

# Perceptual Quality Assessment of Omnidirectional Images: Subjective Experiment and Objective Model Evaluation



DUAN Huiyu, ZHAI Guangtao, MIN Xiongkuo, ZHU Yucheng, FANG Yi, and YANG Xiaokang

(Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** Virtual reality (VR) environment can provide immersive experience to viewers. Under the VR environment, providing a good quality of experience is extremely important. Therefore, in this paper, we present an image quality assessment (IQA) study on omnidirectional images. We first build an omnidirectional IQA (OIQA) database, including 16 source images with their corresponding 320 distorted images. We add four commonly encountered distortions. These distortions are JPEG compression, JPEG2000 compression, Gaussian blur, and Gaussian noise. Then we conduct a subjective quality evaluation study in the VR environment based on the OIQA database. Considering that visual attention is more important in VR environment, head and eye movement data are also tracked and collected during the quality rating experiments. The 16 raw and their corresponding distorted images, subjective quality assessment scores, and the head-orientation data and eye-gaze data together constitute the OIQA database. Based on the OIQA database, we test some state-of-the-art full-reference IQA (FR-IQA) measures on equirectangular format or cubic format omnidirectional images. The results show that applying FR-IQA metrics on cubic format omnidirectional images could improve their performance. The performance of some FR-IQA metrics combining the saliency weight of three different types are also tested based on our database. Some new phenomena different from traditional IQA are observed.

**Keywords:** perceptual quality assessment; omnidirectional images; subjective experiment; objective model evaluation; visual saliency

DOI: 10.12142/ZTECOM.201901007

<http://kns.cnki.net/kcms/detail/34.1294.TN.20190313.0855.002.html>, published online March 13, 2019

Manuscript received: 2019-01-16

## 1 Introduction

Omnidirectional content could provide observers with immersive perception with the help of Head-Mounted Displays (HMDs). As an important component of virtual reality (VR), natural immersive videos provide the viewers with real-world scenes. The omnidirectional visual experience makes the user experience more immersive compared to traditional VR content generated by computer-aided 3D modeling. Therefore, we mainly consider natural immersive content, i.e., omnidirectional images, in this paper.

Because of the immersive experience providing by VRHMD, it is exciting to experience omnidirectional contents. However, due to the limitation of the photographic apparatus, transmis-

sion bandwidth, and display devices, etc, the content viewed by observers usually cannot live up to the expectation. As a consequence, it is significant to study the quality of experience (QoE) in VR environments. Many traditional image quality assessment (IQA) databases have been constructed by researchers, such as Live Mage Quality Assessment Database (LIVE) [1], TID2008—a database for evaluation of full-reference visual quality assessment metrics [2], categorical image quality (CSIQ) database [3], and quality assessment considering viewing distance and image resolution (VDID) [4], and some works related to assessing the quality of omnidirectional visual contents have also been done, such as [5]–[9]. However as far as we know, the databases relevant to omnidirectional image quality assessment are very few. And there is no database includ-

ing both subjective evaluation scores and eye movement data. So we have constructed one omnidirectional IQA (OIQA) database [10]. For traditional videos or images quality assessment, many efforts have been made on designing human visual system (HVS) based IQA metrics [11]–[15]. Visualizing immersive videos or omnidirectional (360-degree, equirectangular, VR) images [5] is different from traditional 2D videos or images. Observers are supposed to be in the central position of a sphere when visualizing immersive contents. The results in [5] and [16] illustrate that the view-port visualized by observers usually only occupies a portion of the whole omnidirectional images or videos. Because of the immersive experience, the visual attention of observers in the view-port of omnidirectional videos or images is different from the visual attention in plane 2D videos or images. Therefore, it is significant to study the new method to evaluate the quality of images and videos combining visual saliency in VR environment.

In this paper, we also explore the method of using human visual preferences to assess the quality of omnidirectional images. We test our ideas based on our OIQA database, which includes 16 raw reference omnidirectional images and their corresponding 320 degradation images under four kinds of distortion types. These distortions are JPEG [17] compression, JPEG2000 [18] compression, Gaussian blur, and Gaussian noise. The head and eye movement data are also collected in OIQA database. We first discuss the influence of intrinsic distortion of equirectangular projection. For comparison, the IQA metrics are tested on cubic images and we think cubic images have almost no such distortion. The performance of some FR-IQA metrics combining the saliency weight of three different types is also tested based on our database. Three different kinds of saliency maps include the global head movement characteristic map, global eye viewing preference map, and ground-truth visual saliency map.

The remainder of this paper is arranged as follows. We introduce the subjective omnidirectional IQA in Section 2. In Section 3, we evaluate several state-of-the-art FR-IQA models on the OIQA database and combine human visual preference in some IQA models. Some inspiring observations are proposed. We summarize and conclude the whole paper in Section 4.

## 2 Subjective Quality Assessment of Omnidirectional Images

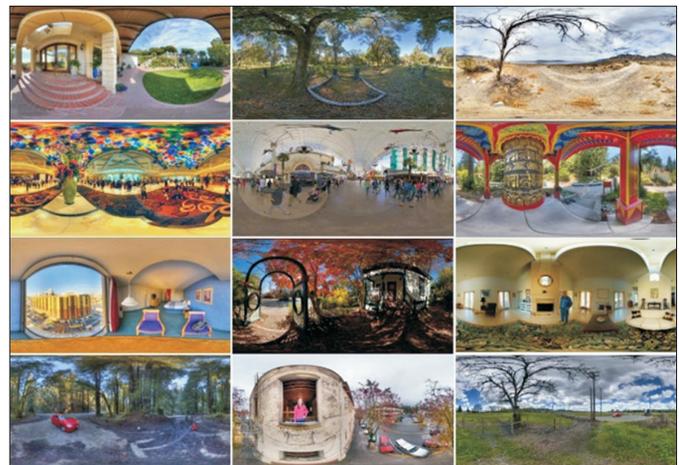
The image collection and quality degradation processes are first introduced in this section. Next, we introduce our experimental methodology to conduct subjective quality rating and capture head or eye movement data. Finally, we process and analyze the collected visual attention data and subjective quality ratings and present some conclusions we have observed.

### 2.1 Original and Distorted Equirectangular Images

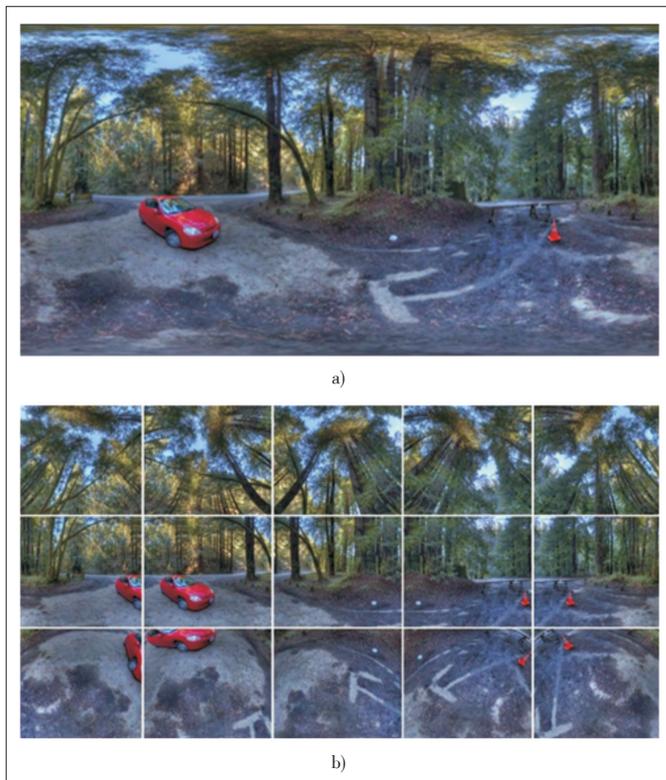
We collect 16 raw images which are captured by profession-

al photographers and available under Creative Commons (CC) copyright. The collected images are representative and have diversified textures. We show several sample raw images in **Fig. 1**. We zoomed in and carefully checked all raw images to avoid easily observed artifacts. This procedure can avoid the “intrinsic artifacts”. All of raw images have close resolutions which range from 11 332×5 666 to 13 320×6 660, and close perceptual quality. This procedure can reduce the influence of the original content’s quality on subjective ratings. We introduce four types of distortions to raw images, with five distortion levels for each type. The four types of distortions we introduced are JPEG compression, JPEG2000 compression, Gaussian blur, and white Gaussian noise (WGN), respectively, which are four commonly encountered distortions.

JPEG and JPEG2000 are the two commonly used compression methods to simulate the artifacts introduced during compression. In this paper, we introduced these two methods with five compression levels each to the raw images. The five levels are manually set to cover a wide perceptual quality range. Since omnidirectional contents are generally created, stored, compressed and transmitted in equirectangular format, we compress all images in equirectangular format directly and get degraded images. We also introduced another two commonly encountered distortions, which are Gaussian blur and WGN. In this paper, we mainly consider the blur and noise introduced during capturing. Omnidirectional images are usually captured by multiple cameras (e.g. camera array) and then stitched. To simulate the blur and noise introduced during capturing, the raw images are split into 15 small blocks. **Fig. 2a** represents the raw image. **Fig. 2b** represents the 15 split images of the raw image in Fig. 2a. These split images represent the scenes captured by each camera of the camera array. To simulate the distortions introduced at the sensor of each camera, we add Gaussian blur and WGN to 15 split images respectively. Then these split images with distortions added are projected back to one equirectangular image. Following these procedures, we



▲ **Figure 1.** Some source images in the omnidirectional image quality assessment database.



▲ Figure 2. One source image and 15 split images: a) the source image; b) 15 split images.

add the Gaussian blur and Gaussian noise to images more uniformly compared with adding distortions to equirectangular images directly. We also introduce five levels of blur and noise distortions to generate images with varying distortions. Thus we have 336 images in total in our OIQA database, including 16 raw images and their corresponding 320 distorted images.

## 2.2 Equipment and Software

The equipment we used to show the omnidirectional images to the subjects is HTC VIVE. This HMD has high-precision tracking ability and excellent graphics display technology. Specifically, the resolution of this display is 1 080×1 200 per eye and 2 160×1 200 combined with a field of view (FOV) of about 100 horizontal degrees and about 110 vertical degrees. The refresh rate of the HTC VIVE is 90 Hz. Additionally, the sensor of this HMD could provide head-orientation data at the same rate as the frame rate. In order to obtain eye movement data, a small eye-tracker named aGlass [19] is installed into the HMD. aGlass is an excellent VR eye-tracker with an error less than 0.5°. We also develop an interactive software based on Unity to display omnidirectional images and collect rating scores, head-orientation data and eye movement data.

## 2.3 Subjects

The total number of the subjects participated in our experiments was 20, including 5 females and 15 males. The age of

subjects ranged from 18 years to 30 years with an average of 24 years. All of the subjects reported normal or corrected-to-normal vision. As illustrated in [20], visually induced motion sickness (VIMS) in virtual reality environment could make the quality of experience (QoE) worse. And all of the subjects in our experiments reported that they did not have travel sickness.

## 2.4 Subjective Experiment Methodology

With the help of HTC VIVE, aGlass and the software, the experiments were conducted to obtain subjective rating score, head-orientation data and eye-movement data at the same time. Subjects were asked to seat in a rolling chair and be free to rotate the chair. This procedure is to ensure that the whole omnidirectional image could be visualized by observers. At the start of each experiment, the subjects were asked to calibrate the eye tracker which is installed in the HMD. Then, the visual attention data of the 16 raw omnidirectional images were collected. Each image was displayed for 20 seconds with a five-second gray screen displayed before showing the following omnidirectional image. In order to collect natural viewing visual attention data, all the subjects were asked to look around at least one circle in this step. Next, in order to make the subjects familiar with the distortions types and levels of the database, we conducted a training procedure. Finally, we conducted the formal quality rating experiment and collected rating scores. All images were displayed in a random order in this step. To avoid VIMS and fatigue, the subjects had enough rest time every 10 minutes during the experiment.

## 2.5 Data Processing and Analysis

Three types of data are collected, including raw subjective quality scores given by subjects, head movement data and eye gaze data, through the subjective experiment. In this section, we discuss the processing and analyzing procedure of these three kinds of data.

### 2.5.1 Subjective Quality Score Processing and Analysis

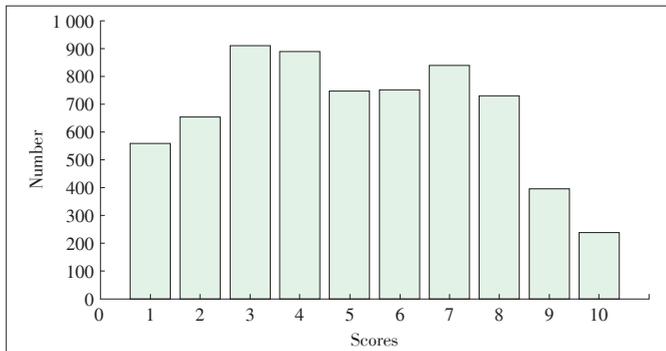
We first process the subjective rating scores of images. The mean opinion scores (MOS) are computed by the following formula:

$$MOS_j = \frac{\sum_{i=1}^N m_{ij}}{N}, \quad (1)$$

where  $N$  is the number of subjects and  $m_{ij}$  is the score assigned by subject  $i$  to image  $j$ . We also use the  $3\sigma$  principle to remove the outliers, which are scores far away from the average value. Fig. 3 illustrates the histogram of the distribution of subjective quality scores. Obviously, the subjective rating scores are distributed across all perceptual quality range.

### 2.5.2 Visual Attention Data Processing and Analysis

Head-orientation and eye-movement data within 20 seconds of 16 raw images were collected. To make the data more intui-



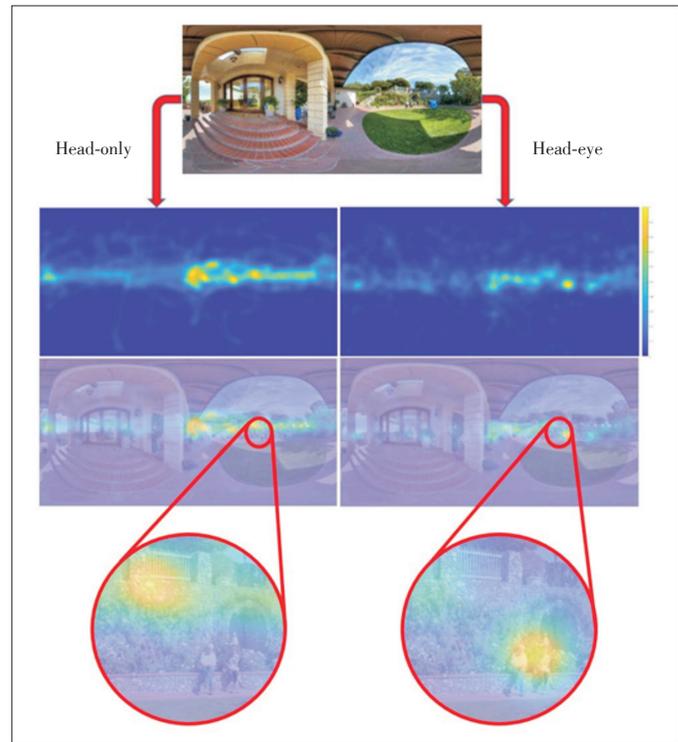
▲ Figure 3. Histogram of the subjective quality scores.

tive, we projected the view direction data and the eye-movement data from the 3D sphere space to the 2D equirectangular image. Head-only (view direction centered) saliency maps and head-eye saliency maps were created using the view direction information and the eye movement information respectively. We followed the method in [21] to get the saliency map. We applied a Gaussian filter of  $3.34^\circ$  of visual angle [22], [23] to fixation maps of the view-port image. Then the viewport images with spread fixations were back-projected into the sphere-map and then to the final equirectangular visual attention map. The OIQA database that includes the view-direction information, eye-fixation maps head saliency maps, and head-eye saliency maps was then released.

As shown in Fig. 4, for an image in OIQA database, we display its corresponding head-only saliency map and head-eye saliency map. Comparing the two saliency maps, we can see that the salient regions mainly centralize in the middle part of the equirectangular image nearby the equator. When viewing omnidirectional images in HMD, the top and bottom regions of an equirectangular image, i.e., the north and the south pole regions of the sphere, are less observed by the human subjects. Moreover, only a small part of the whole scene can be observed by users when viewing omnidirectional images in HMD. From an overall perspective, the head-only saliency map is similar with the head-eye saliency map. However, in details, there are many differences. Therefore, no matter in two-dimensional space or in three-dimensional space, it is reasonable and significant to assess the quality of images combining the visual saliency information.

### 2.5.3 Global Viewing Direction Bias

From Fig. 4, we can see that whether in head-only saliency map or in head-eye saliency map, the salient regions are all located around the equator of map. On the one hand, observers are more comfort when viewing the horizontal direction in HMD. On the other hand, when shooting panoramic images, salient scenes or objects are usually near the equator. Therefore we believe that it is one of the bottom layer features when viewing omnidirectional images using HMD. Fig. 5 shows the scatter diagrams of global viewing direction (head or eye) weight



▲ Figure 4. The head-only saliency map and head-eye saliency map.

proportion along with latitude, clustering over all subjects and all omnidirectional images. One-, two- and three-term gaussian fitting curves are also plotted in this figure. From the figure, we can see that three-term gaussian fitting curves can get relative good fitting performance. Three-term gaussian fitting curves can be plotted by:

$$f(x) = \alpha_1 e^{-\left(\frac{x-\beta_1}{\gamma_1}\right)^2} + \alpha_2 e^{-\left(\frac{x-\beta_2}{\gamma_2}\right)^2} + \alpha_3 e^{-\left(\frac{x-\beta_3}{\gamma_3}\right)^2}. \quad (2)$$

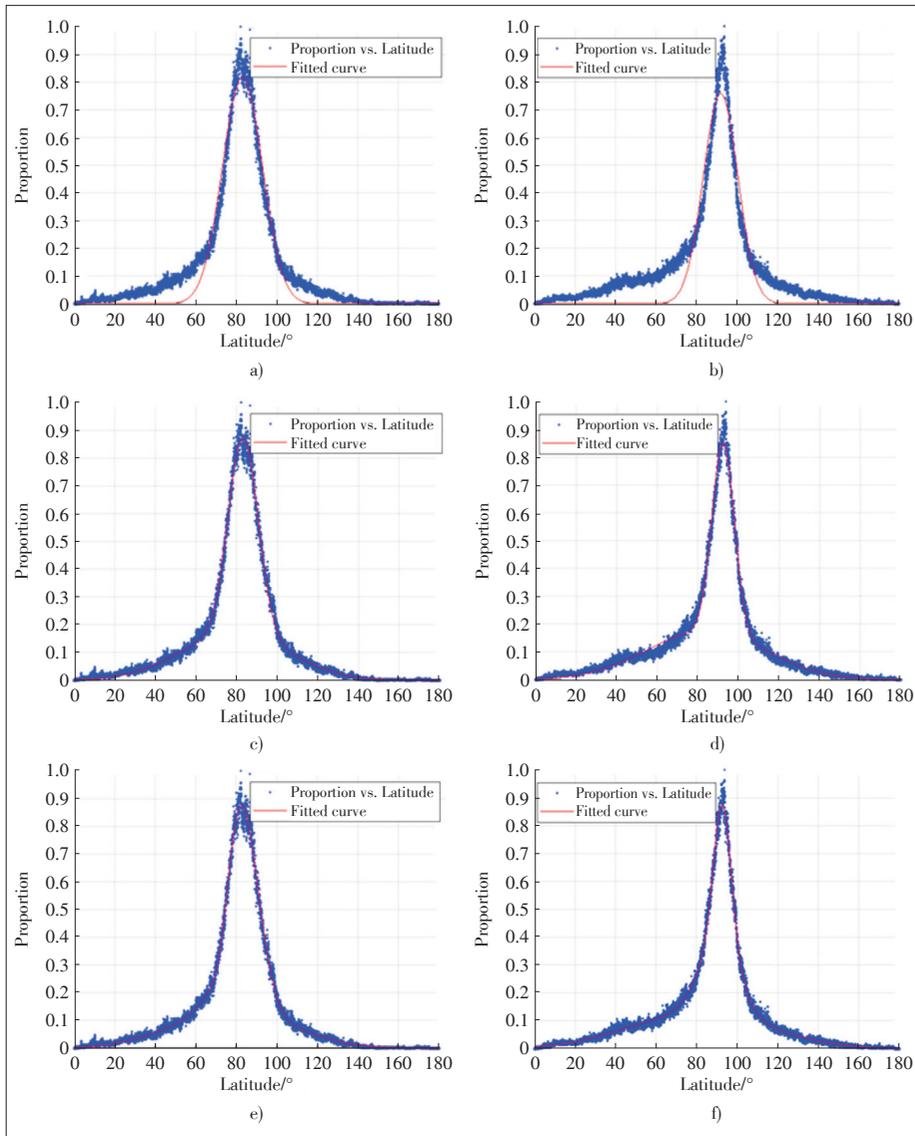
Table 1 shows the coefficient of fitting curves. The first row lists the nine parameters in Equ. (2). The second and third rows list nine coefficient values of global head movement direction and global eye viewing direction fitting curves, respectively. From the fitting curves, we could get the global viewing direction bias when viewing omnidirectional images. This global viewing direction bias can be used to generate global saliency weight, which is shown in Fig. 6. The global viewing bias of omnidirectional images can be used in their perceptual quality assessment. We will discuss this method in the following.

## 3 Comparison of Objective Quality Assessment on the OIQA database

### 3.1 Experimental Protocol

#### 3.1.1 FR-IQA Measures

After the experiment, we compared the performance of 11



▲ **Figure 5.** The scatter diagrams of head movement direction or eye viewing direction weight proportion along with latitude, clustering over all the subjects and all the omnidirectional images; a) One-term fitting curves of head movement direction; b) one-term fitting curves of eye viewing direction; c) two-term fitting curves of head movement direction; d) two-term fitting curves of eye viewing direction; e) three-term fitting curves of head movement direction; f) three-term fitting curves of eye viewing direction.

▼ **Table 1.** The coefficient values of head movement direction fitting curves and eye viewing direction fitting curves

| Parameter            | $\alpha_1$ | $\beta_1$ | $\gamma_1$ | $\alpha_2$ | $\beta_2$ | $\gamma_2$ | $\alpha_3$ | $\beta_3$ | $\gamma_3$ |
|----------------------|------------|-----------|------------|------------|-----------|------------|------------|-----------|------------|
| Fitting value (head) | -0.2404    | 72.75     | 7.53       | 0.7957     | 81.19     | 12.53      | 0.1353     | 77.78     | 40.94      |
| Fitting value (eye)  | 0.4943     | 92.75     | 6.05       | 0.2622     | 91.03     | 13.53      | 0.1288     | 81.26     | 48.26      |

state-of-the-art objective FR-IQA measures, which include 1) feature similarity (FSIM) [24], 2) gradient magnitude similarity deviation (GMSD) [25], 3) GSM [25], 4) gradient similarity (GSI) [26], 5) information content weighted structural similarity

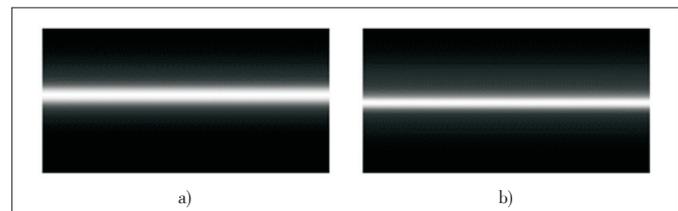
(IW-SSIM) [27], 6) mean squared error (MSE), 7) Multiscale structural similarity (MS-SSIM) [28], 8) peak signal-to-noise ratio (PSNR), 9) structural similarity (SSIM) [29], 10) visual information fidelity (VIF) [30], 11) visual saliency-induced index (VSI) [31]. When calculating the performance, we firstly mapped the predictions of the IQA models to subjective quality ratings through a five-parameter logistic function [32]–[34]:

$$f(x) = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 x + \beta_5, \quad (3)$$

in which  $x$  denotes the predicted scores;  $f(x)$  represents the corresponding mapped score;  $\beta_i$  ( $i=1, 2, 3, 4, 5$ ) are the parameters to be fitted. Then the mapped scores are compared with the subjective scores to measure the performance of the IQA models. In this paper, we use Pearson's Linear Correlation Coefficient (PLCC), root mean square error (RMSE) and Spearman rank correlation coefficient (SRCC) as criteria to evaluate the performance of algorithms. **Table 2** lists the performance of aforementioned IQA models under these three criteria.

### 3.1.2 Combining Human Visual Preference

The FR-IQA metrics are not only calculated on equirectangular images, but also calculated on cubic images. **Fig. 7** shows the omnidirectional images in equirectangular format or cubic format. We supposed that the cubic format omnidirectional images could simulate the view-port which the subjects really saw in HMD. It is obvious that cubic images have less distortion than equirectangular images in this figure. The omnidirectional images in cubic format are similar with traditional 2D images so we think the tra-



▲ **Figure 6.** The global saliency weight map generated from global viewing direction bias: a) Global head movement characteristic map; b) global eye viewing preference map.

ditional IQA metrics could perform better on images of this kind of format. We also combined the saliency weight information aforementioned on part of FR- IQA models to compare the influence of saliency map with different accuracy on the evaluation results. Three kinds of saliency maps, including global head movement characteristic map, global eye viewing preference map and ground-truth visual saliency map are discussed in this paper. The results are shown in Table 2.

### 3.2 Performance Comparison

As show in Table 2a, FSIM, GSI and VSI perform better than other metrics. Although their performance is pretty well, we still believe that the performance could be improved better, e.g., using saliency or other human visual preference to OIQA. In this paper, we first compare the SSIM metrics with or without the pre-processing method, which includes a low-pass filter and downsampling process. SSIM2 is the metric with pre-processing procedure while SSIM1 without in Table 2. It is obvious that this pre-processing method contribute a lot to the performance promotion of SSIM in the OIQA database. Except for these models, other state-of-the-art IQA models perform not well, and they undergo some performance drop when transferring from traditional images to omnidirectional images.



▲ Figure 7. Omnidirectional images of equirectangular format or cubic format: a) Equirectangular image; b)–g) cubic images, which are Front, Right, Back, Left, top and bottom in sequence.

▼ Table 2. Performance of FR-IQA models in terms of PLCC, SRCC and RMSE. The best three performing metrics are highlighted with bold font

a) Assessing the perceptual quality of omnidirectional images on equirectangular format images

| Metrics          | PLCC          | SRCC          | RMSE          |
|------------------|---------------|---------------|---------------|
| <b>FSIMc</b>     | <b>0.9188</b> | <b>0.9140</b> | <b>5.6800</b> |
| GMSD             | 0.7412        | 0.7378        | 9.6574        |
| GMSM             | 0.6768        | 0.6642        | 10.590        |
| <b>GSI</b>       | <b>0.9008</b> | <b>0.8924</b> | <b>6.2473</b> |
| IW-MSE           | 0.6207        | 0.7328        | 11.280        |
| IW-PSNR          | 0.7371        | 0.7328        | 9.7223        |
| IW-SSIM          | 0.7805        | 0.7766        | 8.9934        |
| MSE              | 0.3279        | 0.4971        | 13.590        |
| MS-SSIM          | 0.6745        | 0.6653        | 10.621        |
| PSNR             | 0.5060        | 0.4971        | 12.408        |
| SSIM1            | 0.5271        | 0.3479        | 12.225        |
| SSIM2            | 0.8888        | 0.8800        | 6.5917        |
| VIF              | 0.7878        | 0.7867        | 8.8614        |
| VIF <sub>p</sub> | 0.7555        | 0.7501        | 9.4246        |
| <b>VSI</b>       | <b>0.9087</b> | <b>0.9055</b> | <b>6.0059</b> |

b) Assessing the perceptual quality of omnidirectional images on cubic format images

| Metrics          | PLCC          | SRCC          | RMSE          |
|------------------|---------------|---------------|---------------|
| <b>FSIMc</b>     | <b>0.9316</b> | <b>0.9278</b> | <b>5.2273</b> |
| GMSD             | 0.7120        | 0.7042        | 10.101        |
| GMSM             | 0.6448        | 0.6393        | 10.996        |
| <b>GSI</b>       | <b>0.9215</b> | <b>0.9162</b> | <b>5.5878</b> |
| IW-MSE           | 0.6165        | 0.7110        | 11.327        |
| IW-PSNR          | 0.7179        | 0.7054        | 10.014        |
| IW-SSIM          | 0.7799        | 0.7755        | 9.0045        |
| MSE              | 0.3919        | 0.5693        | 13.235        |
| MS-SSIM          | 0.6699        | 0.6651        | 10.681        |
| PSNR             | 0.5621        | 0.5603        | 11.898        |
| SSIM1            | 0.4462        | 0.3870        | 12.875        |
| SSIM2            | 0.8843        | 0.8740        | 6.7175        |
| VIF              | 0.7725        | 0.7716        | 9.1351        |
| VIF <sub>p</sub> | 0.7761        | 0.7699        | 9.0723        |
| <b>VSI</b>       | <b>0.9236</b> | <b>0.9192</b> | <b>5.5158</b> |

c) Assessing the perceptual quality of omnidirectional images combining head movement direction information (Fig. 6a)

| Metrics      | PLCC          | SRCC          | RMSE          |
|--------------|---------------|---------------|---------------|
| <b>FSIMc</b> | <b>0.9118</b> | <b>0.9049</b> | <b>5.9061</b> |
| GMSM         | 0.6530        | 0.7035        | 10.895        |
| MSE          | 0.3420        | 0.5026        | 13.518        |
| PSNR         | 0.4123        | 0.3958        | 13.106        |
| SSIM1        | 0.4481        | 0.3663        | 12.861        |
| <b>SSIM2</b> | <b>0.8967</b> | <b>0.8844</b> | <b>6.3664</b> |
| <b>VSI</b>   | <b>0.9009</b> | <b>0.8946</b> | <b>6.2451</b> |

d) Assessing the perceptual quality of omnidirectional images combining eye viewing direction information (Fig. 6b)

| Metrics      | PLCC          | SRCC          | RMSE          |
|--------------|---------------|---------------|---------------|
| <b>FSIMc</b> | <b>0.9148</b> | <b>0.9078</b> | <b>5.8090</b> |
| GMSM         | 0.7350        | 0.7283        | 9.7549        |
| MSE          | 0.3407        | 0.5175        | 13.525        |
| PSNR         | 0.4364        | 0.4154        | 12.943        |
| SSIM1        | 0.4665        | 0.3849        | 12.725        |
| <b>SSIM2</b> | <b>0.8988</b> | <b>0.8849</b> | <b>6.3075</b> |
| <b>VSI</b>   | <b>0.9064</b> | <b>0.9005</b> | <b>6.0783</b> |

e) Assessing the perceptual quality of omnidirectional images combining original saliency map from subjects

| Metrics      | PLCC          | SRCC          | RMSE          |
|--------------|---------------|---------------|---------------|
| <b>FSIMc</b> | <b>0.9113</b> | <b>0.9015</b> | <b>5.9245</b> |
| GMSM         | 0.7508        | 0.7449        | 9.5026        |
| MSE          | 0.3475        | 0.5317        | 13.489        |
| PSNR         | 0.4858        | 0.4534        | 12.574        |
| SSIM1        | 0.5083        | 0.4077        | 12.388        |
| <b>SSIM2</b> | <b>0.8927</b> | <b>0.8779</b> | <b>6.4817</b> |
| <b>VSI</b>   | <b>0.9086</b> | <b>0.9027</b> | <b>6.0071</b> |

FR-IQA: full reference image quality assessment  
 FSIM: feature similarity  
 GMSD: gradient magnitude similarity deviation  
 GMSM: gradient magnitude similarity mean  
 GSI: gradient similarity  
 IW-MSE: information content weighted mean squared error  
 IW-PSNR: information content weighted peak signal-to-noise ratio  
 IW-SSIM: information content weighted structural similarity  
 MSE: mean squared error

MS-SSIM: multiscale structural similarity  
 PLCC: Pearson's Linear Correlation Coefficient  
 PSNR: peak signal-to-noise ratio  
 RMSE: root mean square error  
 SRCC: Spearman rank correlation coefficient  
 SSIM1: structural similarity  
 SSIM2: structural similarity with pre-processing procedure  
 VIF: visual information fidelity  
 VSI: visual saliency-induced index

There is much room to improve these models.

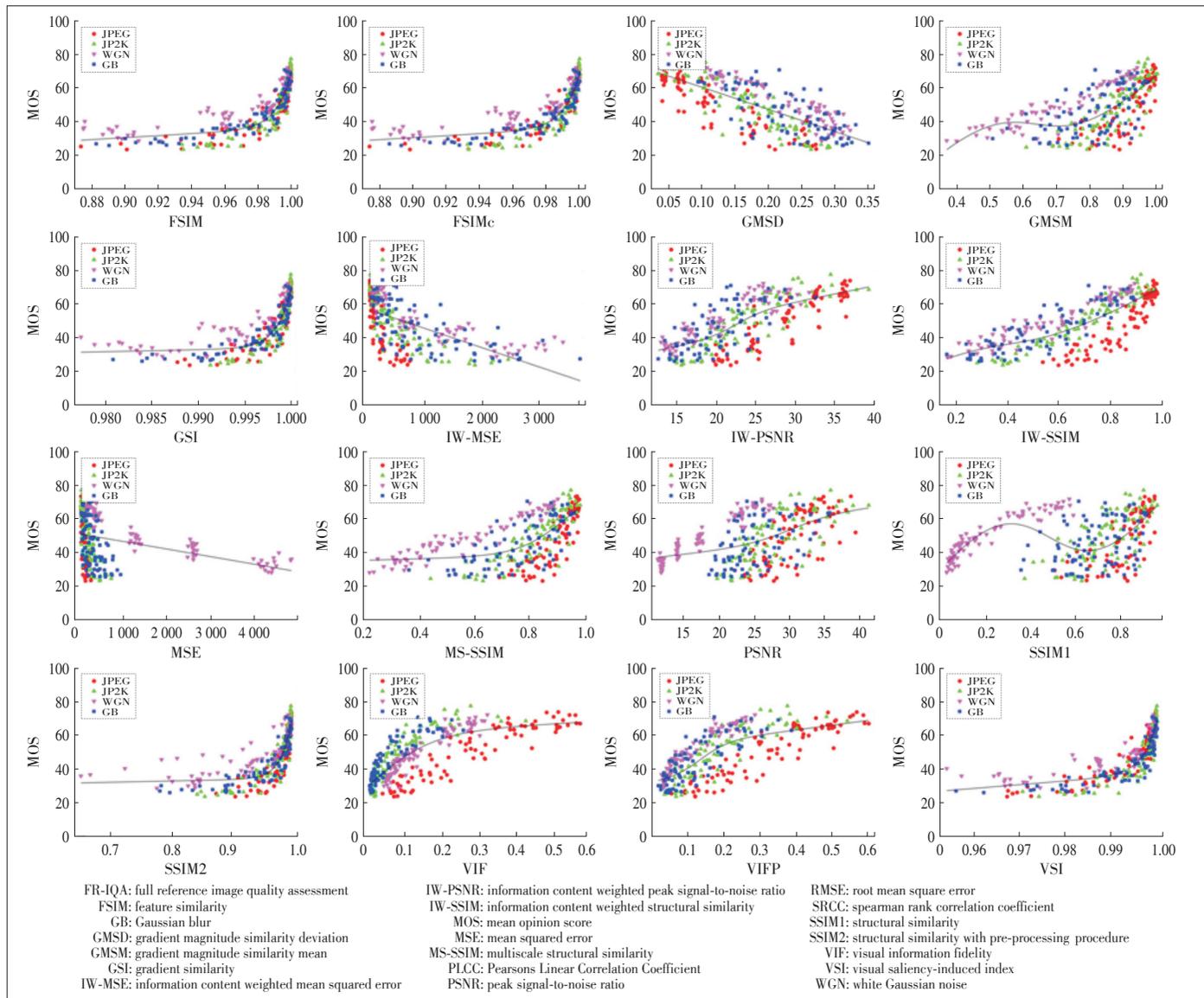
1) Performance Comparison of FR-IQA Metrics on Equirectangular Images and Cubic Images

The performance of these metrics are also calculated on omnidirectional images in cubic format, as shown in Table 2b. The best three performing models in Table 2b are also FSIM, GSI, and VSI. Compared with Table 2a, they have significant performance improvement. The FSIMc calculated on cubic images get the best performance in Table 2. The significant performance improvement also appears in MSE and PSNR metrics. However, for some other FR-IQA metrics, the performance improvement is not obvious. For some FR-IQA metrics, the performance even decreases when calculating on cubic images. It illustrates that this method cannot improve the perfor-

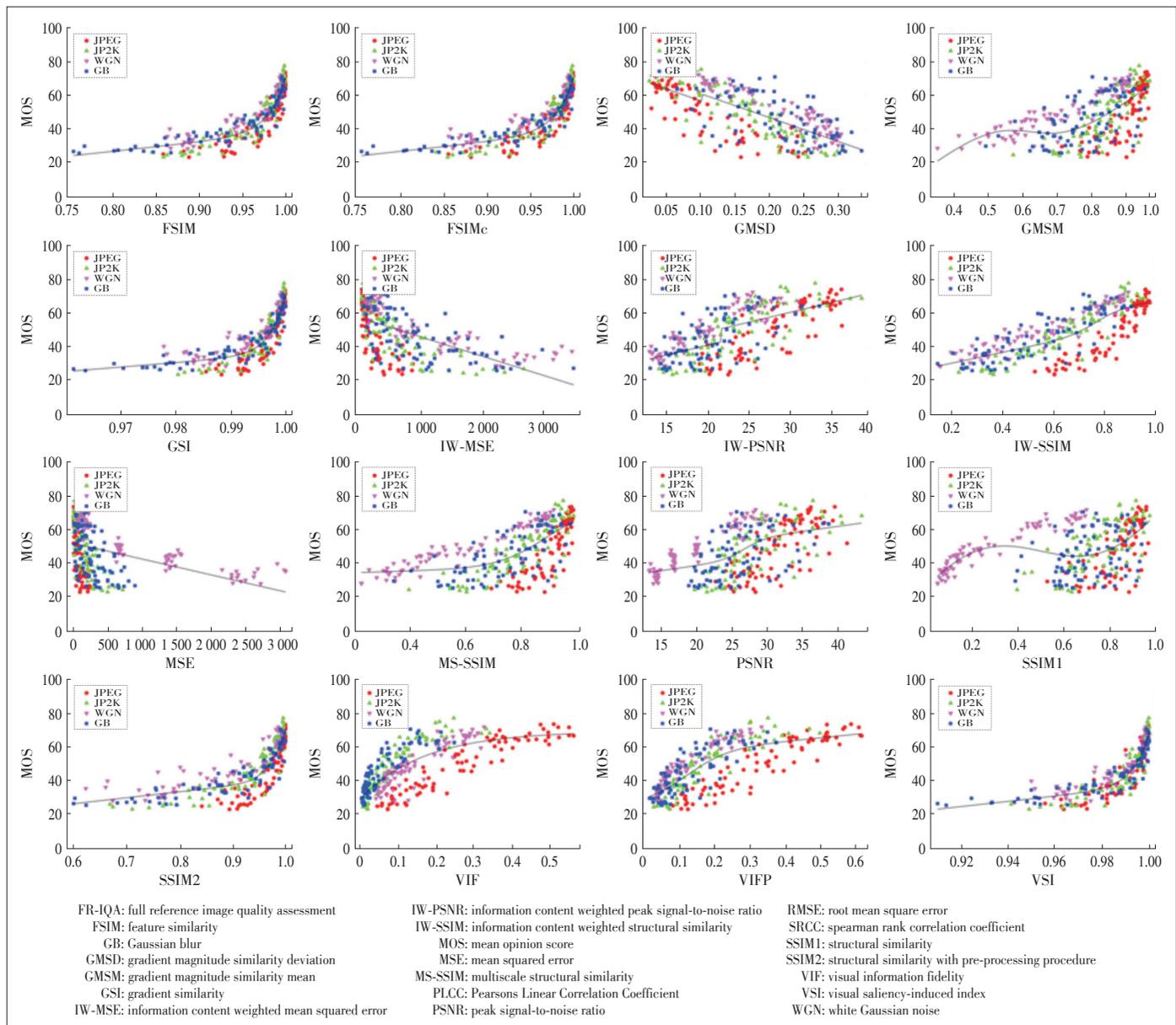
mance of all FR-IQA metrics. We compared the scatter diagrams of FR-IQA metrics on equirectangular images (Fig. 8) and on cubic images (Fig. 9), respectively. Detailed illustration will be discussed in Subsection 3.3.

2) Performance Comparison Combining Saliency Weight of Three Different types

In Tables 2c, 2d and 2e, we display the performance of seven FR-IQA metrics combining saliency weight with three different types, including the global head movement characteristic map, global eye viewing preference map and ground-truth visual saliency map. From these three tables, we find that as the accuracy of saliency map increases, the performance of these FR-IQA metrics increase in general, but the improvement is not impressive. Thus we propose that if the accuracy of



▲ Figure 8. Scatter plots of subjective MOS versus FR-IQA model prediction, including FSIM, FSIMc, GMSD, GMSM, GSI, IW-MSE, IW-PSNR, IW-SSIM, MSE, MS-SSIM, PSNR, SSIM1, SSIM2, VIF, VIFp, and VSI, based on equirectangular format images. The distortion types are JPEG compression (red points), JPEG2000 compression (green points), WGN (magenta points), and GB (blue points).



▲ Fig. 9. Scatter plots of subjective MOS versus FR-IQA model prediction, including FSIM, FSIMc, GMSD, GMSM, GSI, IW-MSE, IW-PSNR, IW-SSIM, MSE, MS-SSIM, PSNR, SSIM1, SSIM2, VIF, VIFp, and VSI, based on cubic format images. The distortion types are JPEG compression (red points), JPEG2000 compression (green points), WGN (magenta points), GB (blue points).

saliency map of omnidirectional images is not assured, researchers could combining the global saliency weight (Fig. 6) in their IQA algorithms. We find from the performance in Table 2 that the performance of some metrics decreases after combining the saliency weight. We think it is because that the intrinsic distortion of equirectangular projection is inconsistent with the saliency map. Relative issues need further research.

### 3.3 Differences Between Omnidirectional IQA and Traditional IQA

We select 16 FR-IQA models (FSIM, FSIMc, GMSD, GMSM, GSI, IW-MSE, IW-PSNR, IW-SSIM, MSE, MS-SSIM,

PSNR, SSIM1, SSIM2, VIF, VIFp, and VSI) and illustrate their scatter plots on equirectangular images (Fig. 8) and on cubic images (Fig. 9). The models we selected contain high performance metrics (such as FSIMc and VSI) and classical IQA metrics (such as SSIM and PSNR). As shown in Fig. 8, the scatter points whose color are magenta represent the distortion type of WGN. Compared with the scatter points of the other three distortion types, it is obvious that these scatter points are always far from the fitted curve. The scatter points of WGN are almost always higher than the scatter points of the other three distortion types. It means that these IQA models have predicted lower quality scores than the ideal values for distortion type WGN.

We believe this phenomenon is partly caused by the subjective ratings and partly caused by the objective IQA models. Another phenomenon we observed from the scatter plots in Figs. 8 and 9 is that the scatter points of JPEG compression do not fit well in some metrics. We think it is mainly caused by the subjective ratings.

#### 1) Subjective Rating Differences

For the exception to the WGN distortion, we see that this phenomenon is observed in various IQA models in Fig. 8. In traditional images, for all kinds of distortions, most IQA models show quite consistent predictions. However, in omnidirectional images, exception to the WGN distortion is observed. Therefore we believe it is partly caused by the subjective ratings. We believe that people prefer high-frequency content when viewing VR stimuli. And this preference leads to the exceptional subjective ratings. Observers will have more comfortable visual experiences when viewing high frequency content. Compared with traditional displays, subjects can only see the view-port image. This limited displaying effects of current VR-HMD also leads to the exceptional subjective ratings. Because the contents are not completed in the view-port and subjects will be annoyed with losing image details. We introduce four types of distortions in this paper, including JPEG compression, JPEG2000 compression, Gaussian noise, and Gaussian blur. Gaussian noise adds high frequency information to the image. The rest three distortions reduce the high frequency information and image details. For the exception to the JPEG compression distortion, we think humans' perceptual assessment of color distortion in VR environment is also different with that in traditional 2D displays. Some following work can be done regarding this phenomenon.

#### 2) Objective Measure Differences

For the exception to the WGN distortion, we also believe that this phenomenon is caused by the intrinsic distortion of equirectangular projection. Fig. 9 shows the scatter diagrams of MOS versus FR-IQA metrics on cubic images and we think cubic images have little intrinsic distortion. We excitingly find from the figure that for some models, such as FSIM, GSI, and VSI, the scatter points of the distortion type WGN are closer to the fitted curves, compared with Fig. 8, although they also a little far away from the fitted curve. Thus the performance of these metrics in Table 2b are better than those in Table 2a. Thus the phenomenon aforementioned is partly caused by the intrinsic distortion of equirectangular projection.

## 4 Conclusion and Future Work

In this paper, we investigate the methodology of assessing the quality of omnidirectional images. We first construct an omnidirectional IQA database. The database includes 16 source images with their corresponding 320 degraded images. We add four most commonly encountered distortions, including JPEG compression, JPEG2000 compression, Gaussian noise, and

Gaussian blur. We collect the subjective quality scores, view-orientation information, and eye-movement data during the experiment. By comparing objective FR-IQA models on the OIQA database, we propose that humans prefer high frequency content and image details in VR HMDs, and the losing of image details can do a lot of harm to the visual experience in the VR case. By comparing the performance of state-of-the-art objective FR-IQA models tested on the equirectangular images and cubic images, respectively, we find that calculating the IQA metrics on cubic images could improve some metrics' performance. Visual saliency information should also be combined in the IQA metrics, and more accurate saliency information will make the performance better.

#### References

- [1] SHEIKH H R, WANG Z, CORMACK L, et al. Live Image Quality Assessment Database Release 2 [EB/OL]. (2005)[2018]. <http://live.ece.utexas.edu/research/quality>
- [2] PONOMARENKO N, LUKIN V, ZELENSKY A. Tid2008-A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics [J]. *Advances of Modern Radioelectronics*, 2009, 10(4): 30-45
- [3] LARSON E C, CHANDLER D. Categorical Image Quality (CSIQ) Database [EB/OL]. (2010)[2018]. <http://vision.okstate.edu/csiq>
- [4] GU K, LIU M, ZHAI G T, et al. Quality Assessment Considering Viewing Distance and Image Resolution [J]. *IEEE Transactions on Broadcasting*, 2015, 61(3): 520-531. DOI: 10.1109/tbc.2015.2459851
- [5] RAI Y, LE CALLET P, GUILLOTTEL P. Which Saliency Weighting for Omnidirectional Image Quality Assessment? [C]/Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 2017: 1-6. DOI: 10.1109/QoMEX.2017.7965659
- [6] UPENIK E, RERÁBEK M, EBRAHIMI T. Testbed for Subjective Evaluation of Omnidirectional Visual Content [C]/2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 2016: 1-5. DOI: 10.1109/PCS.2016.7906378
- [7] YU M, LAKSHMAN H, GIROD B. A Framework to Evaluate Omnidirectional Video Coding Schemes [C]/IEEE International Symposium on Mixed and Augmented Reality, Fukuoka, Japan, 2015: 31-36. DOI: 10.1109/ISMAR.2015.12
- [8] SUN W, GU K, ZHAI G T, et al. CVIQD: Subjective Quality Evaluation of Compressed Virtual Reality Images [C]/IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017: 3450-3454. DOI: 10.1109/ICIP.2017.8296923
- [9] DUAN H Y, ZHAI G T, YANG X K, et al. IVQAD 2017: An Immersive Video Quality Assessment Database [C]/International Conference on Systems, Signals and Image Processing (IWSSIP), Poznan, Poland, 2017: 1-5. DOI: 10.1109/IWSSIP.2017.7965610
- [10] DUAN H Y, ZHAI G T, MIN X K, et al. Perceptual Quality Assessment of Omnidirectional Images [C]/IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 2018: 1-5. DOI: 10.1109/ISCAS.2018.8351786
- [11] ZHAI G T, CAI J F, LIN W S, et al. Cross-Dimensional Perceptual Quality Assessment for Low Bit-Rate Videos [J]. *IEEE Transactions on Multimedia*, 2008, 10(7): 1316-1324. DOI: 10.1109/tmm.2008.2004910
- [12] ZHAI G T, WU X L, YANG X K, et al. A Psychovisual Quality Metric in Free-Energy Principle [J]. *IEEE Transactions on Image Processing*, 2012, 21(1): 41-52. DOI: 10.1109/tip.2011.2161092
- [13] GU K, ZHAI G T, YANG X K, et al. Using Free Energy Principle for Blind Image Quality Assessment [J]. *IEEE Transactions on Multimedia*, 2015, 17(1): 50-63. DOI: 10.1109/tmm.2014.2373812
- [14] GU K, ZHAI G T, LIN W S, et al. No-Reference Image Sharpness Assessment

- in Autoregressive Parameter Space [J]. *IEEE Transactions on Image Processing*, 2015, 24(10): 3218–3231. DOI: 10.1109/tip.2015.2439035
- [15] ZHU W, ZHAI G, SUN W, et al. On the Impact of Environmental Sound on Perceived Visual Quality [C]//Pacific Rim Conference on Multimedia (PCM), Harbin, China, 2017: 723–734
- [16] XU M, LI C, LIU Y F, et al. A Subjective Visual Quality Assessment Method of Panoramic Videos[C]//IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 2017: 517–522. DOI: 10.1109/ICME.2017.8019351
- [17] WALLACE G K. The JPEG still Picture Compression Standard [J]. *IEEE Transactions on Consumer Electronics*, 1992, 38(1): xviii–xxxiv. DOI: 10.1109/30.125072
- [18] SKODRAS A, CHRISTOPOULOS C, EBRAHIMI T. The JPEG 2000 still Image Compression Standard [J]. *IEEE Signal Processing Magazine*, 2001, 18(5): 36–58. DOI: 10.1109/79.952804
- [19] 7invensun. Aglass [EB/OL]. (2017)[2018]. <http://www.aglass.com>
- [20] DUAN H Y, ZHAI G T, MIN X K, et al. Assessment of Visually Induced Motion Sickness in Immersive Videos [C]//18th Pacific-Rim Conference on Multimedia (PCM), Harbin, China, 2017, pp. 662–672. DOI: 10.1007/978-3-319-77380-3\_63
- [21] RAI Y, CALLET P L. A dataset of head and eye movements for 360 degree images [C]//ACM on Multimedia Systems Conference, Taipei, China, 2017: 205–210. DOI: 10.1145/3083187.3083218
- [22] CURCIO C A, SLOAN K R, KALINA R E, et al. Human Photoreceptor Topography [J]. *The Journal of Comparative Neurology*, 1990, 292(4): 497–523. DOI: 10.1002/cne.902920402
- [23] ENGELKE U, ZHANG W, CALLET P L. Perceived Interest Versus Overt Visual Attention in Image Quality Assessment [J]. *Proceedings of SPIE*, vol. 9394, 2015. DOI: 10.1117/12.2086371
- [24] ZHANG L, ZHANG L, MOU X Q, et al. FSIM: A Feature Similarity Index for Image Quality Assessment [J]. *IEEE Transactions on Image Processing*, 2011, 20(8): 2378–2386. DOI: 10.1109/tip.2011.2109730
- [25] XUE W F, ZHANG L, MOU X Q, et al. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index [J]. *IEEE Transactions on Image Processing*, 2014, 23(2): 684–695. DOI: 10.1109/tip.2013.2293423
- [26] LIU A M, LIN W S, NARWARIA M. Image Quality Assessment Based on Gradient Similarity [J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 1500–1512. DOI: 10.1109/tip.2011.2175935
- [27] WANG Z, LI Q. Information Content Weighting for Perceptual Image Quality Assessment [J]. *IEEE Transactions on Image Processing*, 2011, 20(5): 1185–1198. DOI: 10.1109/tip.2010.2092435
- [28] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale Structural Similarity for Image Quality Assessment [C]//Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, USA, 2003: 1398–1402. DOI: 10.1109/ACSSC.2003.1292216
- [29] WANG Z, BOVIK A C, SHEIKH H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity [J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600–612. DOI: 10.1109/tip.2003.819861
- [30] SHEIKH H R, BOVIK A C. Image Information and Visual Quality [J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 430–444. DOI: 10.1109/tip.2005.859378
- [31] ZHANG L, SHEN Y, LI H Y. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment [J]. *IEEE Transactions on Image Processing*, 2014, 23(10): 4270–4281. DOI: 10.1109/tip.2014.2346028
- [32] MIN X K, GU K, ZHAI G T, et al. Blind Quality Assessment Based on Pseudo-Reference Image [J]. *IEEE Transactions on Multimedia*, 2018, 20(8): 2049–2062. DOI: 10.1109/tmm.2017.2788206
- [33] MIN X K, GU K, ZHAI G T, et al. Saliency-Induced Reduced-Reference Quality Index for Natural Scene and Screen Content Images [J]. *Signal Processing*, 2018, 145: 127–136. DOI: 10.1016/j.sigpro.2017.10.025
- [34] MIN X K, MA K D, GU K, et al. Unified Blind Quality Assessment of Compressed Natural, Graphic, and Screen Content Images [J]. *IEEE Transactions on Image Processing*, 2017, 26(11): 5462–5474. DOI: 10.1109/tip.2017.2735192

## Biographies

**DUAN Huiyu** (huiyuduan@sjtu.edu.cn) received the B.E. degree from the University of Electronic Science and Technology of China in 2017. He is currently pursuing the Ph.D. degree at the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, China. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing.

**ZHAI Guangtao** received the B.E. and M.E. degrees from Shandong University, China in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, China in 2009. From 2008 to 2009, he was a visiting student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, where he was a post-doctoral fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen - Nuremberg, Germany. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing. He received the National Excellent Ph. D. Thesis Award from the Ministry of Education of China in 2012.

**MIN Xiongkuo** received the B.E. degree from Wuhan University, China in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, China in 2018. From 2016 to 2017, he was a visiting student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a post-doctoral fellow with Shanghai Jiao Tong University. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He received the Best Student Paper Award from the IEEE ICME 2016.

**ZHU Yucheng** received the B.E. degree from the Shanghai Jiao Tong University, China in 2015. He is currently pursuing the Ph.D. degree at the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He received Grand Challenge Best Performance Awards in ICME 2017 and 2018.

**FANG Yi** is an undergraduate student at Shanghai Jiao Tong University, China and will receive the B.E. degree in 2019. She will pursue the M.E. degree at the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University. Her research interests include image quality assessment, visual attention modeling, and perceptual signal processing.

**YANG Xiaokang** received the B.S. degree from Xiamen University, China in 1994, the M.S. degree from the Chinese Academy of Sciences, China in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, China in 2000. From 2000 to 2002, he was a Research Fellow with the Centre for Signal Processing, Nanyang Technological University, Singapore. From 2002 to 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he visited the Institute for Computer Science, University of Freiburg, Freiburg im Breisgau, Germany, as an Alexander von Humboldt Research Fellow. He is currently a Distinguished Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, where he is also the Deputy Director of the Institute of Image Communication and Information Processing. His current research interests include image processing and communication, computer vision, and machine learning.