed according to the assigned weights of stretching areas. Zakharchenko et al. [35] proposed the craster parabolic projection PSNR (CPP-PSNR), aiming to calculate the PSNR on the CPP plane.

Compared to 2D image data, a 3D image is composed of two 2D images, i.e., left and right views, which brings more challenges to the objective quality assessment of 3D images. For example, if left and right views have different types and degrees of artifacts, asymmetric distortion happens and it is more difficult to evaluate the asymmetrically distorted 3D images [36]. It is obvious that directly averaging the predicted quality values from left and right views cannot reveal

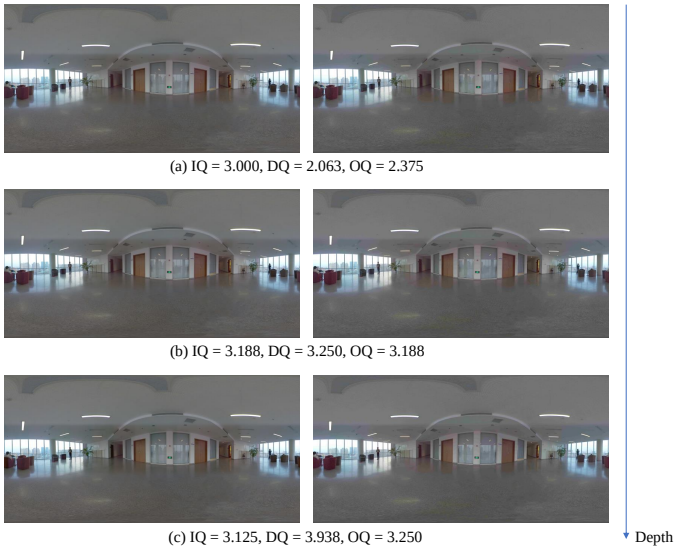

Fig. 2: Examples of JPEG compressed 3D OIs with left and right ERP images. Both left and right views have the same compression level separately (level 1 versus level 4). Subjective ground-truth labels are provided. IQ: image quality, DQ: depth quality, OQ: overall QoE.

the binocular mechanisms of the HVS [37], [38]. Therefore, several 3D IQA methods that take specific 3D characteristics into account have been proposed, e.g., the cyclopean [39] model and weighted SSIM (W-SSIM) as well as weighted FSIM (W-FSIM) [40]. For 3D OIQA, a multi-viewport based quality assessment model was proposed [41]. Furthermore, the stereoscopic omnidirectional image quality evaluator (SOIQE) [42] was developed on the basis of predictive coding theory.

The above-mentioned 3D quality assessment methods are all designed for estimating image quality rather than depth quality which is of great significance in 3D images. Hence, to tackle this problem, the depth perception difficulty index (DPDI) was presented with a prediction model [43]. Moreover, in [44], a depth perception quality metric (DPQM) was proposed. Nevertheless, they are depth perception evaluation criteria for 3D images rather than 3D OIs. In other words, there is a lack of DQA method specifically designed for 3D OIs.

B. Motivations

Typically, an end-to-end 3D omnidirectional content processing pipeline is illustrated in Fig. 1. First, multiple viewports are captured by the camera array, and then stitched onto the sphere for covering the entire field of view (FoV). Second, for ease of encoding, transmission and storage, the 3D OIs are projected to ERP format. Here, different coding artifacts would be introduced in both symmetric and asymmetric manners. Finally, after the server storage and transmission, the back projection converting the ERP format to spherical surface is performed and viewport rendering is used to display the watched scene in HMD. Such process involves the unique characteristics of 3D OIs, which should consider the perceptual combination of 3D and omnidirectional content. Additionally, different from ordinary 2D content, a special dimension in 3D vision is depth perception. Therefore, an effective DQA measure for 3D OIs that jointly considers the key factors of 3D depth perception and omnidirectional properties is highly demanded.

Fig. 2 depicts examples of three 3D OIs with left and right ERP images. Both left and right views have the same JPEG compression level separately. We find that the image quality values are very similar with each other, while the depth quality values significantly increase with more depth perception. Besides, the overall QoE is comprehensively determined by image quality and depth quality.

Motivated by the above analyses, to fill the gaps, we propose a perceptual depth quality measure that not only handles the depth quality prediction of 3D OIs, but also performs well for single-viewport, i.e., traditional 3D images. Our DQI method adopts the interocular discrepancy statistics of local viewports or regions in various frequency subbands with discriminative depth perception information. Moreover, by considering the perceptual HVS characteristics, color decomposition is used to generate different color channels for interocular discrepancy maps. Experiments on 3D OIQA and 3D IQA subjective databases verify the effectiveness of the proposed DQI and its technical components. In addition, we demonstrate that combining our DQI with image quality features can notably improve the performance results of overall QoE measure.

III. Proposed Depth Quality Index

Fig. 3 shows the proposed DQI and the depth-guided overall QoE measure. First, the interocular discrepancy of left and right views can effectively reflect depth information. Second, we use color decomposition to generate the visual signals in perceptual color space. Third, based on the observations of resulted luminance and chroma information, we adaptively extract local viewports and then decompose them into the frequency domain. Finally, interocular discrepancy statistics are adopted to be fed into the depth quality regressor which can predict the depth quality. Additionally, our proposed DQI can also be integrated with classical image quality features to produce the overall QoE.

A. Interocular Discrepancy

With our two eyes viewing the physical world from a slightly different visual perspective, we can feel the sense of depth, which provides the basis of stereopsis [45]. If the camera baseline of left and right views becomes larger, the binocular depth level increases, and vice versa. Moreover, based on some research works on psychophysics and neuroscience, the capability of the HVS to match the corresponding features in left and views determines depth perception [24]–[27]. Thus, to reflect different depth information, we here compute the interocular discrepancy as follows:

$$D = |I_{dl} - I_{dr}|, \tag{1}$$

where I_{dl} and I_{dr} denote left and right view images.

We demonstrate the interocular discrepancy maps with various depth levels in Fig. 4. As shown in this figure, the first and

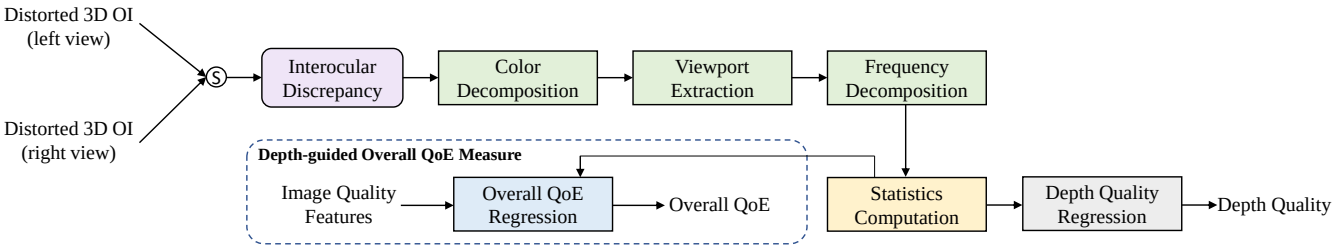

Fig. 3: Diagram of the proposed depth quality measure and the depth-guided overall QoE measure, where S represents subtraction operation.

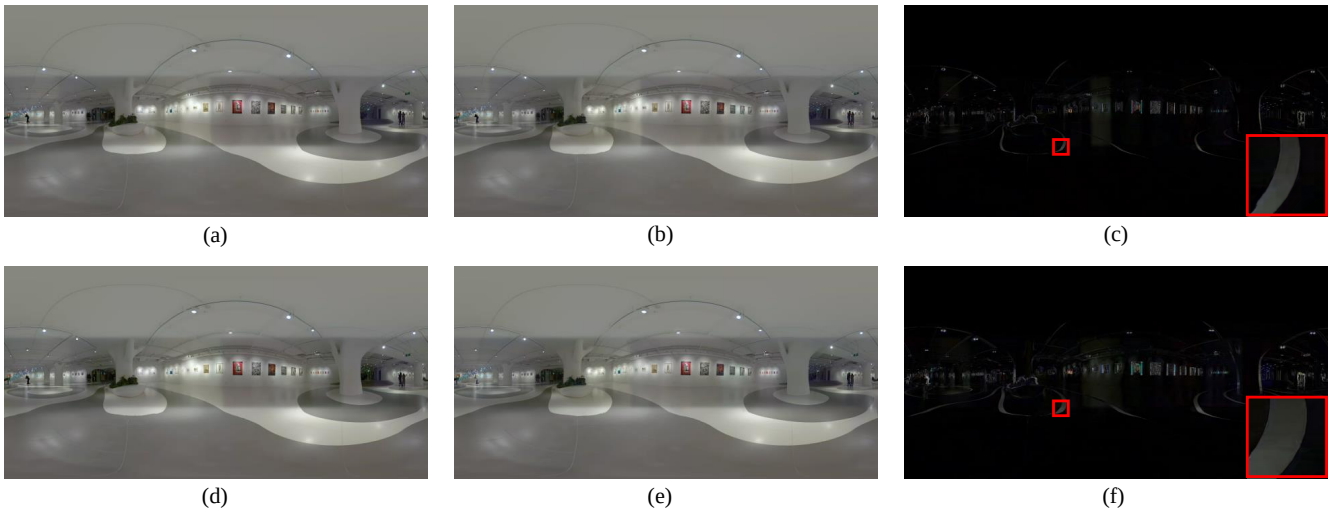

Fig. 4: Interocular discrepancy maps with different depth levels. (a) Left view with medium disparity; (b) Right view with medium disparity; (c) Interocular discrepancy map of (a) and (b); (d) Left view with large disparity; (e) Right view with large disparity; (f) Interocular discrepancy map of (d) and (e).

second rows are two stereopairs with the same compression type and degree. However, (a) and (b), i.e., the 3D OI in the first row has medium disparity, while (d) and (e), i.e., the 3D OI in the second row indicates large disparity. (c) and (f) illustrate the computed interocular discrepancy maps for the corresponding left and right views. We can observe that the interocular discrepancy maps with different depth levels behave differently. Specifically, as seen from the red bounding box, the one that has large disparity delivers more bright regions. Based on this observation, we believe the interocular discrepancy would be a good depth quality indicator.

B. Color Decomposition and Viewport Selection

1) Color Decomposition: Humans perceive vast spaces of color. As significant perceptual information of the HVS, color visual cues help the human brain to better understand the physical world [46]. Many research works have demonstrated the importance of both chroma and luminance on perceptual image quality [47]–[50], but none of them explore this for evaluating the visual quality of 3D OIs. Therefore, we attempt to disentangle the interocular discrepancy map into luminance and chroma components as:

$$D \sim [D_l, D_a, D_b], \tag{2}$$

where D_l is the luminance component, D_a and D_b are two chroma components. $\sim$ indicates the color decomposition operator.

An example is shown in Fig. 5 (a-c). We can see that there exist different visual appearances in these components. Besides, most discriminative information concentrates near the equator. Thus, we choose to extract multiple viewports from the equator of 3D OIs.

2) Viewport Selection: As stated in the section of color decomposition, the selected viewports from the equator of 3D OIs are extracted. In Fig. 6, we give an example of luminance component to show the specific viewport selection process on the sphere and ERP plane, respectively. It should be noted that the adaptive viewport selection process of the other two chroma components is the same as that of the luminance component. In this way, we not only extract multiple viewports containing discriminative depth information, but also avoid the

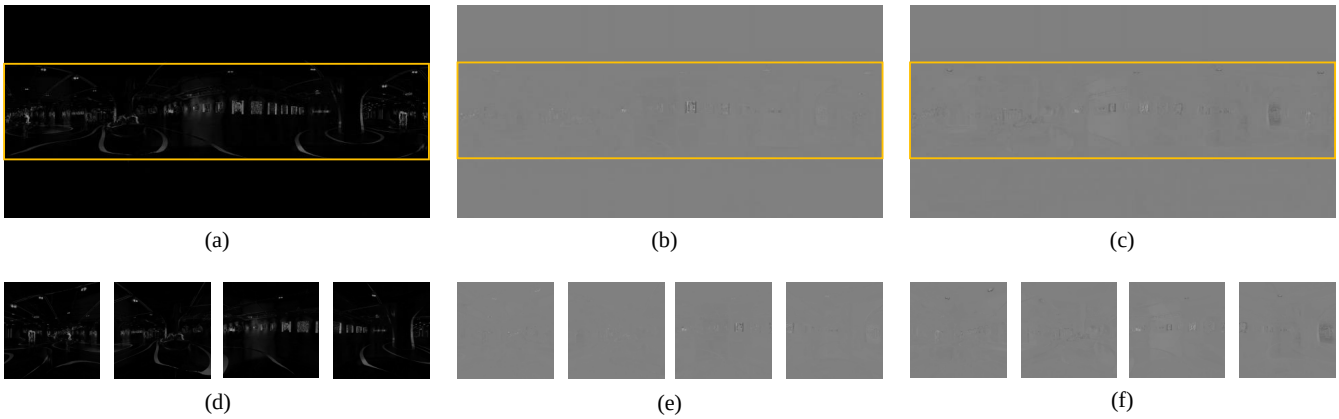

Fig. 5: Demonstration of color decomposition and viewport selection for Fig. 4 (f). (a-c) Luminance and two chroma components in LAB color space, respectively; (d-f) The corresponding viewports extracted from (a-c).

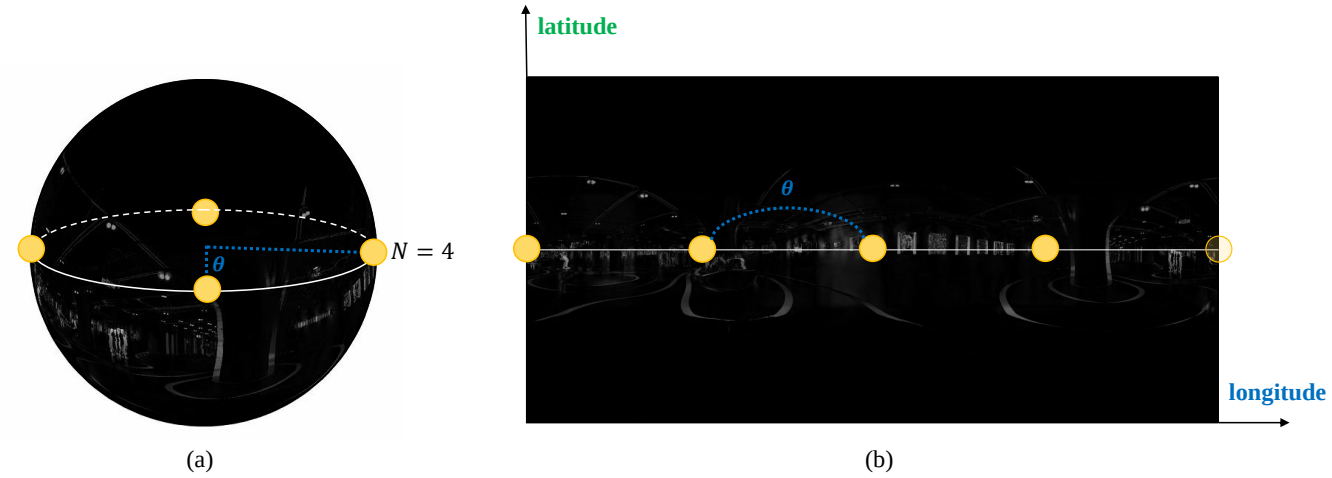

Fig. 6: Specific viewport selection process on the sphere and ERP plane. Here we take the luminance component of Fig. 5 (a) as an example. (a) Sampling viewports on the sphere; (b) Sampling viewports on the ERP plane.

geometric deformation problem. Suppose that we sample N viewports, the interval between adjacent viewports is calculated by:

$$\theta = \frac{360^\circ}{N}. \tag{3}$$

Through the viewport selection process, we can obtain multiple viewports denoted by D_{ln}, D_{an} and D_{bn} $(n = 1, 2, ..., N)$ for the luminance and two chroma components, respectively. In our experiments, we adopt $N = 4$ for simplification.

C. Frequency Decomposition

Based on the frequency independent mechanism in the human brain, neurons are commonly stimulus selective [51]. Therefore, image signals are preferred to be treated in the transform domain with different frequencies, instead of directly being used in the original form of entire image. Among lots of transforms, the wavelet transform shows a promising correlation with the HVS [52]. In addition, motivated by the advantages of discrete Haar wavelet transform (DHWT) in perceptual quality modeling [53], [54], here we choose it to decompose the extracted viewports into subbands. Note that since adding more decomposition levels has no significant performance improvement, we use the single level DHWT which can save computational complexity.

Let W be the DHWT matrix and we then can obtain the decomposed viewports in the Haar wavelet domain. Here, we convert luminance viewports as an example:

$$WD_{ln}W^T = \begin{bmatrix} \widetilde{D}_{ln}^{LL} & \widetilde{D}_{ln}^{HL} \\ \widetilde{D}_{ln}^{LH} & \widetilde{D}_{ln}^{HH} \end{bmatrix}, \tag{4}$$

where the LL, HL, LH and HH represent subbands with low or high frequency along horizontal or vertical direction. We define the aggregation of luminance subbands after frequency decomposition as $\widetilde{D}_{ln}^{s}, s \in \{LL, HL, LH, HH\}$. Similarly, we can obtain the sets of two chroma subbands in the same way as the luminance, denoted by $\widetilde{D}_{an}^{s}$ and $\widetilde{D}_{bn}^{s}$.

(a) (b) (c)

Fig. 7: The histograms of statistics for 3D OIs with the same distortion but different depth levels. (a) Luminance component of interocular discrepancy map from 3D OI with medium depth level; (b) Luminance component of interocular discrepancy map from 3D OI with large depth level; (c) The corresponding histograms of statistics for (a) and (b). Note that we omit the zero disparity images since the pixels in interocular discrepancy maps of these images are almost in black.

As shown in Fig. 7, we illustrate the histograms of statistics for 3D OIs with the same distortion but different depth levels. Here, we take the LL subband of luminance component as an example. From this figure, we find that the histograms of statistics for interocular discrepancy can effectively distinguish various depth quality degrees.

D. Statistics for Quality Regression

Extensive HVS studies have been conducted to demonstrate that the depth perception reveals global properties of visual stimulus [55]. Moreover, many quality assessment works have verified the effectiveness of statistical features [56]–[58]. Based on the observations of various interocular discrepancy maps, we exploit the representative moment (i.e., standard deviation) and entropy intensity in our framework.

Specifically, the standard deviation of luminance subbands is calculated by:

$$d\left(\widetilde{D}_{ln}^{s}\right)=\sqrt{\mathcal{M}\left[\left(\widetilde{D}_{ln}^{s}-\mathcal{M}\left(\widetilde{D}_{ln}^{s}\right)\right)^{2}\right]}, \tag{5}$$

where $\mathcal{M}$ denotes the sample mean operator. Besides, the entropy intensity of luminance subbands can be estimated as:

$$e\left(\widetilde{D}_{ln}^{s}\right)=-\sum_{i}p_{i}^{s}\left(\widetilde{D}_{ln}^{s}\right)\log_{2}p_{i}^{s}\left(\widetilde{D}_{ln}^{s}\right), \tag{6}$$

where p_i^s is the probability of the image pixel equaling i in decomposed subbands, which is computed as follows:

$$p_{i}^{s}=\frac{K_{i}^{s}}{K}, \tag{7}$$

where K_i^s represents the number of pixels equaling to i in decomposed subbands and K is the total number of pixels.

After obtaining the statistics of luminance subbands, we average them across different viewports by:

$$d\left(\widetilde{D}_{l}^{s}\right)=\frac{1}{N}\sum_{n=1}^{N}d\left(\widetilde{D}_{ln}^{s}\right), \tag{8}$$

$$e\left(\widetilde{D}_{l}^{s}\right)=\frac{1}{N}\sum_{n=1}^{N}e\left(\widetilde{D}_{ln}^{s}\right), \tag{9}$$

By repeating the same operations as luminance channel l in Eq. (5-9), the statistics for two chroma components can be achieved, which are denoted by $d\left(\widetilde{D}_{a}^{s}\right)$, $e\left(\widetilde{D}_{a}^{s}\right)$ for chroma channel a and $d\left(\widetilde{D}_{b}^{s}\right)$, $e\left(\widetilde{D}_{b}^{s}\right)$ for chroma channel b. Then, the final statistical feature for depth quality estimation is composed by:

$$\begin{aligned}F_{Depth}=\Big[&d\left(\widetilde{D}_{l}^{s}\right),d\left(\widetilde{D}_{a}^{s}\right),d\left(\widetilde{D}_{b}^{s}\right),\\&e\left(\widetilde{D}_{l}^{s}\right),e\left(\widetilde{D}_{a}^{s}\right),e\left(\widetilde{D}_{b}^{s}\right)\Big],\end{aligned} \tag{10}$$

Finally, the quality index of our proposed DQI is obtained by the following support vector regression (SVR) [59] mapping:

$$Q_{Depth}=f\left(F_{Depth}\right), \tag{11}$$

where $f(\cdot)$ indicates the SVR function.

E. Extended Overall Quality Prediction

As an effective DQA method, our proposed depth quality measure should have the capability to assist overall QoE assessment. Thus, by combining with existing image quality features, we extend the proposed DQI to overall quality prediction, namely the depth-guided overall QoE measure, as shown in Fig. 3. In the proposed depth-guided overall QoE measure framework, we choose the simplest PSNR and a representative structural similarity measure (i.e., MS-SSIM) as the image features. To be specific, the image features of left and right local viewports are calculated as follows:

$$q\left(V_{dl}\right)=\Psi\left(V_{ol},V_{dl}\right), \tag{12}$$

$$q\left(V_{dr}\right)=\Psi\left(V_{or},V_{dr}\right), \tag{13}$$

where Ψ denotes the local PSNR or MS-SSIM. V_{ol} and V_{or} are original local viewports, while V_{dl} and V_{dr} are the corresponding distorted local viewports located at the same spatial position. It should be noted that the viewport selection is followed the same way described in Section III-B. Then, the final feature for overall quality prediction is constituted as:

$$F_{\text{Overall}} = [q(V_{dl}), q(V_{dr}), F_{\text{Depth}}]. \quad (14)$$

Similarly, we can obtain the predicted overall QoE score by the SVR function:

$$Q_{\text{Overall}} = f(F_{\text{Overall}}). \quad (15)$$

With the predicted quality scores and ground-truth subjective ratings, we can compute their correlation as the ultimate performance measurement.

IV. Experimental Results

In this section, we first introduce the experimental settings including the test databases and criteria. Then, we examine the accuracy and validity of our proposed DQI from three aspects: 1) test the performance of the proposed method for 3D OIQA; 2) simplify the viewport selection process and test the proposed DQI for 3D IQA; 3) extend our proposed depth quality measure to overall QoE prediction and verify its performance. In addition, through the whole experiments, different distortion types and the ablation studies of algorithm components are also considered to further demonstrate our proposed DQI.

A. Evaluation Protocols

In order to evaluate and compare the performance of our method with state-of-the-arts, we conduct experiments on the SOLID and Waterloo 3D Depth databases. An introduction of them is shown below:

- The SOLID database [18] includes 276 distorted 3D OIs, in which 84 images are symmetrically distorted and 192 images are asymmetrically distorted. They are impaired from 6 original reference images, involving two distortion types (i.e., JPEG compression and BPG compression) and three depth levels (i.e., zero, medium and large disparity). All images in this database are with the resolution of 8192×4096 for single view and stored in the ERP format. Each 3D OI is associated with three labels, namely image quality, depth perception quality and overall quality. These labels are defined as mean opinion scores (MOSs) ranging from 1 to 5, where higher MOSs indicate better quality. It is worth noting that this database is the only 3D OIQA database that provides depth quality.
- The Waterloo-IVC 3D Depth database [60] originates from 6 pristine texture contents, i.e., Bark, Brick, Flowers, Food, Grass and Water. By considering two depth polarizations and six depth levels, 72 pristine 3D images can be produced. They are degraded by three distortion types containing additive white Gaussian noise, Gaussian blur and JPEG compression. The degradation process can totally result in 1,296 true distorted 3D images, including either inner or outer stereopairs. The image resolution of single view is 480×360. Each distorted 3D image has a depth perception difficulty index (DPDI) as the ground-truth label. Note that higher DPDI represents lower depth perception.

TABLE I: Performance comparison of depth quality prediction on the SOLID database, where DQI- means using traditional non-uniform viewport selection method.

Types	Methods	SROCC	KROCC	PLCC
2D IQA	PSNR	0.0837	0.0596	0.0980
	SSIM [29]	0.1296	0.0884	0.1119
	MS-SSIM [30]	0.0924	0.0656	0.0878
	FSIM [31]	0.1242	0.0850	0.1207
	BRISQUE [56]	0.1768	0.1247	0.1681
	NIQE [63]	0.1461	0.1032	0.1408
	LPSI [64]	0.1644	0.1156	0.1524
	dipIQ [65]	0.0482	0.0339	0.0731
	MEON [66]	0.0988	0.0671	0.1362
	CNNIQA [67]	0.0436	0.0302	0.2560
	TRES [68]	0.0422	0.0300	0.0730
	CLIPIQA [69]	0.0375	0.0218	0.0477
2D OIQA	S-PSNR [33]	0.0753	0.0543	0.0950
	WS-PSNR [34]	0.0752	0.0541	0.0921
	CPP-PSNR [35]	0.0754	0.0547	0.0946
	MFILGN [21]	0.0138	0.0101	0.0533
3D IQA	Cyclopean [39]	0.0513	0.0327	0.1559
	Weighted SSIM [40]	0.1248	0.0854	0.1101
	Weighted FSIM [40]	0.1218	0.0808	0.1214
DQA	DPQM [44]	0.5588	0.4001	0.6524
	Proposed DQI-	0.9222	0.7673	0.9434
	Proposed DQI	**0.9299**	**0.7814**	**0.9482**

For performance comparisons in our experiments, we adopt three commonly used criteria which are Spearman rank-order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), and Pearson linear correlation coefficient (PLCC) [61], [62]. The definitions of these criteria are depicted as follows:

$$SROCC = 1 - \frac{6\sum_{t=1}^{T} k_t^2}{T(T^2-1)}, \quad (16)$$

$$KROCC = \frac{2(P_c - P_d)}{T(T-1)}, \quad (17)$$

$$PLCC = \frac{\sum_{t=1}^{T}(g_t - \bar{g})(o_t - \bar{o})}{\sqrt{\sum_{t=1}^{T}(g_t - \bar{g})^2 (o_t - \bar{o})^2}}, \quad (18)$$

where T is the size of testing set, and k_t represents the rank difference between the subjective and objective scores for the t-th image. Moreover, P_c and P_d indicate the numbers of concordant and discordant pairs. In addition, g_t and o_t denote the t-th subjective score and mapped objective score after nonlinear regression. The mean of all g_t and o_t are $\bar{g}$ and $\bar{o}$, respectively. Among these criteria, SROCC and KROCC reflect the prediction monotonicity, while PLCC reveals the prediction linearity. An excellent objective quality assessment method is expected to obtain SROCC, KROCC and PLCC close to 1.

In the experiments, each database is randomly divided into 80%-20% for training and testing. We perform 1,000 iteration times and the median values are regarded as the final measurement. As recommended by the video quality experts group (VQEG) [70], a nonlinear mapping should be used before calculating the performance by PLCC. Specifically, we employ a monotonic logistic regression function as:

$$Q'(u) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (u - \beta_3)}} \right) + \beta_4 u + \beta_5, \quad (19)$$

TABLE II: Performance results of the proposed DQI for different distortion types.

Distortion Types	SROCC	KROCC	PLCC
BPG	0.9055	0.7498	0.9241
JPEG	0.9249	0.7850	0.9666

TABLE III: Performance results of the proposed DQI for symmetrically and asymmetrically distorted 3D OIs.

Distortions	SROCC	KROCC	PLCC
Symmetric	0.8472	0.6982	0.9098
Asymmetric	0.8700	0.7064	0.9167

where u and $Q'(u)$ are the raw objective quality scores and regressed scores after the nonlinear mapping. $\{\beta_j \mid j = 1, 2, 3, 4, 5\}$ denote five parameters to be fitted.

B. Performance for Depth Quality Measure of 3D OIQA

To demonstrate the effectiveness of our proposed method, we conduct experiments on the 3D OIQA database (i.e., SOLID) and compare the DQI with state-of-the-art quality assessment models: 1) classical full-reference 2D IQA approaches including the PSNR, SSIM [29], MS-SSIM [30] and FSIM [31] as well as no-reference 2D IQA methods consisting of BRISQUE [56], NIQE [63], LPSI [64], dipIQ [65], MEON [66], CNNIQA [67], TRES [68], and CLIPIQA [69]; 2) four 2D OIQA metrics, namely S-PSNR [33], WS-PSNR [34], CPP-PSNR [35], and MFILGN [21]; 3) typical 3D IQA algorithms containing Cyclopean [39], Weighted SSIM [40] and Weighted FSIM [40]; 4) the DQA method called DPQM [44], which is a depth quality measure designed for 3D images. It should be noted that there have been no specifically designed quality metrics for 3D OIs so far.

The performance comparison results of depth quality prediction are reported in TABLE I, where the best performance values are highlighted in bold. From this table, we can find that traditional visual quality assessment models, including 2D IQA, 2D OIQA and 3D IQA methods, fail to evaluate depth quality. This is reasonable because they are designed for assessing image quality rather than depth perception. The image quality targets the perceived quality of pictures, while the depth quality means the ability to deliver an enhanced sensation of depth [17].

Besides, DPQM is a depth quality metric that performs better than those IQA models. However, only 3D images are considered in DPQM, which is not suitable for 3D OIs. To the best of our knowledge, our proposed method is the first one that specifically designs for 3D OIs. The difference between the proposed DQI- and DQI lies in the viewport selection. That is, DQI- adopts the non-uniform viewport sampling method as [41], [42] which results in six viewports (i.e., four on the equator, one at the north pole and one at the south pole), while our proposed DQI is based on the adaptive viewport selection method described in Section III-B. It can be observed that both DQI- and DQI significantly outperform DPQM, demonstrating the advantages of our designed framework. In addition, the proposed DQI is superior to DQI- because the used viewport selection strategy in this work is more likely to reflect the discriminative information of statistical characteristics for interocular discrepancy maps.

TABLE IV: Performance evaluation of the proposed DQI for different interocular discrepancy statistics.

Methods	SROCC	KROCC	PLCC
Standard Deviation	0.9196	0.7646	0.9413
Entropy Intensity	0.8952	0.7324	0.9276
Proposed DQI	**0.9299**	**0.7814**	**0.9482**

TABLE V: Performance evaluation of the proposed DQI for different color decomposition methods.

Methods	SROCC	KROCC	PLCC
HSV	0.9028	0.7402	0.9296
LAB	**0.9299**	**0.7814**	**0.9482**

C. Validity of Various Distortion Scenarios

Since there exist different distortion types and symmetric as well as asymmetric distortions are involved in the SOLID database, we test the performance regards to various distortion scenarios.

The performance results of our method for BPG compression and JPEG compression are shown in TABLE II. As can be seen in the table, the proposed DQI delivers promising performance for the two distortion types. Additionally, from the quantitative numbers of our method for symmetrically and asymmetrically distorted 3D OIs in TABLE III, we can find that DQI can handle both symmetric and asymmetric distortions. These results further verify the effectiveness of our proposed DQI under various distortion scenarios. It should be worth noting that the proposed model shows more superior performance regarding asymmetric distortion. This may be because the 3D OIs with asymmetric distortion have more explicit differences in depth perception.

D. Validity of Different Components

Considering that our proposed DQI is composed of various components, it is interesting to evaluate the performance for individual component.

First, we test the results of our method about different interocular discrepancy statistics, as illustrated in TABLE IV. We can observe that both standard deviation and entropy intensity achieve good performance. Moreover, the combination of them shows the best results. In TABLE V and TABLE VI, we then report the performance values of the proposed DQI for various color decomposition methods and color channels. The two tables show that the LAB color space and luminance channel perform better than the HSV color space and the other two chroma channels, respectively. This is mainly due to the prominent role of luminance in the HVS.

E. Performance for Depth Quality Measure of 3D IQA

Except for 3D OIs, an intuitive idea is to simplify the viewport selection process to see whether the proposed DQI can still be able to evaluate the depth quality of traditional 3D images or not. One choice is to use the entire interocular discrepancy map, which we denote as DQI-. Here, the question is

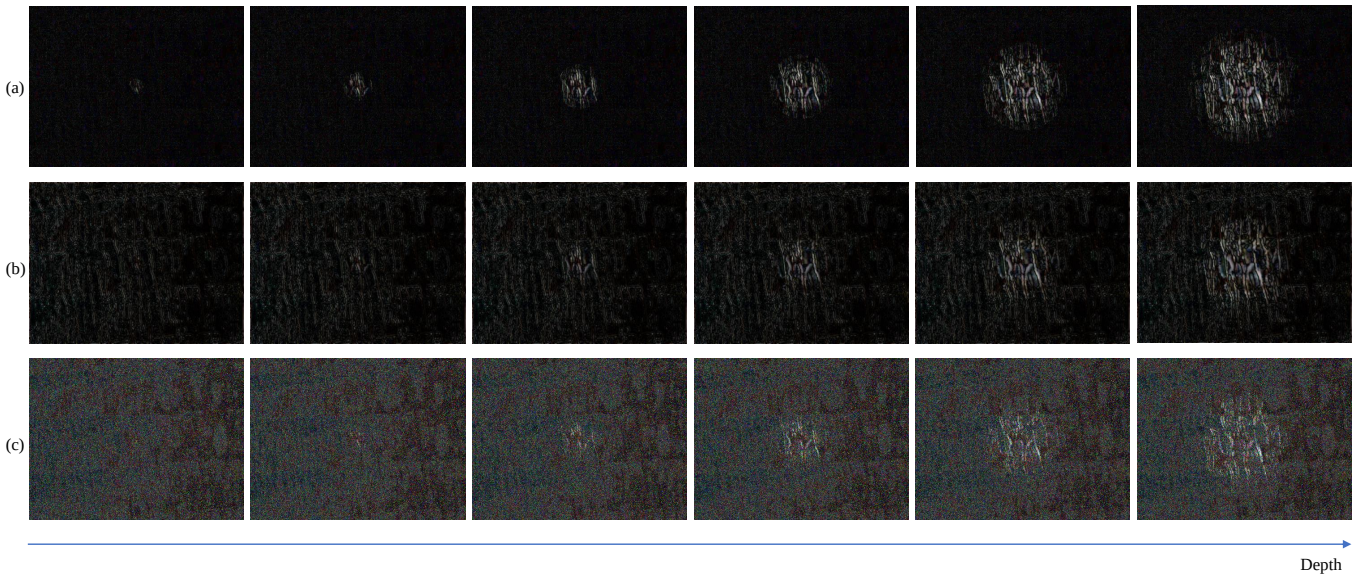

Fig. 8: Variation of interocular discrepancy maps with the increase of depth levels. Each row has the same distortion type and degree, but different depth levels. (a) JPEG compression; (b) Gaussian blur; (c) Additive white Gaussian noise.

TABLE VI: Performance evaluation of the proposed DQI for various color channels.

Channels	SROCC	KROCC	PLCC
l	**0.8871**	**0.7242**	**0.9232**
a	0.8776	0.7040	0.9176
b	0.8450	0.6665	0.8912

TABLE VII: Performance comparison of depth quality prediction on the Waterloo-IVC 3D depth database, where DQI- means using the entire interocular discrepancy map.

Methods	SROCC	KROCC	PLCC
PSNR	0.3600	0.2712	0.3239
SSIM [29]	0.3308	0.2474	0.3937
MS-SSIM [30]	0.4411	0.3276	0.5224
FSIM [31]	0.2925	0.2195	0.4342
DPDI [43]	0.8018	0.6190	0.7970
DPQM [44]	0.6355	0.4631	0.6445
Proposed DQI-	0.7910	0.6099	0.8025
Proposed DQI	**0.8365**	**0.6542**	**0.8526**

that does the whole interocular discrepancy map be necessary for depth quality estimation?

We illustrate the variation of interocular discrepancy maps with the increase of depth levels, as depicted in Fig. 8. We take the image content "Bark" as an example. Note that each row in this figure has the same distortion type and degree, but different depth levels which increase from left to right. We find that the discriminative information focuses on the map center. Let $X \times Y$ be the resolution of interocular discrepancy maps. For luminance and two chroma channels, we extract the map center as follows:

$$D_l{}' = D_l\left(\frac{1}{3}Y : \frac{2}{3}Y, \frac{1}{3}X : \frac{2}{3}X\right), \tag{20}$$

$$D_a{}' = D_a\left(\frac{1}{3}Y : \frac{2}{3}Y, \frac{1}{3}X : \frac{2}{3}X\right), \tag{21}$$

TABLE VIII: Performance comparison of overall quality prediction on the SOLID database, where Proposed- means using traditional non-uniform viewport selection method.

Types	Methods	SROCC	KROCC	PLCC
2D IQA	PSNR	0.5063	0.3512	0.5458
	SSIM [29]	0.7466	0.5459	0.7362
	MS-SSIM [30]	0.6247	0.4447	0.6376
	FSIM [31]	0.7476	0.5452	0.7468
	BRISQUE [56]	0.5206	0.3685	0.5378
	NIQE [63]	0.5870	0.4164	0.6023
	LPSI [64]	0.6608	0.4780	0.6579
	dipIQ [65]	0.5387	0.3803	0.5410
	MEON [66]	0.3952	0.2737	0.4152
	CNNIQA [67]	0.1523	0.0999	0.2401
	TRES [68]	0.2246	0.1542	0.2587
	CLIPIQA [69]	0.1399	0.0942	0.1639
2D OIQA	S-PSNR [33]	0.4754	0.3314	0.5067
	WS-PSNR [34]	0.4705	0.3273	0.4996
	CPP-PSNR [35]	0.4747	0.3308	0.5065
	MFILGN [21]	0.1058	0.0719	0.1582
3D IQA	Cyclopean [39]	0.6317	0.4295	0.6921
	Weighted SSIM [40]	0.7206	0.5204	0.7181
	Weighted FSIM [40]	0.7356	0.5333	0.7384
OQA	Proposed-	0.9203	0.7714	0.9271
	Proposed	**0.9301**	**0.7891**	**0.9359**

$$D_b{}' = D_b\left(\frac{1}{3}Y : \frac{2}{3}Y, \frac{1}{3}X : \frac{2}{3}X\right). \tag{22}$$

The above operation is employed to replace the viewport selection process, resulting in the proposed DQI. We conduct experiments on the Waterloo-IVC 3D depth database. TABLE VII provides the performance comparison results. It can be seen that our proposed DQI is superior to other state-of-the-art quality assessment models, which validates the potential advantages of our method for evaluating the depth quality of 3D images.

TABLE IX: Ablation test results of the depth-guided overall QoE measure.

Methods	SROCC	KROCC	PLCC
Local PSNR	0.5176	0.3663	0.5719
Local MS-SSIM [30]	0.6144	0.4493	0.6658
Proposed DQI	0.8248	0.6497	0.8377
Local PSNR+DQI	0.9240	0.7763	0.9265
Local MS-SSIM+DQI	**0.9301**	**0.7891**	**0.9359**

F. Performance for Overall Quality Measure of 3D OIQA

Apart from the depth quality measure, we further extend to a depth-guided overall QoE measure. Specifically, we integrate the proposed DQI into existing image quality features. Here, we adopt the well-known MS-SSIM as a representative.

In TABLE VIII, we show the performance comparison of overall quality prediction on the SOLID database, where OQA means overall quality assessment. The "Proposed-" and "Proposed" denote the non-uniform viewport sampling method and our proposed viewport selection strategy, respectively. We can see that our extended depth-guided overall QoE measure outperforms state-of-the-art approaches, and the proposed viewport selection method can boost the performance in a sense.

G. Ablation Study of Overall Quality Prediction

Since two factors are considered in our depth-guided overall quality measure (i.e., image quality and depth quality), we conduct the ablation study of overall quality prediction by using either image quality features or our proposed DQI.

The performance results are reported in TABLE IX, where we use the PSNR and MS-SSIM as image features. It can be observed that even with the simplest PSNR, the proposed DQI can still improve the results. The important boosting effects of our DQI demonstrate its validity for not only depth quality, but also overall QoE.

V. Conclusion

In this paper, we propose a DQI method for depth quality assessment based on interocular discrepancy statistics. The proposed model is inspired by the perceptual mechanisms of the HVS, where the interocular discrepancy of the left and right views is used to reflect disparity. Motivated by the color perception in the human brain, we divide the interocular discrepancy into multi-color channels. The key viewports or regions are then selected to perform frequency decomposition and statistics computation, which are on the basis of global depth perception and frequency independent principle in the HVS. Extensive experiments demonstrate that the proposed DQI offers accurate and robust evaluation of depth quality for both 3D omnidirectional images and conventional 3D images. Furthermore, we demonstrate that the proposed DQI can be extended to assess the overall quality by combining with existing image quality features, leading to a depth-guided overall QoE measure. Experimental results show that the proposed method outperforms state-of-the-art quality assessment models. Future work includes the incorporation of deeper visual models such as depth polarization and the development of comprehensive QoE assessment models jointly considering image quality, depth quality and visual discomfort aspects.

References

[1] C. Xi, "Virtual reality/augmented reality white paper," in *China Academy of Information and Communications Technology (CAICT)*. Huawei Technologies Co., 2017.

[2] Y. Zhu, G. Zhai, Y. Yang, H. Duan, X. Min, and X. Yang, "Viewing behavior supported visual saliency predictor for 360 degree videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4188–4201, 2021.

[3] 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.

[4] X. Zhou, Y. Zhang, N. Li, X. Wang, Y. Zhou, and Y.-S. Ho, "Projection invariant feature and visual saliency-based stereoscopic omnidirectional image quality assessment," *IEEE Transactions on Broadcasting*, 2021.

[5] R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard, "Visual distortions in 360° videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2524–2537, 2019.

[6] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 2002, pp. IV–3313.

[7] 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.

[8] 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.

[9] C. Tian, X. Chai, G. Chen, F. Shao, Q. Jiang, X. Meng, L. Xu, and Y.-S. Ho, "VSOIQE: A novel viewport-based stitched 360° omnidirectional image quality evaluator," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 10, pp. 6557–6572, 2022.

[10] J. Xu, J. Li, X. Zhou, W. Zhou, B. Wang, and Z. Chen, "Perceptual quality assessment of internet videos," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1248–1257.

[11] 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 *IEEE International Workshop on Multimedia Signal Processing*, 2018, pp. 1–6.

[12] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual quality assessment of omnidirectional images," in *IEEE international symposium on circuits and systems*, 2018, pp. 1–5.

[13] J. Li, K. Yu, Y. Zhao, Y. Zhang, and L. Xu, "Cross-reference stitching quality assessment for 360 omnidirectional images," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2360–2368.

[14] M. Chen, Y. Jin, T. Goodall, X. Yu, and A. C. Bovik, "Study of 3d virtual reality picture quality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 89–102, 2019.

[15] Y. Qi, G. Jiang, M. Yu, Y. Zhang, and Y.-S. Ho, "Viewport perception based blind stereoscopic omnidirectional image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[16] B. Series, "Subjective methods for the assessment of stereoscopic 3DTV systems," 2015.

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

[18] 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.

[19] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Transactions on Broadcasting*, vol. 57, no. 1, pp. 1–14, 2010.

[20] 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*, vol. 31, no. 5, pp. 1724–1737, 2020.

[21] W. Zhou, J. Xu, Q. Jiang, and Z. Chen, "No-reference quality assessment for 360-degree images by analysis of multifrequency information and local-global naturalness," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[22] J. Fu, C. Hou, W. Zhou, J. Xu, and Z. Chen, "Adaptive hypergraph convolutional network for no-reference 360-degree image quality assessment," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 961–969.

[23] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006.

[24] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors," *Science*, vol. 249, no. 4972, pp. 1037–1041, 1990.

[25] P. Neri, A. J. Parker, and C. Blakemore, "Probing the human stereoscopic system with reverse correlation," *Nature*, vol. 401, no. 6754, pp. 695–698, 1999.

[26] A. J. Parker, "Binocular depth perception and the cerebral cortex," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 379–391, 2007.

[27] S. Henriksen, S. Tanabe, and B. Cumming, "Disparity processing in primary visual cortex," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1697, p. 20150255, 2016.

[28] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2017.

[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 *Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, 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] "Equirectangular projection." [Online]. Available: https://en.wikipedia.org/wiki/Equirectangular_projection

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

[34] 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.

[35] 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.

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

[37] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *IEEE European Signal Processing Conference*, 2007, pp. 2110–2114.

[38] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–13, 2009.

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

[40] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3d images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[41] J. Xu, Z. Luo, W. Zhou, W. Zhang, and Z. Chen, "Quality assessment of stereoscopic 360-degree images from multi-viewports," in *IEEE Picture Coding Symposium*, 2019, pp. 1–5.

[42] 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.

[43] J. Wang, S. Wang, K. Ma, and Z. Wang, "Perceptual depth quality in distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1202–1215, 2016.

[44] 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.

[45] F. A. Kingdom, "Binocular vision: The eyes add and subtract," *Current Biology*, vol. 22, no. 1, pp. R22–R24, 2012.

[46] R. M. Boynton, "Human color perception," in *Science of Vision*. Springer, 1990, pp. 211–253.

[47] J. Preiss, F. Fernandes, and P. Urban, "Color-image quality assessment: From prediction to optimization," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1366–1378, 2014.

[48] D. Lee and K. N. Plataniotis, "Towards a full-reference quality assessment for color images using directional statistics," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3950–3965, 2015.

[49] Lee, Dohyoung and Plataniotis, Konstantinos N, "Toward a no-reference image quality assessment using statistics of perceptual color descriptors," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3875–3889, 2016.

[50] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *Journal of Vision*, vol. 17, no. 1, pp. 32–32, 2017.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] J. Read, "Early computational processing in binocular vision and depth perception," *Progress in Biophysics and Molecular Biology*, vol. 87, no. 1, pp. 77–108, 2005.

[56] 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.

[57] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2014.

[58] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[59] 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.

[60] J. Wang, S. Wang, and Z. Wang, "Depth perception of distorted stereoscopic images," in *IEEE International Workshop on Multimedia Signal Processing*, 2015, pp. 1–6.

[61] 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*, vol. 30, no. 11, pp. 4114–4128, 2019.

[62] 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.

[63] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[64] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *IEEE International Conference on Image Processing*, 2015, pp. 339–343.

[65] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[66] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.

[67] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

[68] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1220–1230.

[69] J. Wang, K. C. Chan, and C. C. Loy, "Exploring clip for assessing the look and feel of images," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 2555–2563.

[70] J. Antkowiak, T. Jamal Baina, F. V. Baroncini, N. Chateau, F. FranceTelecom, A. C. F. Pessoa, F. Stephanie Colonnese, I. L. Contin, J. Caviedes, and F. Philips, "Final report from the video quality experts group on the validation of objective models of video quality assessment," 2000.

BIOGRAPHY

Wei Zhou (S'19–M'21–SM'24) is an Assistant Professor at Cardiff University, United Kingdom. Dr Zhou was a Postdoctoral Fellow at University of Waterloo, Canada. Wei received the Ph.D. degree from the University of Science and Technology of China in 2021, joint with the University of Waterloo from 2019 to 2021. Wei was a visiting professor at Dalian University of Technology, visiting scholar at National Institute of Informatics, Japan, a research assistant with Intel, and a research intern at Microsoft Research and Alibaba Cloud.

Dr Zhou is now an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems, Pattern Recognition, and Neurocomputing. Wei's research interests span multimedia computing, perceptual image processing, and computational vision.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree from The University of Texas at Austin in 2001. He is currently a Canada Research Chair and Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image and video processing and coding; visual quality assessment and optimization; computational vision and pattern analysis; machine learning; multimedia communications; and biomedical signal processing. He has more than 200 publications in these fields with over 100,000 citations (Google Scholar).

Dr. Wang serves as a member of IEEE Image, Video and Multidimensional Signal Processing Technical Committee (2020-2022) and IEEE Multimedia Signal Processing Technical Committee (2013-2015), a Senior Editor of IEEE Journal of Selected Topics in Signal Processing (2022-present), a Senior Area Editor of IEEE Transactions on Image Processing (2015-2019), and an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (2016-2018), IEEE Transactions on Image Processing (2009-2014), and IEEE Signal Processing Letters (2006-2010), among other journals. He was elected a Fellow of Royal Society of Canada: Academy of Science in 2018, and a Fellow of Canadian Academy of Engineering in 2016. He is a recipient of 2021 Technology Emmy Award, 2016 IEEE Signal Processing Society Sustained Impact Paper Award, 2015 Primetime Engineering Emmy Award, 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, 2013 IEEE Signal Processing Magazine Best Paper Award, and 2009 IEEE Signal Processing Society Best Paper Award.