IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 24, 2019, accepted November 11, 2019, date of publication November 18, 2019,
date of current version November 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953983

Impacts of Retina-Related Zones on Quality Perception of Omnidirectional Image

HUYEN T. T. TRAN[1], DUC V. NGUYEN[1], (Student Member, IEEE), NAM PHAM NGOC[2], (Member, IEEE), TRANG H. HOANG[3], TRUONG THU HUONG[3], (Member, IEEE), AND TRUONG CONG THANG[1], (Senior Member, IEEE)

[1]Department of Computer and Information Systems, The University of Aizu, Aizuwakamatsu 965-8580, Japan
[2]Director of Engineering and Technology Program, Vin University Project, Hanoi 100000, Vietnam
[3]School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi 100000, Vietnam

Corresponding author: Huyen T. T. Tran (tranhuyen1191@gmail.com)

This work was supported in part by the Competitive Fund of The University of Aizu, Japan.

ABSTRACT Virtual Reality (VR), which brings immersive experiences to viewers, has been gaining popularity in recent years. A key feature in VR systems is the use of omnidirectional content, which provides 360-degree views of scenes. In this work, we study the human quality perception of omnidirectional images, focusing on different zones surrounding the foveation point. For that purpose, an extensive subjective experiment is carried out to assess the perceptual quality of omnidirectional images with non-uniform quality. Through experimental results, the impacts of different zones are analyzed. Moreover, twenty-five objective quality metrics, including foveal quality metrics, are evaluated using our database. It is quantitatively shown that the zones corresponding to the fovea and parafovea of human eyes are extremely important for quality perception, while the impacts of the other zones corresponding to the perifovea and periphery are small. Besides, most of the investigated metrics are found to be not effective enough to reflect the quality perceived by viewers. Our database has been made available to the public.

INDEX TERMS Omnidirectional content, subjective quality assessment, foveation feature, virtual reality, image quality, image processing.

I. INTRODUCTION

In order to bring immersive experiences to viewers, virtual reality (VR) systems employ omnidirectional content which contains 360-degree views of scenes. Unlike traditional content displayed using a flat screen, omnidirectional content is usually consumed using Head Mounted Displays (HMDs). Also, only a small part of the full content (called *viewport*) corresponding to the current viewing direction is actually seen by the viewer at a moment [1].

Because omnidirectional (or 360-degree) content has very high bitrate, a key challenge in omnidirectional content delivery is how to optimize system resources while still ensuring satisfactory user experiences. For that, many encoding and delivery solutions have been proposed in the literature, where the (estimated) viewport is provided with high quality and the remaining part with low quality [2]–[4]. Moreover, in VR systems, foveated imaging, which decreases quality of zones far from the viewer's foveation point [5], [6], can be used to further reduce resource consumption [6], [7]. However, the estimated viewing direction could be very different from the actual one when the system delay is large [8]. Even the viewer may suddenly turn to look at the back. In these cases, the actual viewport may have low quality in the central part and high quality in the periphery. In other words, the central part may have higher quality (called scenario *S*#1) or lower quality (called scenario *S*#2) than the periphery, both resulting in omnidirectional content with non-uniform quality.

It is well-known that human visual acuity is spatially variable [9], [10]. In particular, when a person gazes at a point on an image, called *foveation point*, a zone closer to this point is perceived to be sharper than the others. This is because that the human eyes have higher sensitivities to distortions in the central than in the periphery. Hence, the understanding of the impacts of different zones on the perceptual quality is obviously of indispensable necessity in the context of omnidirectional content.

The associate editor coordinating the review of this manuscript and approving it for publication was Ke Gu.

In the literature, there are only a few existing studies on subjective quality assessments of images/videos with non-uniform quality [7], [11], [12]. However, most of these studies are devoted to traditional content [11], [12]. In [11], each image is divided into four zones of equal widths. The quality levels of these zones are gradually decreased with a fixed step size. It is found that, when the step size is small, the difference of perceptual quality between the non-uniform and uniform videos is insignificant. In addition, the maximum value of the step size without causing significant quality differences depends on content characteristics. In [12], each image is divided into three zones, which are foveal, blending, and peripheral zones. Experimental results show that participants barely notice quality decreases at the peripheral zones of the eccentricity larger than 7.5 degrees. Also, an evaluation of four subjective quality assessment methods is presented. It is indicated that the Absolute Category Rating (ACR) method is the best method for subjective quality assessments of non-uniform images. Different from [11], [12], our study focuses on quality perception of omnidirectional images.

In the literature, there have been some studies on subjective quality assessments of omnidirectional content [13]–[16]. In these studies, various distortion types such as compression and Gaussian blur are employed to generate images rated in experiments. In particular, the authors in [13] consider 4 distortion types, namely JPEG compression, JPEG2000 compression, Gaussian blur, and Gaussian noise. In [14], only one distortion type of H.265/HEVC compression is used. The study in [15] utilizes three distortion types of JPEG compression, JPEG2000 compression, and HEVC-intra. In [16], the distorted images are generated by down sampling and JPEG compression. However, since the distortions are uniform in [13]–[16], the databases do not contain non-uniform quality images. Also, the foveation feature of the human eyes is not taken into account in constructing these databases.

The work in [7] is the only previous study on omnidirectional content with non-uniform quality. In [7], the authors focus on answering the question of how to spatially reduce image quality without causing impacts on user perception. For that purpose, they propose to divide an omnidirectional image into three zones according to three regions of the human retina, namely the macula, the near periphery, and the far periphery. The image quality corresponding to each region is decreased step by step until participants notice a perceptual difference. The encoding parameters obtained just before that point are modeled and then used as a guide for spatially reducing image quality without perceptual loss. It is shown that this approach could save loading time by about 90% in comparison to a conventional approach using uniform quality. However, the impacts of different zones are not quantitatively quantified in [7]. In this study, we quantify the impacts of five zones corresponding to five regions of the human retina, namely the fovea, the parafovea, the perifovea, the near periphery, and the far periphery. In addition, only one scenario (i.e., *S*#1) of spatial quality changes is considered in [7]. Also, there is no performance evaluation of existing metrics conducted in [7]. In this paper, we construct a large database consisting of not only scenario S#1 but also scenario S#2. By using this database, an extensive evaluation of twenty-five objective quality metrics is also performed.

Over several decades, a large number of objective quality metrics have been proposed [17]–[21]. Some of these metrics take into account the foveation feature, hereafter referred to as *foveal quality metrics* [20], [21]. However, all these metrics are specific to traditional content. There has been no existing foveal quality metric for omnidirectional content so far.

In our previous study [22], a comparison between eight state-of-the-art quality metrics has been conducted. Experimental results show that PSNR turns out to be the most effective metric for quality assessment of omnidirectional videos. However, it is worth to note that images used in that study have uniform quality. As shown later in this paper, PSNR is actually not effective when the quality is spatially variable. To the best of our knowledge, no extensive evaluation of objective quality metrics for omnidirectional images with non-uniform quality has been conducted in the literature.

In this study, our purposes related to user perception of omnidirectional content in VR systems include:

- Subjective study on the impacts of retina-related zones on quality perception of omnidirectional images.
- Performance evaluation of existing objective quality metrics, especially foveal quality metrics, for omnidirectional images having non-uniform quality.

To that end, our major contributions are as follows. First, we present a detailed description of a VR viewing geometry and the human retina. This description helps in designing subjective experiments and in calculating parameters used in foveal quality metrics. Second, we carry out an extensive subjective experiment with 512 stimuli of non-uniform quality. The quality zones of the stimuli are designed based on five regions of the human retina. To the best of our knowledge, this is the first database of omnidirectional images aiming at the impacts of the five zones related to the human retina. Third, using a simple zone-weighted formulation, we quantify, for the first time, the impacts of different zones on the perceptual quality. It is quantitatively found that the zones corresponding to the fovea and parafovea of the human retina are extremely important for quality perception. Also, the impacts of zones are strongly affected by content characteristics. Fourth, we evaluate the correlation of twenty-five objective quality metrics against subjective scores. Experimental results indicate that most of these metrics, even the foveal ones, are not very effective when the viewport quality is spatially variable.

The remainder of the paper is organized as follows. A description of a VR viewing geometry and the human retina is presented in Section II. Section III presents the details of the subjective experiment. The analysis of perceptual behaviors using the experimental results is provided in Section IV. Then, an evaluation of quality metrics is presented in Section V. Section VI concludes the paper and provides an outlook on future work.

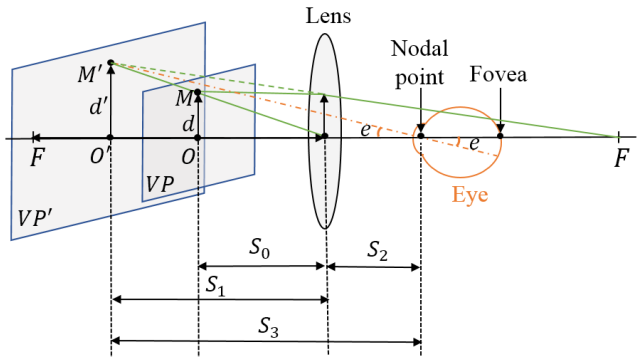


FIGURE 1. Typical viewing geometry in VR systems.

II. OVERVIEW

In this section, the viewing geometry in VR systems is first presented. Then, the regions in human retina are described.

A. VIEWING GEOMETRY IN VR SYSTEMS

Fig. 1 illustrates a typical viewing geometry in VR systems. Assume that VP is the displayed viewport, the lens in the HMD produces a virtual viewport VP' that is further formed on the retina in the human eyes. Eccentricity e (degrees) is used to measure the angular distance from the central gaze direction to any point in the virtual viewport VP'.

Let F (units of length) be the focal length of the lens. S_0, S_1, and S_2 (units of length) respectively denote the distances from the lens to the displayed viewport VP, the virtual viewport VP', and the eye. Based on lens equations, the distance from the lens to the virtual viewport S_1 is computed by

$$S_1 = S_0 \times \frac{F}{F - S_0}. \tag{1}$$

Then, the distance from the eye to the virtual viewport is calculated by

$$S_3 = S_1 + S_2. \tag{2}$$

Let $W_p \times H_p$ (pixels) and $W_l \times H_l$ (units of length) respectively be the width and height of the displayed viewport VP in pixels and units of length. The width of the virtual viewport VP' in pixels and units of length is respectively given by the following equations.

$$W'_p = W_p. \tag{3}$$

$$W'_l = W_l \times \frac{F}{F - S_0}. \tag{4}$$

Also, the height of the virtual viewport VP' in pixels and units of length is respectively calculated by

$$H'_p = H_p \tag{5}$$

and

$$H'_l = H_l \times \frac{F}{F - S_0}. \tag{6}$$

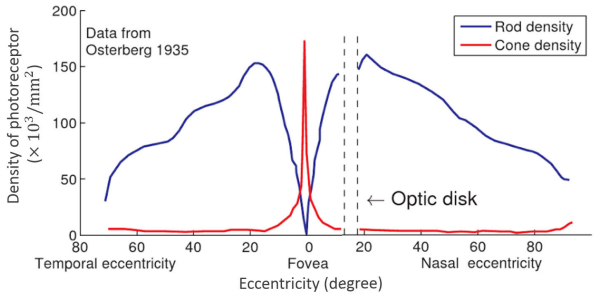


FIGURE 2. Density of photoreceptors in the retinal [23].

Assume that the foveation point is the center $O' = (x_{O'}, y_{O'})$ (pixels) in the virtual viewport VP'. Point $O = (x_O, y_O)$ (pixels) in the displayed viewport VP corresponding to point O' is determined by

$$x_O = x_{O'} \tag{7}$$

and

$$y_O = y_{O'}. \tag{8}$$

Let M be a point at the position of (x_M, y_M) (pixels) in the displayed viewport VP. The position of the virtual point $M' = (x_{M'}, y_{M'})$ (pixels) corresponding to point M is

$$x_{M'} = x_M \tag{9}$$

and

$$y_{M'} = y_M. \tag{10}$$

The distance in pixels from pixel M' to the foveation point O' is

$$d' = \sqrt{\left(\frac{(x_{M'} - x_{O'}) \times W'_l}{W'_p}\right)^2 + \left(\frac{(y_{M'} - y_{O'}) \times H'_l}{H'_p}\right)^2}. \tag{11}$$

The eccentricity e of point M' in the virtual viewport VP' is given by

$$e(x_{M'}, y_{M'}) = tan^{-1}\left(\frac{d'}{S_3}\right). \tag{12}$$

It should be noted that parameters of a point on the virtual viewport are what actually used in a foveal quality metric. Moreover, given the knowledge of the human visual system, the points on the virtual viewport can be divided according to the regions of the retina.

B. REGIONS IN HUMAN RETINA

In the human retina, there are two types of photoreceptors, namely rods and cones, each plays an important role in human visual system. In particular, cones function most effectively in relatively bright light and are responsible for color vision and visual acuity. Meanwhile, rods have higher sensitivities to light, and thus they function mainly in dim light.

Fig. 2 shows the density of photoreceptors in the human retina. It can be seen that most cones are concentrated at

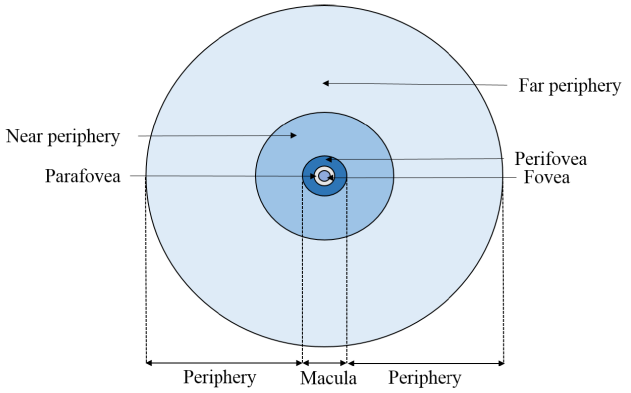


FIGURE 3. Five regions of the retina.

the center of the retina, whereas rods are located away from the center. Visual information from photoreceptors are then collected by the so-called ganglion cells. The optic disk is where axons from ganglion cells exit the retina and convey visual information to the brain.

Based on the ganglion cell layer, the retina of the human eyes can be divided into two main parts, namely macula and periphery [24], as illustrated in Fig. 3. In particular, the ganglion cell layer in the macula is several cells thick. Meanwhile, the periphery is only one ganglion cell thick. The macula is further divided into three regions, called fovea, parafovea, and perifovea. The periphery is in turn divided into two regions, namely near periphery and far periphery [24], [25]. These five regions of the retina are briefly described below. It is worth noting that there has been no standard definition of boundaries between these regions so far [26]. In our research, the boundaries are determined based on [26]–[29].

The fovea is a small central region of the macula that represents 5 degrees of the central visual field or an eccentricity interval between 0 degree and 2.5 degrees. This region consists of densely packed cones. In addition, it has a layer of ganglion cells, which can be up to eight cells thick. Therefore, the fovea vision has the highest sensitivity to fine details.

The fovea is surrounded by the parafovea belt corresponding to an eccentricity interval between 2.5 degrees and 4 degrees. In the parafovea, rods are more numerous. Meanwhile, the thickness of the ganglion cell layer decreases from eight to four cells at its outer edge [25].

The region next to the parafovea is the perifovea with the corresponding eccentricity interval between 4 degrees and 9 degrees. In this region, the density of rods is higher than that of cones. The thickness of ganglion cell layer reduces to one cell at its peripheral edge [25].

In the periphery, the region corresponding to an eccentricity interval between 9 degrees and 30 degrees is the near periphery, and the rest is the far periphery. The dividing line corresponding to the eccentricity of 30 degrees is selected based on several features of visual performance. In particular, letter visual acuity decreases linearly with eccentricity from 0 degree to 30 degrees. For eccentricities larger than 30 degrees, the decrease is much steeper [9].

Based on the above description of the viewing geometry and the retina, stimuli used in the following subjective experiment are designed so that the zones in the virtual viewports will correspond to the five regions of the retina. It is worth noting that, in this paper, we focus on the contributions (or weights) of different zones in the perceptual quality, rather than the quality-reducing trends as in [7], [11], [12].

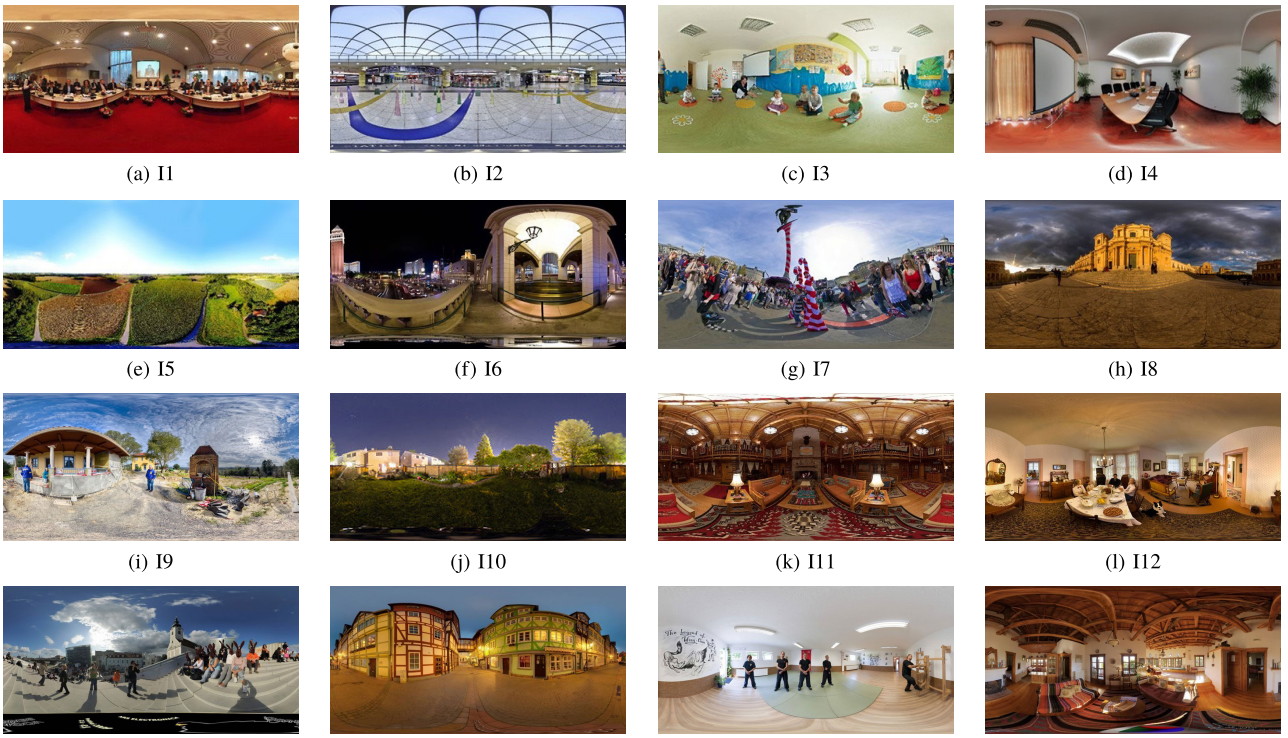
III. EXPERIMENT DESCRIPTION

For the experiment, we used sixteen omnidirectional images, denoted by *I1*–*I16*, as shown in Fig. 4. Two images *I5* and *I7* were obtained on Flickr under of the Creative Commons (CC) copyrights. Image *I11* was from the Salience360 Dataset [30], [31]. The other images were selected from the SUN 360 Database [32], [33]. The characteristics of these images are described in Table 1. It can be seen that the selected images cover various categories of capturing environment and presence of human. All these images were down sampled to the resolution of 8192×4096. We asked 10 participants to freely observe the source images and then point out attractive objects. Based on the obtained results, we selected a foveation point corresponding to a viewport for each image.

In order to generate stimuli of non-uniform quality, each image was first spatially divided into five zones, denoted Z_1, Z_2, Z_3, Z_4, and Z_5. In particular, each zone represents an eccentricity interval as shown in Table 2. It can be seen that zones Z_1, Z_2, Z_3, Z_4, and Z_5 respectively correspond to the fovea, parafovea, perifovea, near periphery, and far periphery in the retina. Fig. 5 illustrates the boundaries of the zones in the viewports used in our experiment.

As described in Section I, we consider two basic scenarios of spatial quality changes. In the first scenario (S#1), the center has higher quality than the periphery; and in the second scenario (S#2), the center has lower quality than the periphery. For each scenario, we used four quality variation patterns as shown in Table 3. In patterns *P1*, *P2*, *P3*, and *P4*, which belong to scenario S#1, the number of high quality zones gradually increases from 1 to 4. In the remaining patterns (i.e., *P5*, *P6*, *P7*, and *P8*), which belong to scenario S#2, the number of high quality zones gradually reduces from 4 to 1.

In this study, we used one high quality level corresponding to the quality level of the source images, and four low quality levels corresponding to four blurring levels. These blurring levels were generated using Gaussian filters with a fixed filter size of 50 and four different standard deviations σ. For scenario S#1, the four σ values are 2, 4, 8, and 12. For scenarios S#2, the four σ values are 1, 2, 4, and 6. The difference between the two scenarios is due to the fact that blurring in zones close to the foveation point is easier to be perceived than in the others. The source and blurred images were then blended into stimuli of non-uniform quality. Specifically, the high quality zones in the stimuli consist of pixels of the source images, and the low quality zones are comprised



(a) I1 (b) I2 (c) I3 (d) I4

(e) I5 (f) I6 (g) I7 (h) I8

(i) I9 (j) I10 (k) I11 (l) I12

(m) I13 (n) I14 (o) I15 (p) I16

FIGURE 4. Sixteen omnidirectional images used in our experiment.

TABLE 1. Features of source images.

Image	Description
I1	indoor scene, large conference room, containing human faces
I2	indoor scene, train station in Japan, containing human faces
I3	indoor scene, small kindergarten classroom, containing human faces
I4	indoor scene, meeting room, without presence of human
I5	outdoor scene, natural landscape, daytime, without presence of human
I6	outdoor scene, balcony, nighttime, without presence of human
I7	outdoor scene, festival, daytime, containing human faces
I8	outdoor scene, outside of a cathedral, at sunset, containing human faces
I9	outdoor scene, small farm, daytime containing human faces
I10	outdoor scene, large green garden, nighttime without presence of human
I11	indoor scene, large wooden living room, without presence of human
I12	indoor scene, dinner table, containing human faces
I13	outdoor scene, large square, daytime, containing human faces
I14	outdoor scene, quiet street, nighttime, without presence of human
I15	indoor scene, gym room, containing human faces
I16	indoor scene, living room, without presence of human

TABLE 2. Eccentricity intervals of zones.

Zone	Z_1	Z_2	Z_3	Z_4	Z_5
Eccentricity interval (degrees)	[0,2.5)	[2.5,4)	[4,9)	[9,30)	[30,+∞)

of pixels of the blurred images. Similar to [12], to prevent noticeable boundaries between low and high quality zones, belts with the width of 5 degrees between two adjacent zones having a quality switch were used as transition belts. The quality levels in these belts smoothly change using a linear function. Totally, our database consists of 512 stimuli, which were rated in the below tests.

To display the stimuli, we used a device set of a Samsung Galaxy S6 smartphone and a Samsung Gear VR headset with the 96 degree field of view. The Samsung Galaxy S6 has the screen resolution of 2560×1440 and the display size of 5.1 inches. For the Samsung Gear VR headset, the focal length of the lens is $F = 62$mm, and the distances from the lens to the displayed viewports and the eyes are approximately $S_0 = 25$mm and $S_2 = 10$mm respectively.

In the tests, we used the Absolute Category Rating method [34], which is shown the best method in [12]. Before doing actual tests, participants were trained to get accustomed to the devices and the rating procedure. In addition, they were instructed to appropriately adjust devices to obtain the best experience. During the test process, the stimuli were randomly displayed one at a time. Note that, for a stimulus, the corresponding viewport displayed on HMD was fixed during the test. Participants were asked to look straight ahead at each viewport displayed directly in front of them to keep focusing on the center, where has an attractive object such as a human face or a flower vase. After stabilizing the gaze direction, each participant verbally gave a score with the grade scale from 1 (bad) to 5 (excellent) which was recorded by an assistant.

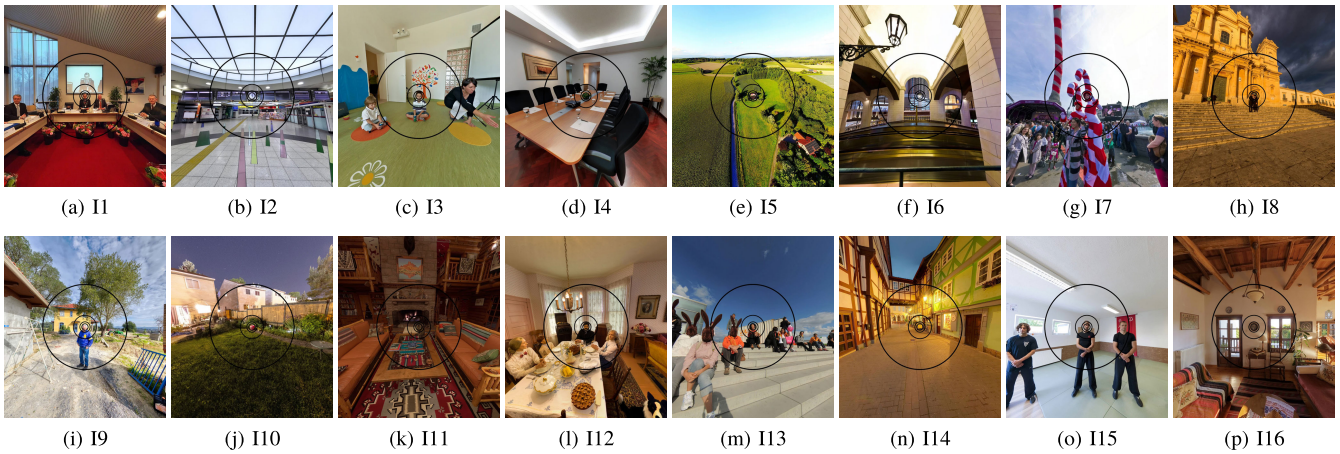


FIGURE 5. Boundaries of zones in viewports used in our experiment.

TABLE 3. Quality variation patterns (HQ: High quality and LQ: Low quality).

Scenario	Pattern	Quality levels of zones				
		Z_1 [0°,2.5°)	Z_2 [2.5°,4°)	Z_3 [4°,9°)	Z_4 [9°,30°)	Z_5 [30°,+∞)
S#1	*P1*	HQ	LQ	LQ	LQ	LQ
	P2	HQ	HQ	LQ	LQ	LQ
	P3	HQ	HQ	HQ	LQ	LQ
	P4	HQ	HQ	HQ	HQ	LQ
S#2	*P5*	LQ	HQ	HQ	HQ	HQ
	P6	LQ	LQ	HQ	HQ	HQ
	P7	LQ	LQ	LQ	HQ	HQ
	P8	LQ	LQ	LQ	LQ	HQ

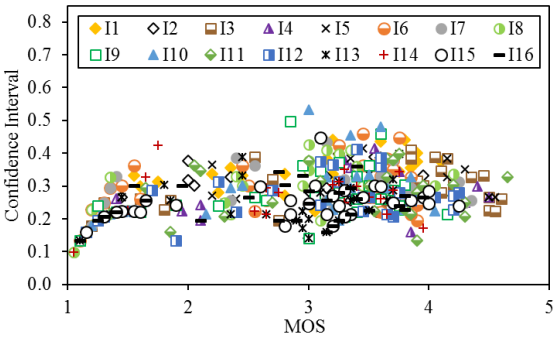


FIGURE 6. 95% confidence intervals of MOS values.

For each stimulus, the viewing duration was decided by the participants themselves to obtain more reliable rating scores. Commonly, the participants spent about 5 seconds for rating a stimulus and then took a break of 5 seconds. To avoid the negative impacts of fatigue and boredom, the tests were divided into 12 sessions conducted in different weeks. Each participant took part in only two sessions. The duration of each session was no more than 10 minutes. There were totally 125 participants between the ages of 20 and 30. A screening analysis of the obtained results was performed following Recommendation ITU-T P.913 [34], and five participants were rejected. After discarding the scores of these five participants, each stimulus was scored by 20 valid participants. The mean opinion score (MOS) of a stimulus is the average score of the valid participants.

The 95% confidence intervals of the MOS values are shown in Fig. 6. We can see that the scores cover fully the value range from 1 to nearly 5. Generally, the confidence intervals are smaller at the two ends of the grade scale. This is because the participants are more confident in rating stimuli of very high (or low) quality.

IV. ANALYSIS OF PERCEPTUAL BEHAVIORS IN ZONES

A. QUANTIFYING IMPACTS OF ZONES

In this part, we present a zone-weighted formulation which will be used to analyze the impacts of different zones on the perceptual quality of omnidirectional images. In general, the virtual viewport is divided into K zones $\{Z_k | 1 \leq k \leq K\}$, each consists of N_k pixels with the corresponding eccentricities $e \in [e_{k-1}, e_k)$. Currently, we use $K = 5$ as described in Section III. Each zone Z_k is then assigned a weight $\{w_k | 1 \leq k \leq K\}$ representing the impact of that zone on human perception of quality. Note that $\sum_{k=1}^{K} w_k = 1$.

Let $V(x_M, y_M)$ and $G(x_M, y_M)$ respectively be the values of pixel $M = (x_M, y_M)$ in the displayed viewports of the original and distorted images. The values of the corresponding pixel $M' = (x_{M'}, y_{M'})$ in the virtual viewports of the original and distorted images are respectively calculated by the following equations.

$$V'(x_{M'}, y_{M'}) = V(x_M, y_M). \tag{13}$$

$$G'(x_{M'}, y_{M'}) = G(x_M, y_M). \tag{14}$$

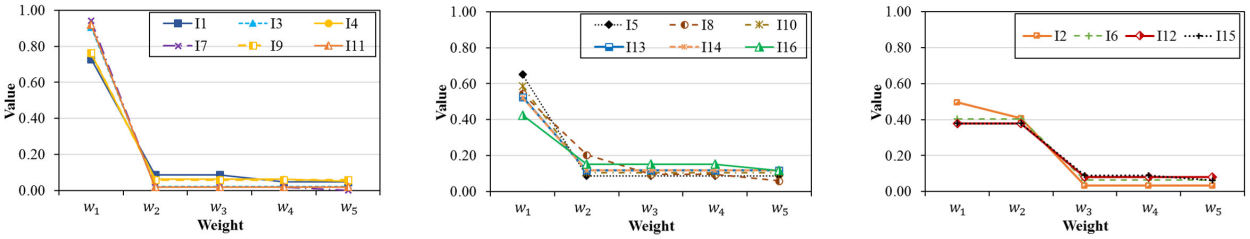
The mean squared error (MSE) of pixels in zone Z_k is computed by (15), as shown at the bottom of the next page.

The zone-weighted formulation, called *ZWF*, is given by

$$ZWF = 10 \log_{10} \left(\frac{MAX^2}{\sum_{k=1}^{K} (w_k \times MSE_k)} \right), \tag{17}$$

TABLE 4. Performance of fitting between the *ZWF* formulation and MOS.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16
PCC	0.988	0.992	0.992	0.969	0.980	0.995	0.981	0.992	0.987	0.988	0.986	0.977	0.977	0.991	0.985	0.976
SROCC	0.972	0.981	0.982	0.961	0.966	0.968	0.983	0.987	0.967	0.969	0.972	0.955	0.914	0.939	0.957	0.961
RMSE	0.146	0.131	0.144	0.274	0.236	0.100	0.201	0.118	0.149	0.153	0.160	0.211	0.168	0.118	0.151	0.175



(a) I1, I3, I4, I7, I9, and I11 (b) I5, I8, I10, I13, I14, and I16 (c) I2, I6, I12, and I15

FIGURE 7. Weights of zones for each source image.

where *MAX* is the maximum possible pixel value. Here we set *MAX* to 255 as the bit depth of pixels is 8 bits in our experiment.

In some previous studies [22], [35], it was shown that four-parameter and five-parameter logistic functions are good mappings between objective quality metrics and MOS. In this work, we employed the following five-parameter logistic function to map the *ZWF* values and the MOS values in our database.

$$y = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5, \tag{18}$$

where $\{\beta_i | i \in \{1, 2, \ldots, 5\}\}$ are parameters to be fitted. The values of the parameters β_i's and the weights w_k's were determined by means of least squares fitting as in [36].

TABLE 5. Weights of zones for each source image.

Image	Weight				
	w1	*w2*	*w3*	*w4*	*w5*
I1	0.728	0.088	0.088	0.048	0.048
I2	0.495	0.407	0.033	0.033	0.032
I3	0.905	0.024	0.024	0.024	0.024
I4	0.759	0.063	0.063	0.063	0.052
I5	0.650	0.087	0.087	0.087	0.087
I6	0.404	0.404	0.064	0.064	0.064
I7	0.941	0.019	0.019	0.019	0.003
I8	0.545	0.204	0.095	0.095	0.061
I9	0.763	0.059	0.059	0.059	0.059
I10	0.587	0.103	0.103	0.103	0.103
I11	0.919	0.020	0.020	0.020	0.020
I12	0.379	0.377	0.081	0.081	0.081
I13	0.526	0.119	0.118	0.118	0.118
I14	0.523	0.119	0.119	0.119	0.119
I15	0.379	0.379	0.090	0.090	0.062
I16	0.426	0.152	0.152	0.152	0.117

B. DISCUSSION

To quantify the impact of each zone taking into account the effects of content characteristics, the weights w_k's are derived for each source image by fitting using the above five-parameter logistic function with the stimuli of that image only. The obtained values of the weights are shown in Fig. 7 and Table 5. The correlation coefficients including Pearson Correlation Coefficient (PCC), Spearman's rank ordered correlation coefficient (SROCC), and Root Mean Square Error (RMSE), which are used to quantify the performance of the fitting between the *ZWF* formulation and the MOS, are shown in Table 4. We can see that, for all the source images, the PCC and SROCC values are very high and the RMSE values are very low. In particular, the lowest PCC and SROCC values are respectively 0.969 and 0.914 while the highest RMSE value is 0.274. This means that the fitting to obtain the weights is reliable.

From Table 5, it can be seen that, except w_1 and w_2, all the other weights are small (i.e., ≤ 0.152). That means the zones outside the eccentricity of 4 degrees have little impacts on the perceptual quality. Among the weights, w_1 is usually highest,

$$MSE_k = \frac{\sum_{x_{M'}=1}^{W'_p} \sum_{y_{M'}=1}^{H'_p} [V'(x_{M'}, y_{M'}) - G'(x_{M'}, y_{M'})]^2 \times R_k(x_{M'}, y_{M'})}{\sum_{x_{M'}=1}^{W'_p} \sum_{y_{M'}=1}^{H'_p} R_k(x_{M'}, y_{M'})}, \tag{15}$$

where

$$R_k(x_{M'}, y_{M'}) = \begin{cases} 1, & \text{if } e_{k-1} \leq e(x_{M'}, y_{M'}) < e_k \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

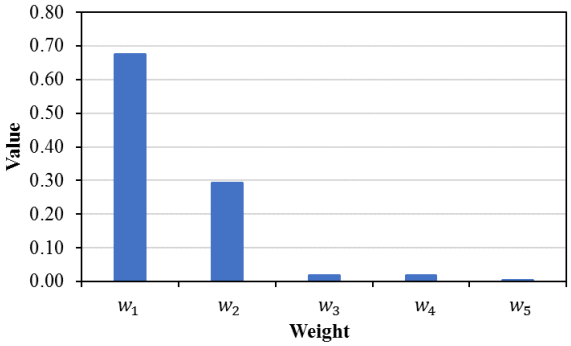


FIGURE 8. Weights of zones for all source images.

which is consistent with the fact that the fovea region of the retina has the highest cone density. Also, because $w_1 \geq w_2 \geq w_3 \geq w_4 \geq w_5$, distortions closer to the center have more significant effects on the perceptual quality than distortions far from the center.

Based on Fig. 7, it is interesting that the value of w_1 actually varies in a wide range. Also, with some images, the value of w_2 is insignificant. More specifically, with images *I1*, *I3*, *I4*, *I7*, *I9*, and *I11*, the values of w_1 are very high (i.e., $\geq$ 0.728). This may be because the participants focus primarily on attractive objects such as small faces at the center of the viewports. Such phenomenon was also observed in [11]. In particular, it was found that a talking face is strongly attractive to human attention [11]. In addition, it can be seen that, in the viewports of images *I3*, *I7*, and *I11*, there are no other interesting objects near the center. Meanwhile, with images *I1*, *I4*, and *I9*, the participants may also pay some attention to other objects near the center (e.g., another face in image *I1*), so the values of w_1 are slightly lower than those of images *I3*, *I7*, and *I11*.

With images *I5*, *I8*, *I10*, *I13*, *I14*, and *I16*, the center's object is not very clear (e.g., small faces in image *I8*) or not very attractive (e.g., a house in image *I5*), resulting in lower values of w_1. Especially, with images *I2*, *I6*, *I12*, and *I15*, the values of w_2 are comparable to those of w_1. In these images, the participants may look at a large central area rather than zone Z_1 only. The reason is that, in image *I2*, *I12*, and *I15*, the objects at the center such as a lock in image *I2* are larger than zone Z_1; and in image *I6*, the object at the center does not stand out from the neighboring area.

From the above, we can see that the perceptual quality is affected by two key factors. The first is the sensitivity of the human eyes. Especially, in the considered context, zones Z_1 and Z_2 are much more important than the other zones. The second is content characteristics. In particular, the values of w_1 and w_2 vary widely according to 1) the attractiveness and 2) the size of the central object, as well as 3) the presence of neighboring objects.

In order to better understand, the weights are obtained by fitting using the stimuli of all the images as shown in Fig. 8. It can be seen that the behavior of the weights are similar to that derived for each image. In particular, w_1 is highest. w_3, w_4, and w_5 are quite small (i.e., $\leq$0.017). The correlation coefficients are respectively 0.836 for PCC, 0.800 for SRCC, and 0.546 for RMSE. Obviously, the performance reduces in comparison to fitting for each image. This result again suggests that there is a significant impact of content characteristics on the perceptual quality.

V. EVALUATION OF QUALITY METRICS

In this part, by using our database, we evaluate the performances of twenty-five existing objective quality metrics (OQM). The goal is to examine whether existing metrics, especially foveal quality metrics, are effective for quality assessments of omnidirectional images with non-uniform quality.

A. DESCRIPTION OF METRICS

Table 6 shows the notations and descriptions of the twenty-five metrics considered in this study. In this table, the PW column indicates whether a metric differentiates the contributions of different pixels; and the FF column indicates whether a metric takes into account the foveation feature of the human eye. Because the implementations of the FWQI, FWSNR, FPSNR, and F-SSIM metrics are not publicly available, we implemented them based on the corresponding publications [20], [21], [47], [48]. For the remaining metrics, we used the implementations provided by the original authors.

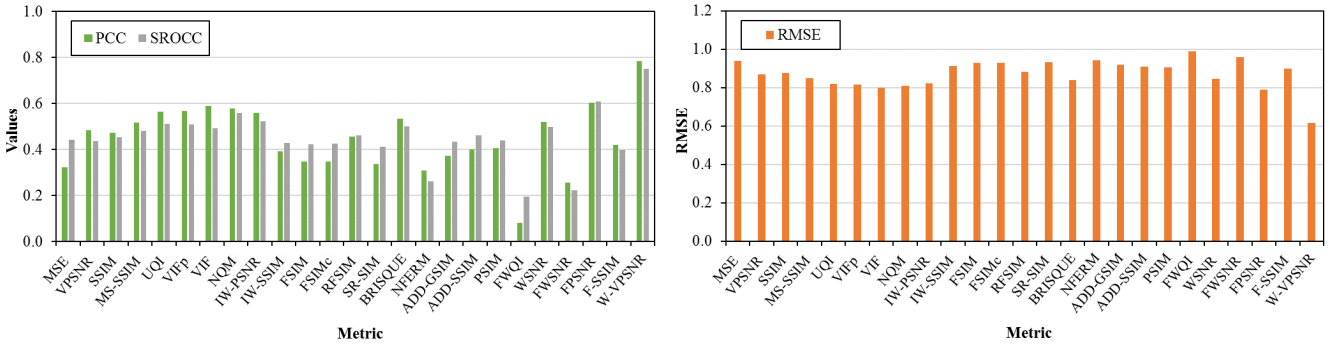
It is worth noting that, except the W-VPSNR metric, all of the other metrics were proposed to calculate for all pixels in a traditional image. In this study, these metrics and the W-VPSNR metric were calculated for viewports only (i.e., visible pixels) of the omnidirectional images to reflect what is actually watched by viewers. To extract the viewports, we used 360Lib software developed by Joint Video Experts Team (JVET) [50]. In addition, geometric parameters in these metrics were calculated based on the equations presented in Subsection II-A.

Obviously, (17) can be used to build an objective quality metric. In our recent study [49], an objective quality metric (called W-VPSNR) is proposed based on this formulation. In this metric, the weights w_k's are obtained by curve-fitting using the MOS values of training stimuli generated from three source images. Three other images are used to generate test stimuli. The selection of the training and test images among six source images is repeated 20 times. The weights w_k's corresponding to the selection having the highest PCC value is recommended to use in the W-VPSNR metric. Note that, since the three images *I1, I2,* and *I6* are used in [49] to obtain the recommended weights, these three images are not used in this part for fairness in the performance evaluation of the metrics.

In order to evaluate the performances of the OQM metrics, we used three performance metrics of Pearson Correlation Coefficient (PCC), Spearman's rank ordered correlation coefficient (SROCC), and Root Mean Square Error (RMSE). Similar to [35], a nonlinear regression was applied to map the OQM values to the MOS values using the

TABLE 6. Descriptions of objective quality metrics tested in this study. PW: Whether or not the metric differentiates pixels' contributions. FF: Whether or not the metric takes into account the foveation feature.

Metrics	PW	FF	Description
MSE	No	No	Mean Squared Error, Calculated based on visble pixels of a viewport with equal weights
VPSNR	No	No	Viewport-PSNR, Calculated based on visble pixels of a viewport with equal weights
SSIM [17]	No	No	Structural SIMilarity, Calculated based on the concept of structural similarity
MS-SSIM [18]	No	No	Multi-scale SSIM, Calculated based on similar measures computed at different resolutions (or multi-scales) of a viewport
UQI [19]	No	No	Universal Image Quality, Modeling any distortion as a combination of three different factors including loss of correlation, luminance distortion, and contrast distortion
VIFp [37]	No	No	Visual Information Fidelity in the pixel domain (*VIFp*) and the wavelet domain (*VIF*), Calculated based on the connections between image information and visual quality
VIF [37]	No	No	
NQM [38]	No	No	Noise Quality Measure, Signal-to-Noise Ratio of the restored distorted image with respect to the model restored image
IW-PSNR [39]	Yes	No	Information content Weighted PSNR, Combining information content weighting with PSNR measures
IW-SSIM [39]	Yes	No	Information content Weighted SSIM, Combining information content weighting with MS-SSIM measures
FSIM [40]	Yes	No	Feature similarity, Combining low-level feature weighting with local similarity measures
FSIMc [40]	Yes	No	Feature similarity incorporating the chromatic information, Combining low-level feature weighting with local similarity measures
RFSIM [41]	Yes	No	Riesz Transforms based Feature Similarity, Combining low-level feature weighting based on Riesz Transforms with local similarity measures
SR-SIM [42]	Yes	No	Spectral Residual based Similarity, Calculated based on a spectral residual visual saliency model
BRISQUE [43]	Yes	No	Blind/Referenceless Image Spatial QUality Evaluator, Using scene statistics of locally normalized luminance coefficients to quantify possible losses of "naturalness" in the image due to the presence of distortions
NFERM [44]	Yes	No	No-reference Free Energy-Based Robust Metric, Using the free-energy-based brain theory and classical human visual system (HVS)-inspired features
ADD-SSIM [45]	Yes	No	Analysis of Distortion Distribution-based (ADD) ADD-SSIM and Analysis of Distortion Distribution-based (ADD) Gradient SIMilarity index (GSIM), Taking into account the distribution of distortion position, distortion intensity,frequency changes, and histogram alteration
ADD-GSIM [45]	Yes	No	
PSIM [46]	Yes	No	Perceptual SIMilarity (PSIM), Combining the gradient magnitude similarities at two scales, the color information similarity, and a reliable perceptual-based pooling
FWQI [47]	Yes	Yes	Foveated Wavelet image Quality Index (FWQI), Calculated based on wavelet coefficients in the discrete wavelet transform domain using the foveation-based error sensitivity model as a weighting function.
WSNR [38]	Yes	Yes	Weighted Signal-to-Noise Ratio, the ratio of the average weighted signal power to the average weighted noise power, where the weighting function is the contrast sensitivity function
FWSNR [20]	Yes	Yes	Foveal Weighted Signal-to-Noise Ratio, Combining weighting for each pixel by the local frequency at that pixel with WSNR measures
FPSNR [48]	Yes	Yes	Foveal Peak Signal-to-Noise Ratio, Combining weighting for each pixel by the local frequency at that pixel with PSNR measures
F-SSIM [21]	Yes	Yes	Foveal-SSIM, Combining weighting for each macroblock based on the local frequency of pixels in that macroblock with SSIM measures
W-VPSNR [49]	Yes	Yes	Weighted Viewport PSNR, Calculated based on a weighted sum of image zones' distortions



(a) PCC and SROCC (b) RMSE

FIGURE 9. Performances of objective quality metrics calculated with all the stimuli (except *I1, I2,* and *I6*).

five-parameter logistic function (i.e., (18)) mentioned in Subsection IV-A.

B. DISCUSSION

Fig. 9 and the last columns of Table 7, Table 8, and Table 9 show the PCC, SROCC, and RMSE values of the OQM metrics when fitting with all the MOSs of the stimuli. It can be seen that all the metrics have low performances (i.e., PCC $\leq$ 0.784, SROCC $\leq$ 0.751, and RMSE $\geq$ 0.616). Even the foveal quality metrics (namely FWQI, WSNR, FWSNR, FPSNR, F-SSIM, and W-VPSNR) have low PCC values (i.e., from 0.082 to 0.784). This means that the investigated metrics are not very effective to assess the perceptual quality of omnidirectional images with non-uniform quality.

Similar to the previous analysis related to the *ZWF* formulation, it is important to understand the performances of the metrics for each source image. Table 7, Table 8, and Table 9 show the performances of the metrics when fitting with the stimuli of each source image. It can be seen that, for most of the metrics, the PCC and RMSE values are drastically variable across different source images. The bold numbers show the metrics having the highest performance for each source image.

Among the investigated metrics, the W-VPSNR metric has the highest PCC and SROCC values and the lowest RMSE values for nine source images (i.e., *I3, I4*, *I5, I7*, *I8*, *I9, I10, I11,* and *I12*). In addition, its performance for the remaining images is quite good (i.e., PCC $\geq$ 0.947, SROCC $\geq$ 0.881, and RMSE $\leq$ 0.259). This result means that the W-VPSNR metric is rather effective to compare the perceptual quality between stimuli of the same source image.

Regarding the FPSNR metric, its performance is quite good for eight source images *I5, I8, I10, I12, I13, I14, I15,*

TABLE 7. PCC values of metrics calculated with all the stimuli and with the stimuli of each source image (except *I1, I2,* and *I6*). The bold numbers show the metrics having the highest performance for each source image.

Metrics	*I3*	*I4*	*I5*	*I7*	*I8*	*I9*	*I10*	*I11*	*I12*	*I13*	*I14*	*I15*	*I16*	*All*
MSE	0.758	0.635	0.675	0.609	0.695	0.561	0.594	0.736	0.515	0.389	0.572	0.744	0.609	0.324
VPSNR	0.629	0.568	0.706	0.548	0.563	0.530	0.605	0.495	0.619	0.495	0.633	0.487	0.425	0.484
SSIM [17]	0.433	0.426	0.530	0.382	0.377	0.370	0.437	0.272	0.340	0.267	0.357	0.303	0.302	0.473
MS-SSIM [18]	0.520	0.489	0.600	0.434	0.449	0.436	0.514	0.325	0.378	0.302	0.419	0.396	0.549	0.519
UQI [19]	0.722	0.644	0.675	0.544	0.585	0.534	0.652	0.469	0.527	0.477	0.552	0.568	0.463	0.564
VIFp [37]	0.707	0.640	0.698	0.612	0.607	0.580	0.667	0.527	0.544	0.489	0.568	0.586	0.510	0.568
VIF [37]	0.713	0.660	0.704	0.636	0.624	0.579	0.657	0.531	0.544	0.501	0.557	0.588	0.520	0.591
NQM [38]	0.727	0.703	0.690	0.695	0.686	0.598	0.653	0.579	0.512	0.582	0.494	0.651	0.550	0.579
IW-PSNR [39]	0.533	0.515	0.585	0.490	0.497	0.405	0.612	0.392	0.393	0.498	0.542	0.428	0.391	0.561
IW-SSIM [39]	0.487	0.477	0.553	0.434	0.430	0.409	0.470	0.323	0.366	0.296	0.399	0.346	0.343	0.394
FSIM [40]	0.430	0.433	0.475	0.385	0.382	0.350	0.406	0.279	0.331	0.269	0.367	0.288	0.286	0.347
FSIMc [40]	0.430	0.433	0.476	0.385	0.382	0.350	0.406	0.279	0.331	0.269	0.367	0.287	0.286	0.347
RFSIM [41]	0.549	0.506	0.621	0.505	0.517	0.475	0.493	0.403	0.431	0.368	0.453	0.361	0.395	0.456
SR-SIM [42]	0.400	0.423	0.449	0.346	0.387	0.331	0.401	0.278	0.346	0.248	0.356	0.288	0.296	0.338
BRISQUE [43]	0.831	0.806	0.269	0.748	0.876	0.478	0.546	0.787	0.758	0.720	0.716	0.749	0.711	0.536
NFERM [44]	0.852	0.716	0.530	0.720	0.825	0.349	0.518	0.848	0.770	0.576	0.705	0.775	0.757	0.310
ADD-SSIM [45]	0.463	0.630	0.571	0.580	0.415	0.391	0.487	0.317	0.386	0.276	0.364	0.372	0.367	0.374
ADD-GSIM [45]	0.520	0.592	0.573	0.615	0.468	0.420	0.506	0.346	0.362	0.314	0.434	0.385	0.386	0.401
PSIM [46]	0.492	0.477	0.584	0.442	0.456	0.431	0.490	0.349	0.386	0.315	0.416	0.345	0.357	0.407
FWQI [47]	0.013	0.088	0.105	0.004	0.127	0.131	0.151	0.020	0.111	0.147	0.299	0.108	0.068	0.082
WSNR [38]	0.530	0.513	0.621	0.493	0.495	0.632	0.495	0.390	0.585	0.485	0.371	0.639	0.401	0.520
FWSNR [20]	0.600	0.568	0.593	0.558	0.481	0.662	0.617	0.691	0.430	0.396	0.510	0.538	0.425	0.256
FPSNR [48]	0.610	0.842	0.903	0.500	0.944	0.615	0.920	0.792	0.874	0.949	0.980	0.976	**0.984**	0.604
F-SSIM [21]	0.530	0.412	0.515	0.401	0.380	0.364	0.403	0.273	0.314	0.254	0.344	0.285	0.293	0.422
W-VPSNR [49]	**0.982**	**0.982**	**0.980**	**0.932**	**0.982**	**0.989**	**0.985**	**0.977**	**0.973**	**0.968**	**0.984**	**0.979**	0.947	**0.784**

TABLE 8. SROCC values of metrics calculated with all the stimuli and with the stimuli of each source image (except *I1, I2,* and *I6*). The bold numbers show the metrics having the highest performance for each source image.

Metrics	*I3*	*I4*	*I5*	*I7*	*I8*	*I9*	*I10*	*I11*	*I12*	*I13*	*I14*	*I15*	*I16*	*All*
MSE	0.759	0.512	0.458	0.575	0.735	0.466	0.520	0.836	0.412	0.242	0.547	0.828	0.669	0.442
VPSNR	0.616	0.424	0.576	0.475	0.474	0.536	0.622	0.474	0.601	0.386	0.604	0.369	0.290	0.438
SSIM [17]	0.523	0.440	0.467	0.492	0.439	0.404	0.452	0.279	0.286	0.218	0.247	0.331	0.302	0.454
MS-SSIM [18]	0.553	0.428	0.462	0.493	0.471	0.417	0.470	0.285	0.298	0.212	0.277	0.374	0.483	0.481
UQI [19]	0.590	0.486	0.458	0.532	0.485	0.429	0.479	0.336	0.365	0.293	0.287	0.386	0.348	0.512
VIFp [37]	0.588	0.476	0.459	0.530	0.484	0.446	0.493	0.351	0.365	0.314	0.305	0.373	0.330	0.510
VIF [37]	0.588	0.474	0.451	0.533	0.483	0.441	0.487	0.345	0.364	0.314	0.303	0.371	0.322	0.494
NQM [38]	0.679	0.545	0.571	0.637	0.629	0.573	0.577	0.513	0.397	0.558	0.336	0.647	0.527	0.561
IW-PSNR [39]	0.551	0.465	0.422	0.544	0.505	0.467	0.556	0.376	0.334	0.367	0.452	0.429	0.403	0.524
IW-SSIM [39]	0.522	0.423	0.466	0.479	0.441	0.389	0.438	0.285	0.276	0.194	0.240	0.346	0.304	0.428
FSIM [40]	0.524	0.443	0.454	0.490	0.448	0.392	0.442	0.286	0.285	0.226	0.254	0.333	0.284	0.424
FSIMc [40]	0.524	0.443	0.454	0.490	0.448	0.402	0.442	0.286	0.285	0.226	0.254	0.333	0.284	0.425
RFSIM [41]	0.529	0.433	0.457	0.494	0.473	0.424	0.432	0.338	0.314	0.237	0.280	0.329	0.316	0.463
SR-SIM [42]	0.508	0.401	0.437	0.463	0.450	0.403	0.416	0.269	0.271	0.199	0.252	0.352	0.303	0.412
BRISQUE [43]	0.820	0.821	0.294	0.724	0.925	0.217	0.522	0.794	0.677	0.729	0.708	0.757	0.772	0.501
NFERM [44]	0.871	0.635	0.518	0.725	0.799	0.350	0.541	0.895	0.720	0.518	0.756	0.811	0.772	0.261
ADD-SSIM [45]	0.522	0.469	0.427	0.566	0.416	0.371	0.449	0.288	0.290	0.167	0.212	0.361	0.316	0.436
ADD-GSIM [45]	0.564	0.486	0.437	0.586	0.480	0.371	0.452	0.305	0.298	0.215	0.279	0.368	0.316	0.462
PSIM [46]	0.516	0.420	0.489	0.486	0.453	0.414	0.420	0.285	0.276	0.220	0.254	0.321	0.289	0.440
FWQI [47]	−0.098	−0.156	0.323	0.129	0.241	0.231	0.296	0.152	0.355	0.421	0.555	0.271	0.268	0.196
WSNR [38]	0.533	0.402	0.538	0.481	0.507	0.551	0.440	0.372	0.493	0.341	0.236	0.550	0.412	0.497
FWSNR [20]	0.497	0.447	0.550	0.484	0.371	0.604	0.467	0.740	0.308	0.284	0.470	0.379	0.261	0.223
FPSNR [48]	0.608	0.896	0.933	0.458	0.944	0.628	0.932	0.793	0.892	**0.916**	**0.953**	**0.957**	**0.968**	0.609
F-SSIM [21]	0.531	0.381	0.419	0.451	0.397	0.353	0.397	0.230	0.240	0.164	0.196	0.280	0.251	0.399
W-VPSNR [49]	**0.965**	**0.968**	**0.966**	**0.930**	**0.975**	**0.974**	**0.963**	**0.966**	**0.927**	0.881	0.881	0.934	0.905	**0.751**

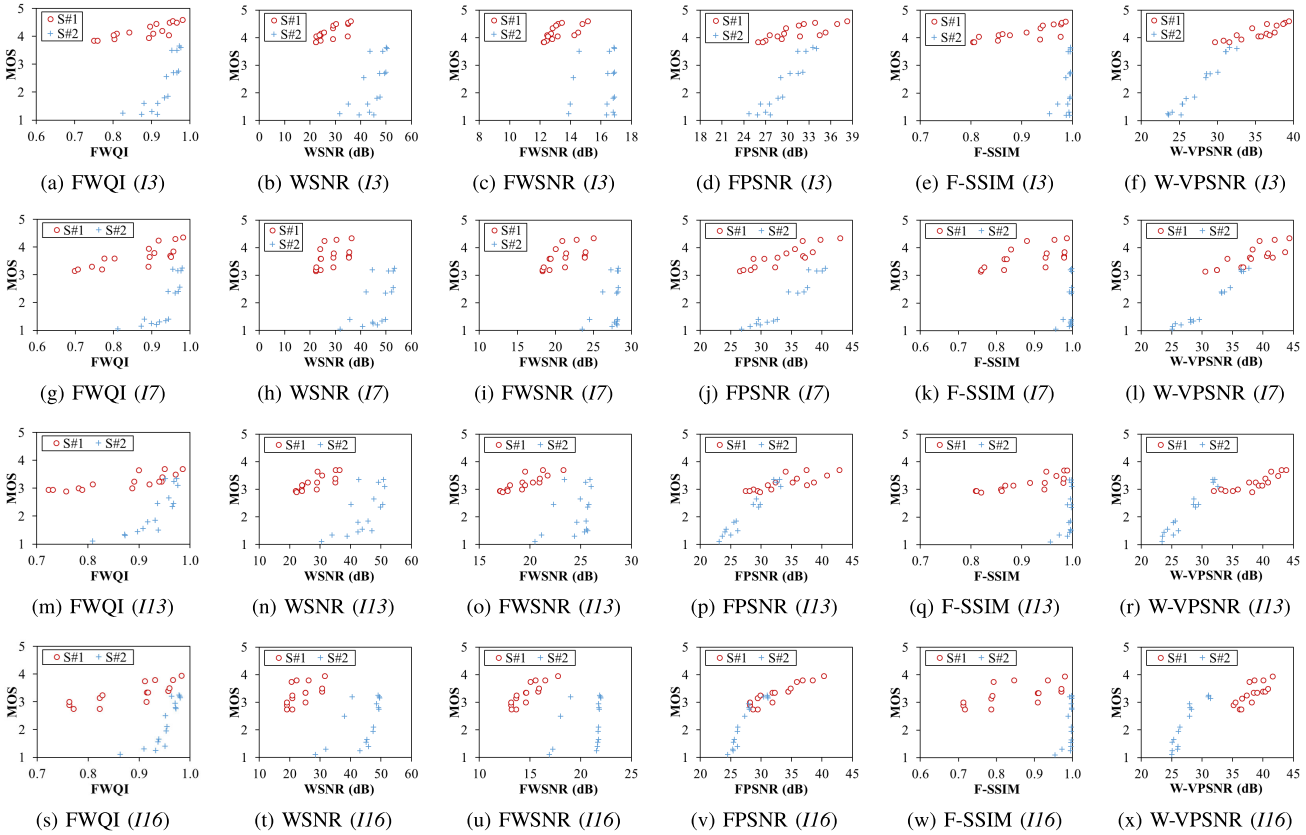
and *I16* (i.e., PCC $\geq$ 0.874, SROCC $\geq$ 0.892, and RMSE $\leq$ 0.508). Especially, with images *I13, I14, I15* and *I16*, its SROCC values are higher than those of the W-VPSNR metric (i.e., 0.916 vs. 0.881, 0.953 vs. 0.881, 0.957 vs. 0.934, and 0.968 vs. 0.905). However, its performance is very low for two images *I3* and *I7* (i.e., PCC $<$ 0.70, SROCC $<$ 0.70, and RMSE $>$ 0.90), even lower than that of the MSE metric (i.e., PCC: 0.610 vs. 0.758 and 0.500 vs. 0.609; SROCC: 0.608 vs. 0.759 and 0.458 vs. 0.575; RMSE: 0.917 vs. 0.754 and 0.903 vs. 0.827), which is the simplest metric in practice.

As for the other quality metrics, their performances are mostly low. Even the other foveal quality metrics (i.e., except the FPSNRand W-VPSNR metrics) have lower performances than the non-foveal and simple metrics for some source images.

To understand the actual behaviors of the foveal quality metrics that cause low performances, Fig. 10 shows the scatter plots of the values of these metrics versus the MOS values for images *I3, I7*, *I13,* and *I16*. In this figure, we use different legends to differentiate the stimuli of scenario *S*#1, where the center has higher quality, and the stimuli of scenario *S*#2, where the center has lower quality. It is well-known that higher values of these metrics mean higher MOS values and better perceptual quality. From Fig. 10, we can see that the MOS values in scenario *S*#1 are generally higher than those in scenario *S*#2. However, for the WSNR, FWSNR, and F-SSIM

TABLE 9. RMSE values of metrics calculated with all the stimuli and with the stimuli of each source image (except *I1*, *I2*, and *I6*). The bold numbers show the metrics having the highest performance for each source image.

Metrics	*I3*	*I4*	*I5*	*I7*	*I8*	*I9*	*I10*	*I11*	*I12*	*I13*	*I14*	*I15*	*I16*	*All*
MSE	0.754	0.856	0.873	0.827	0.679	0.761	0.786	0.654	0.848	0.720	0.712	0.595	0.640	0.939
VPSNR	0.900	0.912	0.838	0.872	0.781	0.779	0.778	0.839	0.777	0.679	0.672	0.778	0.730	0.868
SSIM [17]	1.043	1.003	1.003	0.963	0.875	0.854	0.879	0.929	0.930	0.752	0.810	0.849	0.769	0.874
MS-SSIM [18]	0.988	0.967	0.947	0.939	0.844	0.827	0.838	0.913	0.916	0.744	0.788	0.818	0.674	0.848
UQI [19]	0.800	0.849	0.872	0.875	0.767	0.777	0.741	0.853	0.840	0.686	0.723	0.733	0.715	0.819
VIFp [37]	0.818	0.852	0.847	0.825	0.751	0.748	0.728	0.820	0.830	0.681	0.714	0.721	0.694	0.816
VIF [37]	0.811	0.833	0.840	0.804	0.739	0.749	0.737	0.818	0.829	0.676	0.720	0.720	0.689	0.800
NQM [38]	0.794	0.788	0.856	0.750	0.688	0.737	0.740	0.787	0.849	0.635	0.754	0.676	0.674	0.809
IW-PSNR [39]	0.979	0.950	0.959	0.909	0.820	0.840	0.773	0.888	0.909	0.677	0.734	0.805	0.742	0.822
IW-SSIM [39]	1.010	0.974	0.986	0.939	0.853	0.838	0.863	0.914	0.920	0.746	0.795	0.836	0.758	0.912
FSIM [40]	1.045	0.999	1.041	0.962	0.873	0.861	0.893	0.927	0.933	0.752	0.807	0.853	0.773	0.930
FSIMc [40]	1.045	0.999	1.040	0.962	0.873	0.861	0.893	0.927	0.933	0.752	0.807	0.853	0.773	0.930
RFSIM [41]	0.967	0.956	0.927	0.900	0.809	0.809	0.850	0.883	0.892	0.726	0.773	0.830	0.741	0.883
SR-SIM [42]	1.060	1.005	1.057	0.978	0.871	0.867	0.895	0.927	0.928	0.757	0.811	0.853	0.770	0.934
BRISQUE [43]	0.643	0.656	1.139	0.692	0.457	0.807	0.819	0.595	0.645	0.542	0.606	0.590	0.567	0.838
NFERM [44]	0.606	0.774	1.003	0.724	0.534	0.861	0.836	0.512	0.632	0.638	0.615	0.563	0.527	0.943
ADD-SSIM [45]	1.025	0.861	0.971	0.849	0.860	0.846	0.853	0.916	0.912	0.750	0.808	0.827	0.750	0.920
ADD-GSIM [45]	0.988	0.893	0.969	0.822	0.835	0.834	0.843	0.906	0.922	0.741	0.782	0.822	0.744	0.909
PSIM [46]	1.008	0.974	0.960	0.935	0.841	0.829	0.852	0.905	0.912	0.741	0.789	0.836	0.753	0.906
FWQI [47]	1.157	1.104	1.176	1.042	0.937	0.911	0.966	0.965	0.983	0.772	0.828	0.885	0.805	0.989
WSNR [38]	0.981	0.951	0.927	0.907	0.821	0.712	0.849	0.889	0.802	0.683	0.806	0.685	0.739	0.847
FWSNR [20]	0.925	0.913	0.952	0.865	0.828	0.688	0.769	0.698	0.893	0.717	0.746	0.751	0.730	0.959
FPSNR [48]	0.917	0.598	0.508	0.903	0.313	0.725	0.383	0.589	0.481	0.247	0.172	0.196	**0.142**	0.791
F-SSIM [21]	0.981	1.010	1.014	0.955	0.874	0.856	0.894	0.929	0.939	0.755	0.814	0.854	0.771	0.900
W-VPSNR [49]	**0.220**	**0.211**	**0.235**	**0.377**	**0.180**	**0.136**	**0.169**	**0.208**	**0.229**	**0.196**	**0.154**	**0.183**	0.259	**0.616**



(a) FWQI (*I3*) (b) WSNR (*I3*) (c) FWSNR (*I3*) (d) FPSNR (*I3*) (e) F-SSIM (*I3*) (f) W-VPSNR (*I3*)

(g) FWQI (*I7*) (h) WSNR (*I7*) (i) FWSNR (*I7*) (j) FPSNR (*I7*) (k) F-SSIM (*I7*) (l) W-VPSNR (*I7*)

(m) FWQI (*I13*) (n) WSNR (*I13*) (o) FWSNR (*I13*) (p) FPSNR (*I13*) (q) F-SSIM (*I13*) (r) W-VPSNR (*I13*)

(s) FWQI (*I16*) (t) WSNR (*I16*) (u) FWSNR (*I16*) (v) FPSNR (*I16*) (w) F-SSIM (*I16*) (x) W-VPSNR (*I16*)

FIGURE 10. Scatter plots of the values of the foveal quality metrics versus the MOS values for images *I3*, *I7*, *I13*, and *I16*: (a)-(f) Image *I3*; (g)-(l) Image *I7*; (m)-(r) Image *I13*; (s)-(x) Image *I16*.

metrics, most of their values in scenario *S*#1 are significant lower than those in scenario *S*#2. For the FWQI metric, with the same MOS value, their corresponding values vary in a wide range. These result in the low performances of these foveal quality metrics. Meanwhile, it can be observed that the higher the MOS values are, the larger the W-VPSNR values become in general. Hence, the W-VPSNR metric achieves quite good performances. Also, the performance of

the FPSNR metric is high with images *I13* and *I16*, but low with images *I3* and *I7*.

From the above analysis, we can see that among the investigated metrics, the W-VPSNR metric is rather effective to evaluate omnidirectional images with non-uniform quality, although in some cases its performance is lower than that of the FPSNR metric. The FPSNR metric has very high performances in certain images, it performs even worse than the simple MSE metric in some other images. Moreover, the performances of all the quality metrics are not good across different images. This suggests that it is necessary to integrate content characteristics in these quality metrics.

VI. CONCLUSION

In this paper, we have conducted subjective and objective quality assessments of omnidirectional images with non-uniform quality focusing on foveation feature of the human eyes. Based on the obtained results and discussions, some findings can be summarized as follows.

- The perceptual quality is affected by two key factors, which are the sensitivity of the human eyes and content characteristics.
- The zones of an image corresponding to the fovea and parafovea of the human eyes are extremely important for the perceptual quality.
- Content characteristics including the attractiveness and the size of central object, as well as the presence of neighboring objects affect the quality perception.
- Most of the twenty-five objective quality metrics considered in this study are not effective to evaluate omnidirectional images with non-uniform quality. However, the foveal metrics are very promising and could be further improved.
- In general, the performances of the investigated metrics vary drastically across different contents.

It is expected that the presented database can help researchers to build effective objective quality metrics which are essential to evaluate encoding and delivery solutions. For future work, further investigations with more content types and quality variation patterns will be conducted to derive better understanding of viewers' perceptual behaviors as well as the performances of existing metrics.

REFERENCES

[1] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "A new adaptation approach for viewport-adaptive 360-degree video streaming," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Taichung, Taiwan, Dec. 2017, pp. 38–44.

[2] J. Chakareski, R. Aksu, X. Corbillon, G. Simon, and V. Swaminathan, "Viewport-driven rate-distortion optimized 360° video streaming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–7.

[3] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in *Proc. IEEE Int. Conf. Image Process.*, Beijing, China, Sep. 2017, pp. 2174–2178.

[4] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 29–42, Mar. 2019.

[5] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, "Foveated 3D graphics," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 164:1–164:10, Nov. 2012.

[6] R. Albert, A. Patney, D. Luebke, and J. Kim, "Latency requirements for foveated rendering in virtual reality," *ACM Trans. Appl. Perception*, vol. 14, no. 4, pp. 25:1–25:13, 2017.

[7] P. Guo, Q. Shen, Z. Ma, D. J. Brady, and Y. Wang, "Perceptual quality assessment of immersive images considering peripheral vision impact," 2018, *arXiv:1802.09065*. [Online]. Available: https://arxiv.org/abs/1802.09065

[8] D. V. Nguyen, H. T. T. Tran, and T. C. Thang, "Impact of delays on 360-degree video communications," in *Proc. TRON Symp. (TRONSHOW)*, Tokyo, Japan, Dec. 2017, pp. 1–6.

[9] J. Besharse and D. Bok, *The Retina and Its Disorders*. New York, NY, USA: Academic, 2011.

[10] S. Lee, A. C. Bovik, and B. L. Evans, "Efficient implementation of foveation filtering," in *Proc. Texas Instrum. DSP Educator's Conf.*, 1999, pp. 1–5.

[11] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of foveated video coding using audio-visual focus of attention," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1322–1331, Nov. 2011.

[12] C.-F. Hsu, A. Chen, C.-H. Hsu, C.-Y. Huang, C.-L. Lei, and K.-T. Chen, "Is foveated rendering perceivable in virtual reality?: Exploring the efficiency and consistency of quality assessment methods," in *Proc. 25th ACM Int. Conf. Multimedia*, Mountain View, CA, USA, Oct. 2017, pp. 55–63.

[13] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual quality assessment of omnidirectional images," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Florence, Italy, May 2018, pp. 1–5.

[14] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[15] 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.

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

[17] 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.

[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.

[19] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[20] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.

[21] H. Ha, J. Park, S. Lee, and A. C. Bovik, "Perceptually unequal packet loss protection by weighting saliency and error propagation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1187–1199, Sep. 2010.

[22] H. T. T. Tran, C. T. Pham, N. P. Ngoc, A. T. Pham, and T. C. Thang, "A study on quality metrics for 360 video communications," *IEICE Trans. Inf. Syst.*, vol. E101-D, no. 1, pp. 28–36, 2018.

[23] L. Zhaoping, *Understanding Vision: Theory, Models, and Data*. Oxford, U.K.: Oxford Univ. Press, 2014.

[24] M. Yanoff and J. S. Duker, *Ophthalmology*. Philadelphia, PA, USA: Saunders, 2013.

[25] A. Hendrickson, "Organization of the adult primate fovea," in *Macular Degeneration*. Berlin, Germany: Springer, 2005.

[26] H. Strasburger, I. Rentschler, and M. Juttner, "Peripheral vision and pattern recognition: A review," *J. Vis.*, vol. 11, no. 5, p. 13, May 2011.

[27] V. Roberto, *Intelligent Perceptual Systems: New Directions in Computational Perception*, vol. 745. Springer, 1993.

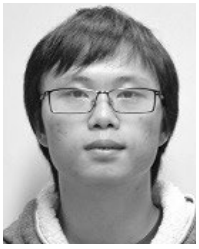
[28] E. Pöppel and L. O. Harvey, Jr., "Light-difference threshold and subjective brightness in the periphery of the visual field," *Psychologische Forschung*, vol. 36, no. 2, pp. 145–161, 1973.

[29] J. A. Jones, J. E. Swan, and M. Bolas, "Peripheral stimulation and its effect on perceived spatial scale in virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 4, pp. 701–710, Apr. 2013.

[30] 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.*, Taipei, Taiwan, Jun. 2017, pp. 205–210. [Online]. Available: https://salient360.ls2n.fr/datasets/training-dataset/

[31] Y. Rai, P. Le Callet, and P. Guillotel, "Which saliency weighting for omni directional image quality assessment?" in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Erfurt, Germany, May 2017, pp. 1–6.
[32] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba, "Recognizing scene viewpoint using panoramic place representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, Jun. 2012, pp. 2695–2702.
[33] Princeton Vision Group. *SUN360 Panorama Database*. Accessed: Sep. 8, 2019. [Online]. Available: https://vision.princeton.edu/projects/2012/SUN360/data/
[34] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, document Rec. P.913 ITU-T, 2014.
[35] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[36] Y.-F. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2473–2486, Jun. 2014.
[37] 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.
[38] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
[39] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[40] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[41] L. Zhang, D. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 321–324.
[42] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1473–1476.
[43] 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.
[44] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
[45] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.
[46] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.
[47] Z. Wang, A. C. Bovik, L. Lu, and J. L. Kouloheris, "Foveated wavelet image quality index," *Proc. SPIE*, vol. 4472, pp. 42–52, Dec. 2001.
[48] S. Lee and A. C. Bovik, "Foveated video image analysis and compression gain measurements," in *Proc. 4th IEEE Southwest Symp. Image Anal. Interpretation*, Apr. 2000, pp. 63–67.
[49] H. T. T. Tran, T. H. Hoang, P. N. Minh, N. P. Ngoc, and T. C. Thang, "A perception-based quality metric for omnidirectional images," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Bangkok, Thailand, Jun. 2019, pp. 1–2.
[50] Joint Video Exploration Team. *360Lib*. Accessed: Oct. 3, 2018. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-2.0.1/

HUYEN T. T. TRAN received the B.E. degree from the Hanoi University of Science and Technology, Vietnam, in 2014, and the M.Sc. degree from The University of Aizu, Japan, in 2017, where she is currently pursuing the Ph.D. degree in computer science and engineering. Her research interests include quality of experience (QoE), multimedia networking, and content adaptation. She has been a recipient of Japanese Government Scholarship (MonbuKagaku-sho) for graduate study, since 2015. In 2017, she received the IEEE Signal Processing Society (SPS) student travel grant.



DUC V. NGUYEN received the B.E. and M.E. degrees in computer science and engineering from The University of Aizu, Japan, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree. He is also a Research Assistant with the Computer Communications Laboratory, The University of Aizu. His research interests include video streaming, virtual reality, and networking. He has been a recipient of Japanese Government Scholarship for graduate study, since 2015.

NAM PHAM NGOC received the B.E. degree in electronics and telecommunication from the Hanoi University of Science and Technology, Vietnam, in 1997, the M.Sc. degree in artificial intelligence from K.U. Leuven, Belgium, in 1999, and the Ph.D. degree in electrical engineering from K.U. Leuven, in 2004. He is currently the Director of Engineering and Technology Program, Vin University Project, Vietnam. His research interests include QoS management at end-systems for multimedia applications, reconfigurable embedded systems, and low-power embedded system design.

TRANG H. HOANG is currently pursuing the bachelor's degree with the Hanoi University of Science and Technology. She is also a Research Assistant with the Embedded Systems and Reconfigurable Computing Laboratory, School of Electronics and Telecommunications. Her research interests include quality of experience (QoE) and multimedia networking. She achieved Global Korea Scholarship (GKS) for ASEAN countries' Science and Engineering Students, in 2018, and Student Exchange Program at National Taipei University of Technology (Taipei Tech), in 2019.

TRUONG THU HUONG received the B.Sc. degree in electronics and telecommunications from the Hanoi University of Science and Technology (HUST), Vietnam, in 2001, the M.Sc. degree in information and communication systems from the Hamburg University of Technology, Germany, in 2004, and the Ph.D. degree in telecommunications from the University of Trento, Italy, in 2007. She came back to work for Hanoi University of Science and Technology as a Lecturer, in 2009, and became an Associate Professor, in 2018. Her research interests are oriented toward network security, artificial intelligence, traffic engineering in next generation networks, QoE/QoS guarantee for network services, green networking, and development of the Internet of Things ecosystems and applications.

TRUONG CONG THANG received the B.E. degree from the Hanoi University of Science and Technology, Vietnam, in 1997, and the Ph.D. degree from KAIST, South Korea, in 2006. From 1997 to 2000, he was a Network Engineer with the Vietnam Post and Telecommunications (VNPT). From 2007 to 2011, he was a member of Research Staff with the Electronics and Telecommunications Research Institute (ETRI), South Korea. He has been an Active Member of Korean and Japanese delegations to standard meetings of ISO/IEC and ITU-T, since 2002. Since 2011, he has also been an Associate Professor with The University of Aizu, Japan. His research interests include multimedia networking, image/video processing, content adaptation, IPTV, and MPEG/ITU standards.