Segmented Spherical Projection-Based Blind Omnidirectional Image Quality Assessment

XUELEI ZHENG, GANGYI JIANG, MEI YU, AND HAO JIANG
Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
Corresponding author: Gangyi Jiang (gyjiang@nbu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871247, Grant 61671258, and Grant 61931022, and in part by the K. C. Wong Magna Fund of Ningbo University.

ABSTRACT In contrast with traditional images, omnidirectional image (OI) has a higher resolution and provides the user with an interactive wide field of view. OI with equirectangular projection (ERP) format, as the default for encoding and transmitting omnidirectional visual contents, is not suitable for quality assessment of OI because of serious geometric distortion in the bipolar regions, especially for blind image quality assessment. In this paper, a segmented spherical projection (SSP) based blind omnidirectional image quality assessment (SSP-BOIQA) method is proposed. The OI with ERP format is first converted into that with SSP format, so as to solve the problem of stretching distortion in the bipolar regions of ERP format, but retain the equatorial region of ERP format. On the one hand, considering that the bipolar regions of the SSP format are circular, a local/global perceptual features extraction scheme with fan-shaped window is proposed for estimating the distortion in the bipolar regions of OI. On the other hand, the perceptual features of the equatorial region are extracted with heat map as weighting factor to reflect users’ visual behavior. Then, the features extracted from the OI’s bipolar and equatorial regions are pooled to predict the quality of distorted OIs. The experiments on two databases, namely CVIQD2018 and MVAQD databases, demonstrate that the proposed SSP-BOIQA method outperforms the state-of-the-art blind quality assessment methods, and is more consistent with human visual perception.

INDEX TERMS Omnidirectional image, blind quality assessment, segmented spherical projection, user’s visual behavior, fan-shaped window.

I. INTRODUCTION

With the rapid development of virtual reality (VR) technologies, omnidirectional visual contents play an important role in VR systems and are used in multimedia, industry, medical care and business fields [1]. During omnidirectional image (OI) processing, compression and transmission, the degradation of the OIs may be generated [2]. Different from the ordinary images being presented on flat display, OIs are viewed with head mounted display (HMD). In this viewing mode, users can only view part of the OI at a certain time as a viewport in HMD, and the distorted viewport will be magnified, which greatly affecting the users’ experience [3]. Therefore, how to evaluate the quality of OIs effectively and objectively is an urgent problem to be solved.

In VR system, OIs are captured and represented in spherical form, however, this form is not easily to be stored or transmitted directly. Thus, the motion picture expert group (MPEG) has developed the omnidirectional media application format (OMAF) for encoding, storing, transmitting and rendering omnidirectional media [4]. In OMAF, equirectangular projection (ERP) is used as the default projection to project the original spherical signal onto a two-dimensional (2D) plane for compression and transmission with the existing encoding standards [5]. But the OI with ERP format (hereinafter, denoted as ERP-image for convenience) will lead to redundancies and geometrics distortion with the latitude, especially in the bipolar regions of ERP-image due to stretching. Hence, it is indispensable to develop the objective omnidirectional image quality assessment (OIQA) methods to evaluate and optimize the VR systems.

Generally, image quality assessment (IQA) metrics can be categorized into the full-reference (FR), reduced-reference
(RR) and blind/no-reference methods. FR-IQA metrics need all the information of the original image, RR-IQA metrics only need part information of the original image, while blind IQA (BIQA) metrics predicate the image quality without any reference information [6]. For FR-IQA metrics, some classic metrics were proposed for ordinary 2D images, such as peak signal-to-noise-ratio (PSNR), structural similarity (SSIM) [7] and visual information fidelity (VIF) [8], and so on. Because of the OI’s particularity which is different from the ordinary image, these FR-IQA metrics cannot estimate OIs’ quality accurately. Thus, Yu et al. [9] proposed spherical PSNR (S-PSNR) on the spherical surface to adapt the traditional PSNR for OIs. Sun et al. [10] noted that the distortion degree of the ERP-image is related to the latitude of corresponding pixels, and proposed weighted to spherically uniform PSNR (WS-PSNR). Unlike S-PSNR, WS-PSNR does not require projection transformation, however, it cannot be calculated across formats. Zakharchenko et al. [11] presented Craster parabolic projection PSNR (CPP-PSNR) to solve this cross-format limitation. Zhou et al. [12] proposed the weighed to spherically uniform SSIM (WS-SSIM), which analyzed the stretching distortion problem in ERP-image. Although the above methods have improved OI’s quality prediction to a certain degree, they did not consider user’s visual perception well. Therefore, Li et al. [13] utilized the characteristics of user behavior and proposed to map the user’s head movement (HM) and eye movement (EM) into deep learning model to estimate the quality of omnidirectional video. Kim et al. [14] constructed a deep-learning-based omnidirectional IQA method through quality predictor and human perception guidance module.

Since the distortion-free reference image is usually difficult to be obtained in practical applications, some blind image quality assessment (BIQA) methods were proposed to predict the quality of ordinary 2D images instead of FR-IQA methods. For example, BRISQUE [15] and OG [16] extract natural statistical features in the spatial domain or gradient domain, and pooling with the traditional machine learning methods. These BIQA methods are called opinion-aware BIQA methods. On the contrast, the opinion-unaware BIQA methods do not require subjective opinions for training, such as NIQE [17], diplIQ [18] and BPRI [19], but they cannot be universally applicable in all distortion types. Although these BIQA metrics have great consistency with the human visual system (HVS) in predicting the quality of ordinary 2D images, they cannot achieve satisfactory results in predicting the OIs’ quality. In addition, most of the objective quality evaluation approaches for OI are mainly based on the ERP format. However, ERP format has severe distortion at the bipolar regions. Since the OI distortion process will change and produce unnatural local anisotropy, especially in the circular regions, a relative gradient orientation (RGO) is used to explain the structural features of OI. Moreover, in order to adapt to image with different geometric shape, an improved circular local binary pattern (CLBP) operator is designed by replacing the square neighborhood with the circle neighborhood.

(1) Considering that the bipolar regions of SSP format are circular, a local/global perceptual feature extraction scheme with fan-shaped window is proposed for estimating the distortion in the OI’s bipolar regions. Since the OI distortion process will change and produce unnatural local anisotropy, especially in the circular regions, a relative gradient orientation (RGO) is used to explain the structural features of OI. Moreover, in order to adapt to image with different geometric shape, an improved circular local binary pattern (CLBP) operator is designed by replacing the square neighborhood with the circle neighborhood.

(2) Considering that the user’s behavior characteristics when viewing OIs through HMD can also affect the subjective rating of the OIs, the perception characteristics represented with heat map or significance map are used to further simulate the human visual system (HVS) so as to improve the performance of the proposed SSP-BOIQA method.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III describes the proposed method in detail. Section IV presents the experimental results and analyses. Finally, the conclusion and future works are given in Section V.

II. RELATE WORKS

In this section, OI’s SSP and ERP formats are briefly described and user’s visual behavior characteristics of viewing OIs with HMD are discussed.

In VR system, the omnidirectional image/video with sphere projection format was first mapped to a 2D plane, and then compression and transmission by using the existing 2D image/video coding standards. ERP format is the default mode in OMAF. As shown in Figs. 1(a) and 1(b), ERP maps the spherical longitude lines and the latitude lines to vertical lines and horizontal lines with constant spacing respectively, so that bipolar regions in the sphere are stretched dramatically at the top and bottom of the ERP-image. Although the mapping relationship in ERP is simple and the boundary is continuous, ERP will cause bit waste in encoding and serious stretching distortion at the bipolar regions. All of these results
In a gap between the omnidirectional image/video in ERP format and the practical sphere projection format viewed through head mounted display (HMD).

SSP [20] develops an approximate equal-area mapping, and it is of compression-friendly shapes with less wasteful pixels. In Fig. 1(c), SSP segments the sphere into three tiles: the north polar, equatorial and south polar regions. The north and south poles are mapped into two circles, and the projection of the equatorial segment is the same as the ERP format. The stretching distortions in the bipolar regions in ERP format, as shown in Fig. 2(a), may lead to errors in some BOIQA methods. The SSP format not only retains the visual and continuous boundary at the equatorial region of the ERP format, but also improves the serious stretching distortion of the bipolar regions, which is more in line with the real feeling of user through HMD, as shown in Fig. 2(b).

In addition, users can select the viewport in HMD to focus on the attractive contents of OIs through their HM, and the EM determines which region will appear in the viewport. Therefore, it is helpful to evaluate the objective quality of OIs by analyzing the viewing behavior. Recently, a number of data sets including HM and EM data for omnidirectional image/video visual attention models have emerged to track and analyze user’s viewing behavior. Abreu et al. [21] established one of the earliest attention dataset for OIs, including the HM data of 32 subjects. Rai et al. [22] provided a visual attention dataset (VAD) containing 60 OIs. Except for visual attention models based on subjective experiments, some saliency prediction methods for omnidirectional image/video have appeared. Zhang and Chen [23] presented a novel two-stream neural network for video saliency prediction, which can learn saliency related spatiotemporal features from human fixations.

In summary, in order to establish effective objective BOIQA metrics, it is necessary to consider the unique perception features of OIs. First of all, even though ERP format is compatible with the existing image/video coding standards, it will result in bit waste in encoding and serious stretching distortion in bipolar regions, the latter will make impact on the accuracy of objective BOIQA with the ERP format. Moreover, the process of viewing OIs through HMD is interactive, and the degree of user’s interest in the content of the OI is inconsistent, which provides a way to evaluate OI quality based on user behavior. Therefore, in this paper, we will develop a SSP-BOIQA method to overcome the problems existing in the bipolar regions of ERP format. At the same time, we also consider the characteristics of user’s behavior, and use the subjective heat map or objective visual attention as a weighting factor to further optimize the quality model.

III. PROPOSED SSP-BOIQA METHOD

From the perspective of OMAF compatibility, this paper proposes a SSP-BOIQA method, which mainly focuses on OI quality prediction in both of the bipolar regions and the equatorial region. The framework of the proposed method is shown in Fig. 3. Firstly, the ERP-image is converted into the SSP-image, so as to solve the stretching distortion problem of the bipolar regions in ERP format. Secondly, in the bipolar regions, considering that the regions are circular in the SSP-image, a local/global perceptual feature extraction scheme with a fan-shaped window is proposed for estimating the distortion in the OIs’ bipolar regions. Thirdly, in the equatorial region, the perceptual features are extracted with heat map as weighting factor to reflect users’ visual behavior. Finally, the extracted feature set is used as the input of random forest to simulate the nonlinear relationship between the feature space and the human opinion score.

For the sake of convenience, let $I_{ERP}$ denote the distorted OI with the ERP format, $I_{SSP}$ be the SSP-image converted from $I_{ERP}$, $I_{SSP} = \{I_N, I_S, I_E\}$, where $\{I_N, I_S\}$ denote the north and south poles. $I_E$ represents the equatorial region of distorted OI. Let $H_{ERP}$ or $H_{SSP}$ denote the heat maps with ERP or SSP format respectively, and $H_{SSP} = \{H_N, H_S, H_E\}$. We extract local/global features from $\{I_N, I_S\}$, called as $\{F_{N1}, F_{S1}, F_{N1}, F_{S1}, F_{N2}, F_{S2}\}$. Similarly, for $I_E$, the structure and contrast features are extracted, called as $\{F_{Estr}, F_{Econ}\}$. The method described in this section takes OI2 in MVAQD database [30] as an example.

A. THE BIPOLAR REGION FEATURES OF SSP-IMAGE

The ERP images have severe stretching distortion in the bipolar regions. To overcome this shortcoming of the ERP format, the SSP format projects the bipolar region into two
up and down circles, which makes it more able to show the original appearance of objects than ERP format. However, there also exists a defect, that is, features in the circular regions are difficult to be extracted by using rectangular window. Therefore, a new feature extraction method based on circular region is proposed. Here, the same features are extracted from the bipolar regions, and the proposed method is described in detail by taking the north polar region, \( I_N \), as an example.

Detail features are important to the visual perception, especially in the flat areas of the bipolar regions where the loss of the details is more likely to be magnified and found. In addition, the block effect in the plane domain produced by compression warps due to the conversion, changes to radial blocking and radial banding distortion. As an example, Fig. 4 shows the radial blocking and radial banding distortions. It can be found that when viewing the bipolar region of the distorted OIs through HMD, the radial blocking and radial banding distortions are similar to those in the SSP-image, which means that SSP-image is more suitable for quality assessment of OI compared with ERP-image.

Considering that the traditional rectangular window does not fit the shape of bipolar regions, and cannot detect the edge of the circular area well to represent this unique distortion type, a fan-shaped window, as shown in Fig. 5, is proposed to replace the traditional rectangular window. The fan-shaped window scheme can effectively be used to describe the bipolar regions’ distortion. Let \( W_{fan} \) denote a fan-shaped window, and \((i,j) \in W_{fan}\), \( \theta \) be the constant central angle of \( W_{fan} \), while \( r \) be the variable ring width of the \( W_{fan} \) to ensure that the area covered by each fan-shaped window is almost equal. The north polar region, \( I_N \), is divided into non-overlapping image blocks with sliding \( W_{fan} \). Then, the mean standard deviation, \( \sigma \), of all image blocks of \( I_N \) is calculated as

\[
\sigma = \frac{1}{N} \sum_{N} \left( \frac{1}{n} \sum_{(i,j) \in W_{fan}} (I_N(i,j) - I_{W_{fan}})^2 \right)^{1/2}
\]

where \( N \) is the number of image blocks, \( n \) is the number of pixels in the corresponding \( W_{fan} \), and \( I_{N,m} \) is the mean of the pixel value of all pixels in the \( W_{fan} \). Then, \( \sigma \) is used as the local detail feature of the bipolar regions of the OI.

Cortical neurons are highly sensitive to anisotropic information in images, and image distortion processes can modify and create unnatural local anisotropies [24]. Hence, it is of great significance to study the direction information of the image to improve BOIQA method, especially in the circular region of image. In this work, relative gradient orientation (RGO) is used to explain the structural features of images. Fig. 6 shows the RGO maps, from which the difference between the original image and its distorted images can be clearly distinguished.
By calculating the gradient of each pixel in horizontal and vertical directions, the feature of the gradient direction in $I_N$ is extracted. Let $\{\mathcal{G}_N(i,j)\}$ denote the gradient direction map of $I_N$, then, $\mathcal{G}_N(i,j)$ can be computed by

$$\mathcal{G}_N(i,j) = \arctan \left( \frac{I_N(i,j) * P_h}{I_N(i,j) * P_v} \right)$$

(2)

where $*$ denotes convolution operator, and $P_h$ and $P_v$ represent gradient operators of horizontal and vertical directions. For example, Prewitt operators or other gradient operators can be used to compute $\mathcal{G}_N(i,j)$. Orientation is relative, it may be measured absolutely, against the frame of reference of the image coordinate system, or it can be measured in a relative manner against the background of local orientations. RGO features can capture departures from the natural distribution of local orientations caused by local degradations of image structure. Therefore, RGO features are used to describe the degradation of the distorted image structure, and the estimated RGO, $\{\mathcal{G}_{N,RGO}(i,j)\}$, can be computed by

$$\mathcal{G}_{N,RGO}(i,j) = \mathcal{G}_N(i,j) - \mathcal{G}_{N,m}(i,j)$$

(3)

where $\mathcal{G}_{N,m}(i,j)$ is the local mean orientation of north polar region in the SSP-image, which is calculated by

$$\mathcal{G}_{N,m}(i,j) = \arctan \left( \frac{G_{N,vm}(i,j)}{G_{N,hm}(i,j)} \right)$$

(4)

where $G_{N,vm}(i,j)$ and $G_{N,hm}(i,j)$ represent the mean of gradient values of the horizontal and vertical directions in a window centered at position $(i,j)$, respectively.

Fig. 7 shows RGO distributions of an OI with regard to five kinds of distortions. It is shown that distortions cause the RGO distribution deviating from that of the original OI, and different distortion types lead to different changes. For example, HEVC, JPEG and JP2K distortions generate block effect due to coding compression, which adds pseudo-edge structure to make the peak higher, while white noise (WN) introduces noise to make the distribution flatter. The distribution conforms to the generalized Gauss distribution (GGD) model with zero mean value, which can be fitted as

$$G(x; \mu, \alpha, \beta) = \frac{\alpha}{2\beta \Gamma(1/\alpha)} \exp \left( -\frac{|x|^{\alpha}}{\beta} \right)$$

(5)

where $\Gamma(\cdot)$ is the gamma function, and expressed as

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0$$

(6)

where $\mu$ is the mean brightness value of the image block, $\alpha$ is the shape parameter, and $\beta$ is the scale parameter. Likewise, $\mu, \alpha$ and $\beta$, which can be estimated with the moment matching-based method, are added to the quality-aware feature set as the global structure features in this work.

Usually, the scene of the OI’s bipolar regions are relatively simple because of the shooting conditions, but the texture information will be changed due to distortions, so the texture information is also worth to be considered. Local binary pattern (LBP) operator [25] can be used to describe the relationship between the center pixel and its surrounding neighbors by computing gray-level differences. The traditional LBP operator is calculated in a square neighborhood, which cannot obviously meet texture features of images with different geometric shapes, especially, in the case of the SSP’s bipolar regions. In order to adapt to bipolar regions of OI with SSP format and extract effective texture information at the edges of these regions, the traditional LBP operator is improved by replacing the square neighborhood with the circular neighborhood.

Specifically, the circular LBP (CLBP) of each pixel in the circular image can be computed by

$$\text{CLBP}_{P,R}^{riu2} = \begin{cases} \sum_{i=0}^{P-1} s \left( \frac{I_N - I_{Nc}}{\bar{R}} \right), & \text{if } u(\text{CLBP}_{P,R}) \leq 2 \\ P + 1, & \text{otherwise} \end{cases}$$

(7)

where $u$ is the uniform measure, and the superscript $riu2$ denotes the rotation invariant “uniform” patterns with $u$ value less than 2, $P$ is the number of pixels in the neighborhood, and $R$ is the neighborhood radius; $I_{Nc}$ denotes the pixel value of the center pixel, and $I_N$ is the pixel value of the $i$th pixel in the neighborhood. $s(\cdot)$ is a threshold function and defined by

$$s(c_1 - c_r) = \begin{cases} 1, & \text{if } c_1 - c_r \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

(8)

CLBP describes the relationship between the pixels of the circular image in the bipolar regions, which can effectively
reflect the complex degradation caused by various distortions. In the bipolar regions, the area of interest to human eyes will make greater contributions to the subsequent extraction of perceptual texture features. Therefore, the histogram of CLBP map is weighted by using heat map as the weight, and the CLBP histogram weighted with heat map is obtained by

\[
h_{\text{CLBP}}(k) = \sum_{(i,j) \in I_N} H_N(i, j)f(\text{CLBP}_{P,K}(i, j), k)
\]

where \(k \in [0, K]\) is the possible CLBP patterns, and \(H_N(i, j)\) denotes the heat map of the north polar region.

Fig. 8 shows the CLBP map of the original OI, its distorted OIs, and the weighted CLBP histogram respectively. It can be seen from the figure that different types of distortion result in different CLBP maps, and the difference in the weighted CLBP histogram is obvious. CLBP map has \(P + 2\) different modes and different types of distortions result in different distributions of CLBP modes. Among them, the mode 0 stands for bright spot, the modes 1-7 represent the edge areas, the mode 8 denotes dark or flat area, and the mode 9 is the non-uniform patterns that are less encountered in natural OIs. For example, HEVC and JPEG compression increase the mode 8 in the map, because these two compression distortions can cause the block effect, resulting in a large number of flat image blocks. WN reduces the modes 2-6 in the map and increases the mode 9, because it reduces the edge information but increases the non-uniform texture which is not common in natural images.

According to the above analysis, the extracted histogram of the CLBP map can effectively summarize the effect of image distortion on texture information, so it is taken as the global texture feature of the bipolar regions.

Finally, the perceptive feature vector extracted from the \(I_N\) has a total of 14 dimensions, denoted as \(F_N = \{F_{NL}, F_{Ng1}, F_{Ng2}\}\), where \(F_{Ng}\) is the 1-dimensional local detailed feature extraction from the \(I_N\), \(F_{Ng1}\) is the 3-dimensional global structure feature vector extracted from the \(I_N\), \(F_{Ng2}\) is 10-dimensional global CLBP feature vector extracted from the \(I_N\). Similarly, 14-dimensional feature vector are also extracted from the \(I_S\), denoted as \(F_S = \{F_{Sl}, F_{Sg1}, F_{Sg2}\}\).

**B. THE EQUATORIAL REGION FEATURES OF SSP-IMAGE**

The equatorial region in the SSP-image is convolved with a 2-D log-Gabor filter denoted by

\[
G_{2D}(\omega, \theta) = \exp \left( -\frac{1}{2\sigma^2} \left( \log \frac{\omega}{\omega_0} \right)^2 \right) \cdot \exp \left( -\frac{(\theta - \theta_k)^2}{2\sigma_\theta^2} \right)
\]

where \(\theta_k = k\pi/K, k = \{0, 1, \ldots, K - 1\}\) is the orientation angle of the filter, \(K\) is the number of orientations, \(\sigma_\theta\) determines the filter’s angular bandwidth. \(\omega_0\) is the central frequency of the filter, and \(\sigma_r\) controls the filter bandwidth.

By modulating \(\omega_0\) and \(\theta_k\) and convolving \(G_{2D}\) with the equatorial region \(I_E\), we get a set of responses of the pixels at the positions of \{(i, j)\} as \([e_n,\theta_k(i, j), o_n,\theta_k(i, j)]\). The local amplitude on scale \(n\) and orientation \(\theta_k\) is

\[
A_n,\theta_k(i, j) = \sqrt{e_n,\theta_k(i, j)^2 + o_n,\theta_k(i, j)^2}
\]

The local energy along orientation \(\theta_k\) is

\[
E_{\theta_k}(i, j) = \sqrt{F_{\theta_k}(i, j)^2 + H_{\theta_k}(i, j)^2}
\]

where \(F_{\theta_k}(i, j) = \sum_n e_n,\theta_k(i, j)\) and \(H_{\theta_k}(i, j) = \sum_n o_n,\theta_k(i, j)\). The 2-D PC at \((i, j)\) is computed as follows

\[
PC_{2D}(i, j) = \frac{\sum_k E_{\theta_k}(i, j)}{\varepsilon + \sum_n \sum_k A_n,\theta_k(i, j)}
\]

where \(\varepsilon\) is a very small normal number, in case the denominator is 0.

Multi-scale representation can better reflect the texture structure characteristics of OIs, so the PC map is sampled four times to obtain five scale feature maps. In this section, the information entropy from five scale feature graphs is taken as the structure feature of the equatorial region in SSP-image.

Compared with the traditional edge detection method based on gradient function, PC can detect the edge of the image well and detect features within a wide range. However, it is contrast invariant, but the local contrast of an image affects the image quality perceived by HVS. When viewing

![FIGURE 8. CLBP map and the corresponding histogram of north polar region under different distortions.](image-url)
the OI through HMD, the user can only keenly observe one part of the OI at a certain moment. Therefore, the heat map is taken as the weight to guide the selection of image patches, and the histogram of the selected image patches is used as contrast feature to compensate phase information, so as to improve the perceptual consistency between the objective assessment results and human visual system.

By using the sliding window, the equatorial region of the OI is divided into several non-overlapping image patches. Let \( W_d \) denote the 64 × 64 sliding window. The heat map is also divided into several corresponding image patches. The weight of an image patch is calculated as \( w(l) = \sum H_E(l) \), where \( H_E \) represents the heat map of the equatorial region. A group of image patches with greater weight are selected, and the histogram features of which are calculated. Then we construct the following histogram set, in which each column is the histogram of one selected image patch, denoted by

\[
H = \begin{bmatrix}
    h_{0,0} & \cdots & h_{0,L-1} \\
    \vdots & \ddots & \vdots \\
    h_{255,0} & \cdots & h_{255,L-1}
\end{bmatrix}
\]

(15)

where \( L \) is the number of image patches.

Principal component analysis (PCA) is used to construct the zero-mean matrix, so that the image patch has zero-mean brightness and its contrast feature is enhanced.

\[
H_0(l) = H(l) - \overline{H}
\]

(16)

where \( H(l) \) represents the histogram of the \( l \)-th image patch, and also the \( l \)-th column vector located in \( H \). \( \overline{H} \) is the average brightness matrix of \( H \).

Then, the covariance matrix, \( Cov \), of \( H \) is calculated by

\[
Cov = \frac{1}{L} \sum_{l=0}^{L-1} H_0(l)H_0(l)^T
\]

(18)

Singular value decomposition (SVD) [27] is used to calculate the eigenvalues of the covariance matrix. The eigenvalues are arranged in order from large to small

\[
\begin{bmatrix}
    \lambda_0 & \cdots & 0 \\
    \vdots & \ddots & \vdots \\
    0 & \cdots & \lambda_{255}
\end{bmatrix}
\]

(19)

The first 20 eigenvalues are selected as the local contrast features of the equatorial region in the SSP-image.

Finally, the perceptive feature vector extracted from the \( I_E \) has a total of 25 dimensions, denoted as \( F_E = \{ F_{Estr}, F_{Econ} \} \), where \( F_{Estr} \) is the 5-dimensional structure feature vector extracted from the \( I_E \), and \( F_{Econ} \) is the extracted 20-dimensional local contrast feature vector of \( I_E \).

C. QUALITY PREDICTION OF OI

The 53-dimensional feature set \( F_f, F_f = \{ F_N, F_S, F_E \} \), extracted from the equatorial region and the bipolar regions are used as the input of the random forest to simulate the nonlinear relationship between the feature space and the human opinion score. Specifically, the feature space is mapped to predict the objective quality \( Q_f \) of OIs by a regression function \( f_m(\cdot) \), which can be expressed as

\[
Q_f = f_m(F_f)
\]

(20)

where \( f_m(\cdot) \) is achieved by machine learning, and \( F_f \) is the extracted feature vector. Since random forest has great prediction accuracy and is not easy to appear the over-fitting phenomenon when performing regression and classification tasks [28], the random forest is utilized to learn the mapping function \( f_m(\cdot) \) in this work.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENTS

To verify the effectiveness of the proposed method, the method is compared with several classical NR-IQA and FR-IQA methods on the following two OI databases.

(1) CVIQD2018 database [29]: The database contains a total of 544 images, including 16 sets of original scenes and distorted images generated by HEVC, AVC and JPEG compression. The MOS value ranges from 0 to 100, and the higher the score, the higher the image quality. It should be noted that there is no heat map in this database, so Zhang’s omnidirectional significance map extraction method [23] is utilized to generate significance map from the reference image of this database instead of the heat map.

(2) MVAQD database [30]: We have also designed a diverse subjective OI database to further verify the effectiveness of the proposed SSP-BOIQA method. It contains

TABLE 1. Distortion level.

| Distortion level | 1 | 2 | 3 | 4 |
|-----------------|---|---|---|---|
| Gaussian Blur(σ) | 0.5 | 1 | 2 | 5 |
| HEVC intra coding(QP) | 27 | 32 | 37 | 42 |
| JPEG compression(Q) | 70 | 50 | 30 | 10 |
| JP2K compression(8bpp) | 0.6 | 0.3 | 0.1 | 0.05 |
| White Noise(σ²) | 0.001 | 0.005 | 0.01 | 0.02 |

FIGURE 9. Original images in MVAQD database. (a) OI11. (b) OI12. (c) OI13. (d) OI14. (e) OI15. (f) OI16. (g) OI17. (h) OI18. (i) OI19. (j) OI10. (k) OI11. (l) OI12. (m) OI13. (n) OI14. (o) OI15.
15 high-quality uncompressed OIs and their heat maps, all from the 360-degree image database (VAD) provided by Rai et al. [22], as shown in Fig. 9.

For the MVAQD database, each original image is distorted with five distortion types at four distortion levels respectively, as shown in Table 1, and therefore the corresponding 20 distorted images are obtained with respect to each of the original image.

In the subjective experiments, 26 subjects voted on the quality of 315 OIs in MV AQD database. The voting is performed by using five-grade quality scale with the following levels: “5-Excellent”, “4-Good”, “3-Fair”, “2-Poor”, and “1-Bad”. After the subjective evaluation process, we have strictly followed the screening criteria described in [31] to implement the subject reliability assessment. Finally, the MOS of the remaining 20 subjects are utilized in the experiments. Fig. 10 describes the MOS values of 15 scenes with different degrees of distortion. With the increase of distortion degree, the subjective quality decreases. Fig. 11 further gives the exemplary images with specific distortions from MVAQD database and their subjective scores. It is seen that the better the image quality, the higher the MOS value, which illustrates the validity of subjective ratings of MVAQD database.

Four commonly used performance criteria, suggested by video quality experts group (VQEG), are employed to evaluate and compare the proposed method with existing IQA methods, namely Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), Kendall’s rank correlation coefficient (KRCC) and Root mean squared error (RMSE) [32]. PLCC, SRCC, KRCC and RMSE measure the accuracy, monotonicity, relevance and error in the process of prediction, respectively. The closer the absolute values of PLCC, SRCC and KRCC are to 1, and the closer RMSE is to 0, the better the performance of an objective quality assessment method is.

B. OVERALL PERFORMANCE COMPARISON

To investigate the effectiveness of the proposed SSP-BOIQA method, it is compared with some popular FR-IQA and BIQA methods on the above two omnidirectional databases. The selected FR-IQA methods are PSNR, SSIM [7], VIF [8], CPP-PSNR [11], S-PSNR [9] and WS-SSIM [12], in which the last three are objective evaluation metrics designed for omnidirectional visual content. The BIQA methods selected include BRISQUE [15], OG [16], NIQE [17], dip_IQ [18] and BPRI [19]. Among them, BRISQUE, OG and SSP-BOIQA proposed in this paper are opinion-aware BIQA methods that requires a large number of subjective opinions for training, while the other three objective evaluation metrics are opinion-unaware BIQA methods. For opinion-aware BIQA methods, 80% of the reference images and their distorted images in the omnidirectional data set were randomly selected as the training set, while the remaining 20% were as the test set, and the distorted images corresponding to the same reference image are allocated to the same set to ensure complete independence of the training and test data. To avoid the contingency of the calculated correlation coefficients, the above training-test process was repeated 1000 times, and the median values of PLCC, SRCC, KRCC and RMSE were taken as the final performance indicators. For FR-IQA and opinion-unaware BIQA methods, we report the mean performance of them on the test set of the opinion-aware BIQA methods for a fair comparison with methods that need training.

Table 2 gives the experimental results on the CVIQD2018 and MVAQD databases, with the best performance indicators for FR-IQA and BIQA highlighted in bold.

From Table 2, the following observations can be obtained.

1) All performance indexes of CVIQD2018 database are generally better than that of MVAQD database. Compared with CVIQD2018 database, which contains only three encoding compression distortions, MVAQD database additionally includes blur, JP2K and WN. Thus, the MVQAD database has better distortion diversity, which makes the distortion difference of this database larger and more complex, and therefore
makes more challenge for objective quality assessment. Some of these methods have significant performance differences between these two databases, which shows their limitation to different types of distortions. For example, BPRI performed poorly on MV AQD database, but well in CVIQD2018 database.

(2) The performance of FR-IQA methods was generally higher than that of the classical BIQA methods, and opinion-aware BIQA methods performed better than opinion-unaware BIQA methods. The FR-IQA methods take both the reference image and the distorted image as the input, which can offset the influence of the geometric distortion of the ERP-image to some extent. Additionally, most opinion-aware BIQA methods that need training are subject to feature regression through random forest, and this has advantages over simple linear combinations because the relationship between features and subjective evaluation is extremely complex.

(3) The performance of FR-OIQA methods is improved slightly compared with the traditional FR-IQA methods. It indicates that using the traditional FR-IQA methods to calculate the quality of an OI in ERP format cannot truly reflect the distortion on the sphere, but calculating the image quality on the plane that closer to the sphere can better eliminate the influence of projection transformation. However, the existing FR-OIQA methods do not take into account user behavior characteristics, which results in an insignificant improvement over the FR-IQA methods. Moreover, SSIM predicts the objective quality score of the distorted image by calculating the structural similarity between the reference image and the distorted image. For CVIQD2018 database which only contains three encoding compression distortions, SSIM is good at detecting the changes of image structure information resulted from the compression, so SSIM performs well on CVIQD2018 database. But for more complex MVAQD database, SSIM is not so outstanding.

By contrast, the proposed SSP-BOIQA method is outstanding for both of the two databases because of its geometric compensation for the ERP-image and the combination of user’s behavior characteristics in HMD.

C. PERFORMANCE COMPARISON OF DIFFERENT FEATURES

The proposed method comprehensively considers geometric compensation and user perception characteristics, divides the OI with SSP format into the equatorial region and the bipolar regions, and extracts different perception features from these two kinds of regions. To compare the contribution of different features extracted from different regions to the
final performance of quality assessment, experiments were conducted on the above two databases for different regional features, as listed in Table 3, where “NQ” and “SQ” respectively represent the extracted perception features of the north and south polar regions, while “EQ” represents the extracted perception features of the equatorial region. For the north and south polar regions where the same perceptual features are extracted, the performance indexes of the south polar region are better than those in the north polar region. This is because the north polar regions of OI in the two databases are mostly sky, indoor ceiling, and so on, which are flatter and simple than the south polar region. This phenomenon is more obvious in CVIQD2018 database, so the performance indexes of SQ in this database are much better than that of NQ. In addition, EQ makes a great contribution to the quality assessment, because equatorial region has rich structure and complex texture, which is the most interesting area for human eyes. Finally, if all the features are combined the performance of quality assessment can be improved significantly, as shown by “All” in Table 3.

Considering that the geometric shapes of different regions in SSP-image are different, and human eyes have different perception degrees for distortions in different regions, we extract different perceptual features from different regions in SSP-image. In order to analyze the performance of different types of features in two databases, as well as the complementary effect between some features, the experiments shown in Table 4 are carried out. In the proposed SSP-BOIQA method, \( \{F_{N1}, F_{Ng1}, F_{Ng2}, F_{S1}, F_{Sg1}, F_{Sg2}\} \) constitute the feature set of the bipolar regions, where \( F_{N1} \) and \( F_{S1} \) are the local detail features, \( F_{Ng1} \) and \( F_{Sg1} \) are the global structure features, while \( F_{Ng2} \) and \( F_{Sg2} \) are the global CLBP features extracted from the bipolar regions. \( \{F_{Estr}, F_{Econ}\} \) constitute the feature set of the equatorial region, where \( F_{Estr} \) is the PC structure feature, and \( F_{Econ} \) is the local contrast feature extracted from the equatorial region.

From Table 4, the following observations can be derived. Firstly, among all the features, the global CLBP features \( \{F_{Ng2}, F_{Sg2}\} \) produce the best performance on the two OI databases. This is because the scene contents of the bipolar region of OI is relatively simple, which makes the texture information change caused by distortions more obvious. Secondly, the performance of feature set of bipolar regions is improved after feature fusion. In the bipolar regions, local features and global features complement each other, while structural features and texture features complement each other. Thirdly, the feature set of equatorial region performs better after the feature fusion. Phase consistency is contrast

### Table 3. Performance of features of different regions.

| Regions | CVIQD2018 database | MVAQD database |
|---------|--------------------|----------------|
|         | PLCC | SRCC | KRCC | RMSE | PLCC | SRCC | KRCC | RMSE |
| NQ      | 0.7102 | 0.6466 | 0.4826 | 9.9659 | 0.6461 | 0.5849 | 0.4182 | 0.8241 |
| SQ      | 0.8422 | 0.8161 | 0.6271 | 7.3824 | 0.6776 | 0.6228 | 0.4518 | 0.7852 |
| EQ      | 0.8742 | 0.8465 | 0.6641 | 6.8838 | 0.7558 | 0.6809 | 0.5094 | 0.7045 |
| All     | 0.9077 | 0.8614 | 0.6954 | 6.1178 | 0.8406 | 0.7838 | 0.6036 | 0.5864 |
TABLE 4. Performance of different types of features.

| Types | Metrics | CVIQD2018 database | PLCC | SRCC | KRCC | RMSE |
|-------|---------|---------------------|------|------|------|------|
| N/SQ  | \(F_{SN}, F_E\) | 0.3700 | 0.3162 | 0.2779 | 15.6295 |
|       | \(F_{SN}, F_{G1}\) | 0.5250 | 0.5584 | 0.4654 | 12.0868 |
|       | \(F_{SN}, F_{G2}\) | 0.8552 | 0.8522 | 0.6506 | 7.2352 |
|       | \(F_E, F_S\) | 0.8662 | 0.8454 | 0.6641 | 7.1069 |
| EQ    | \(F_{SN}\) | 0.7834 | 0.6589 | 0.4814 | 8.7497 |
|       | \(F_{G1}\) | 0.6638 | 0.6244 | 0.4469 | 10.4062 |
|       | \(F_{G2}\) | 0.8742 | 0.8465 | 0.6641 | 6.8838 |
| All   | 0.9077 | 0.8614 | 0.6954 | 6.1178 |

TABLE 5. Effects of heat map or significance map.

| Metrics | Heat Map | CVIQD2018 database | PLCC | SRCC | KRCC | RMSE |
|---------|----------|---------------------|------|------|------|------|
| N/SQ    | \(F_{SN}, F_{G2}\) | × | 0.8507 | 0.8319 | 0.6522 | 7.2958 |
|         | \(F_{SN}, F_{G2}\) | ✓ | 0.8552 | 0.8322 | 0.6506 | 7.2352 |
| EQ      | \(F_{SN}\) | × | 0.6553 | 0.6216 | 0.4378 | 10.5196 |
|         | \(F_{SN}\) | ✓ | 0.6638 | 0.6244 | 0.4469 | 10.4062 |
| Types   | Significance Map | CVIQD2018 database | PLCC | SRCC | KRCC | RMSE |
| N/SQ    | \(F_{SN}, F_{G2}\) | × | 0.7358 | 0.6750 | 0.4972 | 0.7234 |
|         | \(F_{SN}, F_{G2}\) | ✓ | 0.7362 | 0.6721 | 0.5023 | 0.7220 |
| EQ      | \(F_{SN}\) | × | 0.5816 | 0.4792 | 0.3401 | 0.8537 |
|         | \(F_{SN}\) | ✓ | 0.5937 | 0.4903 | 0.3449 | 0.8492 |

D. EFFECT OF HEAT MAP

Users can select the viewport of HMD and focus on the attractive contents of OIs through head movement and eye movement. Therefore, the heat map or significance map is used as the weighting factor in the selection of image patches and the global CLBP feature extraction. Two comparison experiments are implemented, and the results are given in Table 5. In the first group, the directly counted histogram of CLBP is compared with the histogram weighted with heat map or significance map. In the second group, the scheme which randomly selects image patches in the equatorial region is compared with the scheme which selects image patches with heat map or significance map. From Table 5, it is seen that the use of heat map or significance map helps to improve the performance of the proposed SSP-BOIQA method.

E. PARAMETER SENSITIVITY

In the proposed SSP-BOIQA method, the selection of some specific parameters is involved, including the radius \(R\) and the sampling points \(P\) in the improved CLBP, the number of image blocks \(N\) divided by the sliding fan-shaped window in the bipolar regions, the dimension \(C\) of the local contrast feature vector and the scale dimension \(S\) of the structure features in the equatorial region. The selection of these parameters will have an impact on the proposed BOIQA method, so we design comparative experiments to select the best parameters.

Firstly, in order to discuss the effects of the radius \(R\) and the sampling points \(P\) of the improved CLBP for the bipolar regions, we changed the size of \(R\) and \(P\) for experimental comparison. Three groups parameters are chosen to measure PLCC and SRCC for comparison experiment, respectively: \(R = 1, P = 8; R = 2, P = 8; R = 2, P = 16\), the result is shown in Fig. 13(a). It can be found from the figure that there is no significant difference in performance as \(R\) and \(P\) changing. When \(R = 1, P = 8\), the overall performance index is slightly higher. After considering the computational complexity, \(R\) is set to 1 and \(P\) is set to 8 reasonably in the proposed SSP-BOIQA method.

Secondly, the number of image blocks \(N\) determines the degree of fineness to extract the local detail perception.
features in the bipolar regions. If \( N \) is too large, the image block will contain too little content to accurately reflect the local details. However, if \( N \) is too small, the image block size will be too large to reflect the advantages of the improved fan-shaped window. Fig. 13(b) shows the performance comparison with respect to different \( N \). \( r \) and \( \theta \) in Fig. 5 can be changed to control the size of the fan-shaped window, and divide the bipolar regions into image blocks. The \( N_1 \) and \( N_2 \) of Fig. 13(b) are the number of rings and the number of image blocks of a ring, respectively, and three groups of parameters are compared in terms of the corresponding PLCC and SRCC of the quality assessment: \( 8 \times 18 \), \( 8 \times 36 \), and \( 16 \times 36 \). As can be seen from Fig. 13(b), changing \( N \) has a little effect on overall performance. When \( N = 16 \times 36 \), PLCC and SRCC indexes are slightly higher than others. Therefore, \( N \) is set to \( 16 \times 36 \) in this work.

Finally, in order to discuss the dimension \( C \) of the local contrast feature, vector and the scale dimension \( S \) of the complementary multi-scale structure features in the equatorial region, the performance of features is measured on the two databases when \( C = \{ 10, 15, 20, 25 \} \) and \( S = \{ 1, 3, 5 \} \), as shown in Table 6. It is seen that when \( C = 20 \) and \( S = 5 \), the features of the equatorial region achieve good performance for both databases, and the size of \( C \) will not significantly affect the performance of the features.

F. DISCUSSION

In this paper, the SSP-BOIQA method is proposed to compensate the OI geometrically, taking into account the fact that the bipolar regions of the ERP image have obvious stretching distortion and waste a lot of bits for encoding. Moreover, considering the user’s behavior characteristics with HMD, the feature extraction is improved and the perception factor is added. Although the proposed method achieves good results on both MVAQD and CVIQD2018 databases, it may not provide satisfactory results in some challenging situations.

In this work, most of the features selected from the equatorial region and the bipolar regions are based on the improvement of natural statistical features. How to dig deeper features related to human visual perception is important for blind quality assessment of omnidirectional image. Omnidirectional video system is a kind of close eye viewing, its visual perception characteristics is more complex and different from the common video viewing mode. Therefore, how to obtain effective user behavior data through subjective experiments and use it for feature extraction and pooling is worth being studied to further improve the performance of quality assessment in the future.

With the rapid development of VR technology, omnidirectional video (OV) has been widely used instead of OI, because it provides users with more immersive experience. Therefore, the objective quality assessment of OV has become a new challenge. Different from OI, OV focuses on the fluency of presentation and playback. Therefore, how to describe the feature of temporal correlation is a new problem.

V. CONCLUSION

In this paper, a segmented spherical projection based blind omnidirectional image quality assessment (SSP-BOIQA) method is proposed for virtual reality systems. Specifically, the omnidirectional image (OI) is first converted to segmented spherical projection (SSP) format to solve the problem of stretching distortion of the bipolar regions in the equirectangular projection (ERP) format. Considering the different geometric shapes of different regions in SSP-image, and the different perception degree of human eyes to distortions in different regions of the SSP-image, different features are extracted from different regions in the SSP-image. For feature extraction of the bipolar regions, the fact that the bipolar regions are circular images are considered. In addition, heat map or significance map is used to improve the performance of the quality assessment. The experimental results on the MVAQD and CVIQD2018 databases show that the proposed SSP-BOIQA method is superior to the existing image quality assessment methods. However, even though the heat map or significance map can simply simulate the user’s behavior of viewing OI through HMD, it is not enough for quality assessment of OI. Therefore, there is still great room for the improvement of the performance of IQA method and how to dig deeper features related to human visual perception is the focus of future work.

REFERENCES

[1] Y. Fan, Y. Jin, Z. Meng, and X. Zeng, “Pixels and panoramas: An enhanced cubic mapping scheme for videoimage-based virtual-reality scenes,” IEEE Consum. Electron. Mag., vol. 8, no. 2, pp. 44–49, Mar. 2019.
[2] Y. Kusuno, J. Takahashi, and Y. Yong, “A method localizing an omnidirectional image in pre-constructed 3D wireframe map,” in Proc. IEEE/SCS Int. Symp. Syst. Integr. (SiS), Paris, France, Jan. 2019, pp. 325–329.
[3] R. G. D. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatza, B. Adsumilli, and P. Frossard, “Visual distortions in 360-degree videos,” IEEE Trans. Circuits Syst. Video Technol., to be published, doi: 10.1109/tcsvt.2019.2927344.
[4] B. Choi and E. Yip, OMAF PACK-VE: Stereoscopic 360 Video Format and Packing, Standard ISO/IEC JTC1/SC29/WG11, MPEG/M39842v3, 2017.
[5] Z. Chen, Y. Li, and Z. Zhang, “Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation,” Signal Process., vol. 146, pp. 66–78, May 2018.
[6] M. Oszust, “Local feature descriptor and derivative filters for blind image quality assessment,” IEEE Signal Process. Lett., vol. 26, no. 2, pp. 322–326, Feb. 2019.
[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.
[8] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.
[9] M. Yu, H. Lakshman, and B. Girod, “A framework to evaluate omnidirectional video coding schemes,” in Proc. IEEE Int. Symp. Mixed Augmented Reality, Fukuoka, Japan, Sep. 2015, pp. 31–36.
[10] Y. Sun, A. Lu, and L. Yu, “Weighted-to-spherically-uniform quality evaluation for panoramic video,” IEEE Signal Process. Lett., vol. 24, no. 9, pp. 1408–1412, Sep. 2017.
[11] V. Zakharchenko, K. P. Choi, and J. H. Park, “Quality metric for spherical panoramic video,” Opt. Photon. Inf. Process. X, vol. 9970, pp. 1–10, Sep. 2016.
[12] Y. Zhou, M. Yu, H. Ma, H. Shao, and G. Jiang, “Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video,” in Proc. 14th IEEE Int. Conf. Signal Process. (ICSP), Beijing, China, Aug. 2018, pp. 54–57.
A. De Abreu, C. Ozcinar, and A. Smolic, “Look around you: Saliency AHG8: Segmented Sphere

C. Zhang, Y. Lu, J. Li, and Z. Wen,

X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, “Blind quality

K. Ma, W. Liu, T. Liu, Z. Wang, and A. C. Bovik, “Blind image quality

L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, “Blind image quality

A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality

K. Zhang and Z. Chen, “Video saliency prediction based on spatial-

T. Maugey, O. Le Meur, and Z. Liu, “Saliency-based navigation in omni-
directional image,” in Proc. IEEE 19th Int. Workshop Multimedia Signal

C. Li, M. Xu, X. Du, and Z. Wang, “Bridge the gap between VQA and

H. G. Kim, H.-T. Lim, and Y. M. Ro, “Deep virtual reality image

K. Ma, W. Liu, T. Liu, Z. Wang, and A. C. Bovik, “Blind image quality

T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and

T. Maugey, O. Le Meur, and Z. Liu, “Saliency-based navigation in omni-
directional image,” in Proc. IEEE 19th Int. Workshop Multimedia Signal

ISO/IECJTC1/SC29/WG11, JVET-E0025, Geneva, Switzerland, 2017.

A. De Abreu, C. Ozcinar, and A. Smolic, “Look around you: Saliency

T. Maugey, O. Le Meur, and Z. Liu, “Saliency-based navigation in omni-
directional image,” in Proc. IEEE 19th Int. Workshop Multimedia Signal

C. Zhang, Y. Lu, J. Li, and Z. Wen, AHG8: Segmented Sphere

D. Kalman, “A singularly valuable decomposition: The SVD of a matrix,”

P. Kovesi, “Phase congruency: A low-level image invariant,” Psychol.

T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and

GANGYI JIANG (Senior Member, IEEE) received the M.S. degree from Hangzhou University, China, in 1992, and the Ph.D. degree from Ajou University, South Korea, in 2000. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. He has authored over 100 technical articles in refereed journals. His research interests mainly include visual communication and image/video processing, 3D video coding, omni-directional video processing, light field imaging, visual perception, and quality assessment.

HAO JIANG received the M.S. degree from Ningbo University, China, in 2008, where he is currently pursuing the Ph.D. degree in signal and information processing. He is also a Researcher with the Ningbo Key Laboratory of Intelligent Household Appliances, Zhejiang Business Technology Institute, China. His research interests mainly include low complexity video processing and omni-sensor quality assessment.

MEI YU received the B.S. and M.S. degrees from the Hangzhou Institute of Electronics Engineering, China, in 1990 and 1993, respectively, and the Ph.D. degree from Ajou University, South Korea, in 2000. She is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. Her research interests mainly include image/video coding, light field imaging, visual perception, and quality assessment.

XUELEI ZHENG received the B.S. degree in communication engineering from Zhejiang University City College, China, in 2018. She is currently pursuing the M.S. degree with the Faculty of Information Science and Engineering, Ningbo University, China. Her research interests mainly include omnidirectional video processing and quality assessment.

[13] 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 Proc. ACM Multimedia Conf. Multimedia Conf. (MM), Oct. 2018, pp. 22–26.

[14] H. G. Kim, H.-T. Lim, and Y. M. Ro, “Deep virtual reality image quality assessment with human perception guider for omnidirectional image,” IEEE Trans. Circuits Syst. Video Technol., to be published, doi: 10.1109/tcsvt.2019.2898732.

[15] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[16] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, “Blind image quality assessment by relative gradient statistics and adaboosting neural network,” Signal Process., Image Commun., vol. 40, pp. 1–15, Jan. 2016.

[17] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘completely blind’ image quality analyzer,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[18] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, “DipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs,” IEEE Trans. Image Process., vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[19] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, “Blind quality assessment based on pseudo-reference image,” IEEE Trans. Multimedia, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.

[20] C. Zhang, Y. Lu, J. Li, and Z. Wen, AHG8: Segmented Sphere

[21] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘completely blind’ image quality analyzer,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[22] Y. Rai, J. Gutiérrez, and P. Le Callet, “A dataset of head and eye movements for 360 degree images,” in Proc. 8th ACM Multimedia Syst. Conf. (MMSys), 2017, pp. 205–210.

[23] K. Zhang and Z. Chen, “Video saliency prediction based on spatial-temporal two-stream network,” IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 12, pp. 3544–3557, Dec. 2019.

[24] T. Maugéy, O. Le Meur, and Z. Liu, “Saliency-based navigation in omni-directional image,” in Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP), Luton, U.K., Oct. 2017, pp. 1–6.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.

[26] P. Kovesi, “Phase congruency: A low-level image invariant,” Psychol. Res., vol. 64, no. 2, pp. 136–148, Dec. 2000.

[27] D. Kalman, “A singularly valuable decomposition: The SVD of a matrix,” College Math. J., vol. 27, no. 1, pp. 2–23, Jan. 1996.

[28] S.-C. Pei and L.-H. Chen, “Image quality assessment using human visual dog model fused with random forest,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3282–3292, Nov. 2015.

[29] 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 Proc. IEEE 20th Int. Workshop Multimedia Signal Process. (MMSP), Vancouver, BC, Canada, Aug. 2018, pp. 1–6.

[30] MVAQD. (Oct. 2019). Multi-Distortions Visual Attention Quality Dataset. [Online]. Available: https://github.com/Jianghao2019/MVAQD

[31] Methodology for the Subjective Assessment of the Quality of Television Pictures, Standard BT.500-13 (01/2012), ITU-R Recommendation, International Telecommunication Union, 2012.

[32] VQEG. (Jun. 2000). Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. [Online]. Available: http://www.vqeg.org/