Subjective Quality Assessment of Foveated Omnidirectional Images in Virtual Reality

Aliakbar Bozorgian, Marius Pedersen, Jean-Baptiste Thomas, Mohamed-Chaker Larabi

IEEE Open Journal on
Immersive Displays

Received 9 January 2025; revised 12 March 2025; accepted 25 March 2025. Date of publication 28 March 2025;
date of current version 18 April 2025. The review of this article was arranged by Associate Editor Jun Yu.

Digital Object Identifier 10.1109/OJID.2025.3556364

Subjective Quality Assessment of Foveated Omnidirectional Images in Virtual Reality

ALI BOZORGIAN [1], MARIUS PEDERSEN [1] (Member, IEEE), JEAN-BAPTISTE THOMAS [1], AND MOHAMED-CHAKER LARABI [2] (Senior Member, IEEE)

[1]Colourlab, Department of Computer Science, NTNU—Norwegian University of Science and Technology, 7034 Gjøvik, Norway
[2]Laboratory XLIM, Department SIC, Universite de Poitiers, 86000 Poitiers, France

CORRESPONDING AUTHOR: ALI BOZORGIAN (e-mail: ali.bozorgian@ntnu.no).

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT This study presents a novel dataset called "Foveated Omnidirectional Image Quality Assessment" (FOIQA) for the subjective quality evaluation of foveated 2D omnidirectional images. This dataset addresses the limitations of existing datasets by leveraging a high-resolution head-mounted display and a gaze-contingent evaluation approach. We provide individual opinion scores, mean opinion scores, and gaze data associated with both the test and reference images. The utility of our dataset is validated by benchmarking two existing objective foveated image quality metrics. Our results demonstrate that incorporating gaze data into the evaluation framework improves the accuracy of one of the tested objective metrics. The dataset is publicly available at https://doi.org/10.5281/zenodo.14009106.

INDEX TERMS Eye tracking, foveated rendering, near-eye displays, spatial resolution, subjective image quality assessment, virtual reality.

I. INTRODUCTION

Virtual and augmented reality systems predominantly rely on near-eye displays to deliver immersive visual experiences. Ideally, near-eye displays should produce retinal images that closely replicate the physical, optical, spatial, and temporal characteristics of natural scenes with high fidelity [1]. Most consumer-grade near-eye displays offer a spatial resolution between 10 and 30 Pixels Per Degree (PPD). In contrast, a person with 20/20 vision has a resolving power of 1 arcminute, which is equivalent to 60 PPD. This discrepancy is a significant factor that contributes to the degradation of image quality. Consequently, improving the spatial resolution of near-eye displays is of utmost importance; however, it presents considerable challenges.

Visual acuity is not consistent across the visual field. It reaches its maximum in the central region of the retina, known as the fovea, and decreases in the peripheral areas. A cautious objective for the spatial resolution of near-eye displays could be to match the foveal resolution. However, achieving this might be unrealistic for consumer-grade devices, as rendering two separate high-resolution images for each eye at an adequate refresh rate is highly demanding for a GPU. Furthermore, transmitting such high-quality visual content poses additional challenges. However, a perceptual-aware design in hardware and software can capitalize on the constraints of human vision to mitigate the computational demands of rendering and the bandwidth requirements for transmission.

The constraints of human vision relevant in this discussion relate to the variations in optical quality and neural processing between foveal and peripheral vision. As noted previously, the optical performance of the eye and the sampling rate of the retinal neurons decrease in human peripheral vision as the angular distance from the fovea increases [2]. This angular distance is typically referred to as eccentricity in units of visual degrees. This reduction leads to lower contrast sensitivity in peripheral vision.

Foveation techniques leverage this reduction in contrast sensitivity in peripheral vision to facilitate either lower rendering resolution or higher compression [3]. The core concept of foveated techniques is to deliver a sharper image for foveal vision while maintaining a sufficient but less detailed image for peripheral vision. This can be accomplished by either

tracking the user's gaze or by making assumptions about typical gaze behavior. Instances of foveation are present in both the hardware and software design of near-eye displays. In this study, our emphasis is on algorithms for foveated image processing.

With increasing interest in the development and deployment of foveated processing algorithms for immersive visual content, there is a growing focus on assessing the impact of foveation on perceived image quality [4], [5], [6], [7], [8], [9], [10]. Image Quality Assessments (IQAs) can be conducted subjectively through visual experiments or objectively by applying computational models [11], [12]. Although subjective assessments are time-consuming, they provide ground-truth data essential for the development and benchmarking of objective models. However, objective models strive to predict subjective quality ratings using computational methods. These objective models should be accurate, efficient, and ideally interpretable on the basis of the human visual system's anatomy and physiology.

Currently, there are a limited number of subjective datasets available to assess foveated image quality [7], [9], [13]. This study introduces "FOIQA," a new dataset intended for subjective quality evaluation of foveated 2D omnidirectional images. This new dataset aims to rectify certain limitations present in current datasets. We achieve this by employing a high-resolution near-eye display that integrates an eye-tracker. Additionally, we introduce an evaluation approach for objective IQA metrics, which involves viewport extraction and utilizes gaze data to assign weights to metric outputs. Conclusively, we assess two objective IQA metrics with our dataset to showcase its practicality.

II. SURVEY OF EXISTING SUBJECTIVE DATASETS

A. NON-FOVEATED SUBJECTIVE ASSESSMENT DATASETS

Upenik et al. [14] created a testbed for the subjective assessment of omnidirectional media and performed a preliminary experiment. Their testbed application managed various elements of the test procedure, including training sessions, instructions, and duration of stimulus presentation. The absolute category rating with hidden references [15] was used to assess the image quality of six 360-degree scenes subjected to different JPEG compression levels and displayed in both equirectangular and cubicmap projection formats. The reference images and the HMD had a maximum resolution of 3000×1500 and 667×750, respectively. The HMD used in this experiment was, in fact, a mount named MergeVR, which housed an iPhone 6S serving as the screen. A total of 48 participants participated in the experiment; half rated the equirectangular format, while the other half rated the cubic format.

Sun et al. [16] created a comprehensive 360-degree spherical image quality database named CVIQD, which included 5 reference images and 165 test images, which were compressed using JPEG, H.264 / AVC, and H.265 / HEVC techniques. For their visual experiment, they used the single stimulus method [17] to enhance the immersive experience for the participants. These reference images were captured using an Insta360 camera with a resolution of 4096×2048 pixels. The study used a HTC VIVE HMD with a resolution of 1080×1020 pixels and a Field Of View (FOV) of 110°, yielding an angular resolution of 9.81 PPD. They tested the performance of nine objective IQA metrics against the Mean Opinion Scores (MOS) provided by 20 subjects, concluding that multiscale models such as MS-SSIM [18] have a higher correlation with subjective evaluations.

Omnidirectional IQA (OIQA) is another database for 360-degree subjective quality analysis, consisting of 16 reference images and 320 test images that have been degraded using JPEG compression, JPEG2000 compression, Gaussian blur, and Gaussian noise [19]. The image resolution vary from 11332×5666 to 13320×6660. Each test image was shown on an HTC VIVE HMD for a duration of 20 seconds, during which the head movements and gaze directions of the 20 participants were tracked. The HTC VIVE's resolution is 1080×1020 pixels with a FOV of 110°, providing an angular resolution of 9.81 PPD. The authors identified an equator bias after examining the generated saliency maps. Furthermore, they suggested that integrating saliency information into IQA metrics could be beneficial based on the comparison between the objective IQA metrics and the corresponding MOS. Both conclusions confirm earlier findings in the field.

B. FOVEATED SUBJECTIVE DATASETS

LIVE-FBT-FCVR [7] includes two datasets of 2D and 3D foveated videos, each comprising 10 reference videos and 180 foveated videos evaluated by a cohort of 38 subjects. The experimental setup incorporated an HTC VIVE Pro HMD with a resolution of 1440×1600 pixels and a FOV of 110°, yielding an angular resolution of 13.09 PPD. The 2D and 3D video sequences boasted resolutions of 8K and 6K, respectively. The HMD was integrated with a Tobii eye tracker operating at a sampling rate of 120Hz. The collected gaze data facilitated the operation of the gaze-contingent foveated video player; however, data on head and eye movements is not made available by the authors.

Mantiuk et al. [9] compiled a dataset named FovDots comprising animated synthetic stimuli by meticulously controlling pivotal factors influencing foveated artifacts, such as contrast, luminance, and velocity. The experimental procedure adhered to the sequential two-interval forced choice methodology, presenting two animations for two seconds each. A total of 35 participants were recruited, with the task of discerning the sequence that exhibited superior visual quality. For visualization, an HTC VIVE Pro Eye HMD was used, featuring a resolution of 1440×1600 pixels and a FOV of 110°, achieving an angular resolution of 13.09 PPD. Throughout the experiment, participants were instructed to maintain central fixation on the display to ensure accurate foveated image delivery.

Deepfovea is a neural reconstruction model designed for foveated rendering and video compression, utilizing natural

TABLE 1. The Summary of IQA Datasets for Omnidirectional Images

Datasets	Foveated	Device	Display Resolution
Upenik et al. (2016)	No	MergeVR	667 × 750
CVIQD (2018)	No	HTC VIVE	1080 × 1020
OIQA (2018)	No	HTC VIVE	1080 × 1020
LIVE-FBT-FCVR (2021)	Yes	HTC VIVE Pro	1440 × 1600
FovDots (2021)	Yes	HTC VIVE Pro	1440 × 1600
DeepFovea (2019)	Yes	HTC VIVE Pro	1440 × 1600
Ours (FOIQA)	Yes	Varjo VR-3	2880 × 2720

Display resolutions expressed here are for each eye.

image statistics to create a plausible foveated image from a sparsely sampled input image [13]. To validate their approach and compare it with other foveated algorithms, the authors carried out a user study. This study was carried out under two conditions: one group of eight subjects using a large projector screen and another group of five subjects using an HMD. Both setups included eye trackers to monitor the subjects' gaze locations. The projector setup had a horizontal visual coverage of 78° and a resolution of 1920 × 1080 pixels. The used HMD is an HTC Vive Pro with a resolution of 1440 × 1600 pixels and a horizontal FOV of 110°. The authors reported that their proposed method showed lower detectability and Difference Mean Opnion Scores (DMOS) compared to the multiresolution foveation algorithm [20]. A summary of all foveated and non-foveated dataset is listed in Table 1.

III. A NEW SUBJECTIVE QUALITY ASSESSMENT DATASET FOR FOVEATED IMAGES

A. DEVICE

For our visual experiment, we used the Varjo VR-3 HMD. This device features a distinct display configuration with two panels per eye: an LCD panel for the peripheral region (referred to as the 'context display') and a uOLED microdisplay for the foveal region (referred to as the 'focus display'). The context display has a resolution of 2880 × 2720 pixels, offering a 115° horizontal FOV. The right and left context displays do not entirely overlap in the visual spaces they cover, and as per the manufacturer, the angular resolution is 35 PPD. The focus display has a 1920 × 1920 pixel resolution, covering 27° × 27° of the visual field, yielding an angular resolution of 71 PPD. The image from the focus display is optically merged with the image from the context display using a fixed-position beam splitter, then directed through an aspheric lens to the eye. The Varjo VR-3 also includes an automatic InterPupillary Distance (IPD) adjustment mechanism (ranging from 59 to 71mm) and an eye tracker with a 200Hz refresh rate (accuracy below 1°). Despite the focus display's near-ideal angular resolution for the human eye, we deactivated it during our visual experiment because of the noticeable seam between the two displays. Furthermore, having dual overlays would add complexity to the evaluation of IQA metrics. Despite this, at the time of our experiment, the context display provided the highest angular resolution seen in any commercially available HMD.

TABLE 2. Rendering Resolution and Foveated Parameters for Test Images

Index	Context scaling factor	Context texture size	Focus texture size
1	1.0	2192 × 1880	1200 × 1200
2	0.8	1753 × 1504	
3	0.7	1534 × 1260	
4	0.6	1315 × 1128	
5	0.5	1096 × 940	
6	1.0	1660 × 1420	1076 × 1076
7	0.9	1494 × 1278	
8	0.8	1328 × 1136	
9	0.7	1162 × 994	
10	0.6	996 × 852	

B. REFERENCE AND TEST IMAGES

We selected seven equirectangular images from the OIQA dataset [19] as our reference scenes. These images, which are licensed under Creative Commons licenses, have a resolution of 13320 × 6660 pixels. They depict natural outdoor environments with varying degrees of detail (refer to Fig. 1). To obtain test images, we used the dynamic foveation feature of the Varjo Unity XR plugin (version 3.20.0). This foveation algorithm employs two viewports: a dynamic high-resolution viewport of size 27° × 27° that tracks the observer's gaze, and a fixed low-resolution viewport that spans the entire FOV of the HMD. Note that resolution in this context refers to the size of the textures. The algorithm's primary parameters are the "focus scaling factor" and the "context scaling factor", both adjustable from 0 to 1. Lower values reduce the texture size of the corresponding viewport. Additionally, the Varjo Base application offers a general resolution setting (called 'rendering quality') that influences the texture size of both viewports. Ultimately, ten different combinations of overall resolution and context scaling factor were applied to each reference image, producing a total of 70 test images (see Table 2). These combinations were visually evaluated, by the authors, to encompass a range from subtle to noticeable but acceptable foveated artifacts.

C. VISUAL EXPERIMENT METHODOLOGY

A custom omnidirectional image viewer was developed using Unity software (version 2021.3.11f1). All raw reference omnidirectional images were loaded into Unity, compressed using the BC7 format, and resized to 16384 × 8192 pixels using the bilinear resize method. To ensure an immersive experience for the subjects, each omnidirectional image was applied as a two-dimensional texture to a skybox material using Unity's default panoramic shader. This presentation method allowed subjects to explore the scene freely through head movements. For subjective quality assessment, we used the single stimulus method as detailed in [21]. The experiment included 41 participants, of which 27 were male and 14 were female. All participants verbally agreed to take part in the experiment. The ages of the participants ranged from 16 to 60 years, with an average age of 29 years. Among them, 18 had no previous experience, 18 had limited experience,

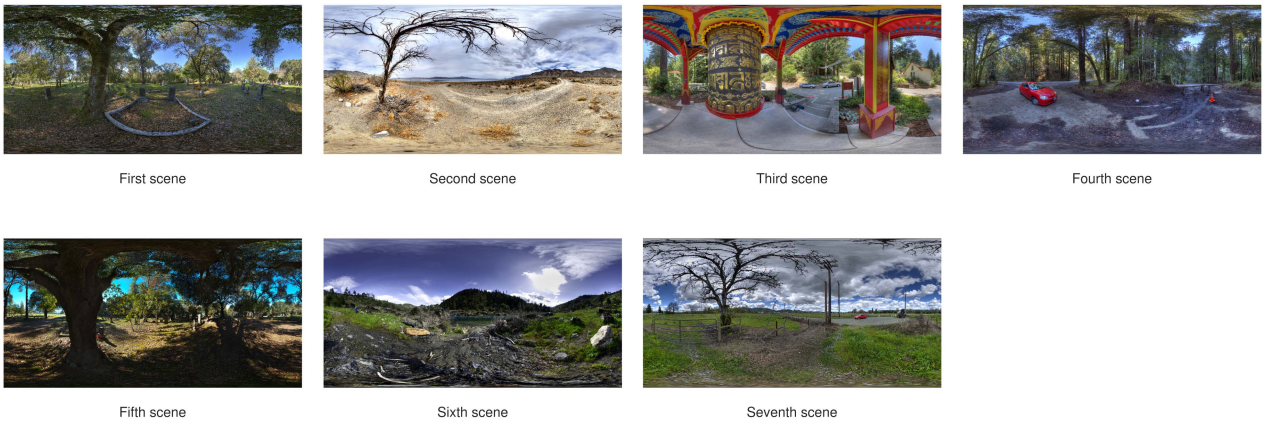

FIGURE 1. 360-degree scenes used for our dataset. All of these images are adopted from the OIQA database [19], have a resolution of 13320 × 6660 pixels and are available in Raw format.

and 3 were occasional users of HMDs. We ensured that all subjects did not have color deficiencies and had normal or corrected to normal vision using the Ishihara test for color vision and the Snellen chart for visual acuity. The HMD eye box was sufficiently large to fit standard eyewear, so subjects who wore prescription glasses were instructed to wear them during the experiment. Before starting, each participant was verbally briefed about the purpose of the study, HMD usage, controller operation, and several safety points. Information about the types of artifacts and visual impairments included in the experiment was withheld to avoid influencing their attention to peripheral areas, which could result in lower detection thresholds. The participants sat on a swivel chair and were instructed to explore the scene before rating it. The HMD's built-in eye tracking was calibrated each time a participant wore the device.

The experiment consisted of two sessions, each preceded by a training phase. During the training phase, the participants evaluated the quality of eight test images that exemplified the range of artifacts present in the study. This phase aimed to help participants familiarize themselves with the rating procedure and eliminate the initial variability in their assessments, which could influence the final results. The main phase included 40 test images, with five images repeated to assess intra-subject consistency. The images were presented in a random sequence in each session. Each image was shown for 15 seconds to provide sufficient exploration time without inducing fatigue from prolonged sessions. Following each test image, an input box appeared in which participants used a controller to adjust a labeled slider and rate the overall quality of the scene (see Fig. 2). The slider featured five labels indicating quality levels (Poor, Bad, Fair, Good, Excellent), with intermediate levels available [22]. Participants were allowed as much time as they needed to complete their scores. Out of 41 subjects, 39 completed both sessions.

D. DATA PROCESSING

During the experiment, we collected three different types of data: subjective opinion scores, head-rotations, and gaze locations. In the upcoming sections, we will review the processing and analysis methods used for each type of information.

FIGURE 2. The input box including an adjustable slider with corresponding labels.

1) SUBJECTIVE OPINION SCORES

In order to determine the MOS, the raw opinion scores were standardized for each subject per session and then averaged among all subjects per test image [7]. The raw opinion score given by the subject ith for the foveated image jth during session kth is denoted as OS_{ijk}. To standardize the opnion scores, z-scores (z_{ijk}) were calculated per session using the following expressions:

$$\mu_{ik}^{MOS} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} OS_{ijk}, \tag{1}$$

$$\sigma_{ik}^{MOS} = \sqrt{\frac{\sum_{j=1}^{N_{ik}} (OS_{ijk} - \mu_{ik}^{MOS})^2}{N_{ik} - 1}}, \tag{2}$$

$$z_{ijk}^{MOS} = \frac{OS_{ijk} - \mu_{ik}^{MOS}}{\sigma_{ik}^{MOS}}, \tag{3}$$

where N_{ik} is the number of foveated images observed by ith subject in the kth session, μ_{ik} is the mean, and σ_{ik} is the standard deviation of the opinion scores collected from ith subject in the kth session. The z-scores span the interval from -3 to 3. To facilitate easier interpretation, we map these z-scores to

a scale of 0 to 100 using the following formula:

$$z'^{MOS}_{ijk} = \frac{100\left(z^{MOS}_{ijk} + 3\right)}{6}. \tag{4}$$

MOS are obtained from z-scores through the following equation:

$$MOS_{jk} = \frac{\sum_{i=1}^{N_{jk}} z'^{MOS}_{ijk}}{N_{jk}}, \tag{5}$$

where N_{jk} is the number of observers who rated the quality of jth foveated image in kth session.

Difference Opinion Scores (DOS) were obtained from raw opinions scores using the following expression:

$$DOS_{ijk} = s^{ref}_{ik} - s_{ijk}, \tag{6}$$

where s^{ref}_{ik} is the subjective quality rating associated with the hidden reference in session kth by subject ith. The DMOS are obtained using following expressions:

$$\mu^{DMOS}_{ik} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} DOS_{ijk}, \tag{7}$$

$$\sigma^{DMOS}_{ik} = \sqrt{\frac{\sum_{j=1}^{N_{ik}} \left(DOS_{ijk} - \mu^{DMOS}_{ik}\right)^2}{N_{ik} - 1}}, \tag{8}$$

$$z^{DMOS}_{ijk} = \frac{DOS_{ijk} - \mu^{DMOS}_{ik}}{\sigma^{MOS}_{ik}}, \tag{9}$$

$$z'^{DMOS}_{ijk} = \frac{100\left(z^{DMOS}_{ijk} + 3\right)}{6}, \tag{10}$$

$$DMOS_{jk} = \frac{\sum_{i=1}^{N_{jk}} z'^{DMOS}_{ijk}}{N_{jk}}. \tag{11}$$

The correlation-based post-processing of z'^{MOS}_{ijk} and z'^{DMOS}_{ijk} resulted in the rejection of 7 and 10 subjects, respectively [17]. The scores associated with these subjects are excluded from the calculation. The associated 95% confidence interval for each MOS is obtained through the following equation [17]:

$$CI = \left[MOS_{jk} - \delta_{jk}, MOS_{jk} + \delta_{jk}\right], \tag{12}$$

where:

$$\delta_{jk} = 1.96 \frac{STD_{jk}}{\sqrt{N_{jk}}}. \tag{13}$$

The standard deviation for each test image, STD_{jk}, is given by:

$$STD_{jk} = \sqrt{\frac{\sum_{i=1}^{N_{jk}} \left(MOS_{jk} - z'^{MOS}_{ijk}\right)^2}{N_{jk} - 1}}. \tag{14}$$

Provided the distribution of individual z-scores meets certain criteria, there is a high probability that the difference between the experimental mean score and the 'true' mean score (in cases where many observers are present) will be smaller than the 95% confidence interval.

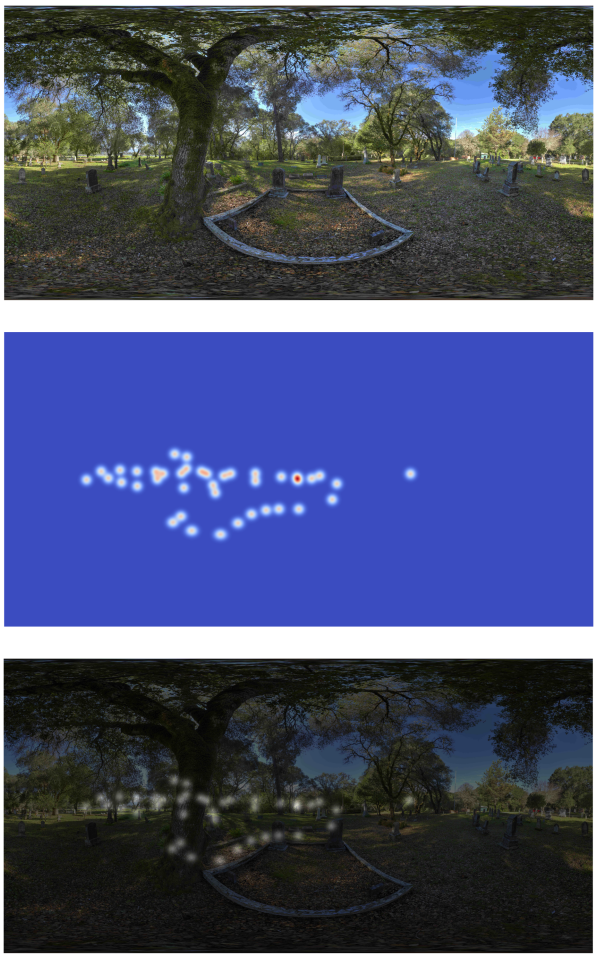

FIGURE 3. Example scene and associated saliency (attention) maps generated using the Salient360 Toolbox [23]. The top panel presents the scene in an equirectangular format. The middle panel displays the saliency map with color coding: cooler hues indicate areas without fixation, while warmer hues highlight fixation points. The bottom panel offers a blended view of the saliency map and the sample scene. Fixation points are detected via the I-VT algorithm [25] and blurred using a Gaussian function with a standard deviation of 2 visual degrees for illustrative purposes.

2) HEAD AND EYE MOVEMENTS

Throughout the main experiment, we recorded the head and eye movements of all subjects. Each data sample was associated with a timestamp and comprised a quaternion representing head-rotation, along with two unit vectors describing the gaze direction of each eye. For the analysis of the collected data, we use the Salient360 toolbox [23], specially designed to handle eye tracking data from virtual reality headsets. This comprehensive toolbox combines eye directions with head-rotations to identify saccades and fixations, and also offers the possibility to generate saliency maps and scanpaths. Fig. 3 shows an example saliency (attention) map.

Our primary objective in recording gaze data was to identify fixation points for integration into the evaluative framework

of objective IQA metrics (see Section IV-B). We hypothesize that fixation locations exert significant influence on subjects' opinion scores. This hypothesis is potentially supported by saccadic suppression, suggesting a reduction in visual sensitivity during saccadic eye movements [24]. The Salient360 Toolbox provides several methodologies for identifying fixations and saccades, among which we used the Velocity-Threshold Identification (I-VT) algorithm for fixation extraction. The core principle of the I-VT algorithm is the significantly higher point-to-point velocity of saccades (i.e., $>$ 300 deg/sec) compared to fixations (i.e., $<$ 100 deg/sec), enabling discrimination between fixations and saccades [25].

E. REFERENCE AND TEST IMAGE EXTRACTION

To make FOIQA accessible for evaluating the performance of objective IQA metric, we need to generate suitable reference and test images. In this context, suitable means that the properties of the reference and test images should closely match the viewing conditions experienced by subjects during the experiment. From this point forward, we refer to the reference and test image pairs as input images. In conventional IQA experiments, where subjects view images on conventional displays from a fixed distance, providing input images to IQA metrics is usually straightforward. In contrast, in experiments where subjects engage with immersive content, allowing for head-rotations and random eye movements, generating suitable input images becomes more complex and necessitates additional assumptions or eye-tracking data.

A typical approach to generate input images involves extracting viewports from the equirectangular image. These viewports emulate what an observer sees at a specific head-direction. The visual content within the viewport changes as the observer moves their head. We define head-directions as unit vectors starting from the center of a unit sphere and intersecting its surface. In our dataset, we create two viewports for each head-rotation: one for the left eye and another for the right eye. Current virtual reality screen capture solutions failed to provide the necessary flexibility for fast and reliable screenshot capture with our required specifications. As a result, we developed two custom applications using Unity and the Varjo native SDK [26]. In the Unity application, we rendered test images for the headset and controlled the camera's direction. The camera's direction controls the head-direction as described above. While Unity handled rendering, we used the multi-app functionality of the Varjo headset to run a concurrent C++ program [27]. In the C++ program, we rendered a transparent scene for the displays, overlayed on the image submitted by Unity to the Varjo compositor. This allowed us to request a mirror window using the Varjo API and DirectX12 [28]. Subsequently, we used the ScreenGrab library in the DirectX Toolkit for DirectX12 to capture the mirror window and save it to disk [29]. We used an inter-process communication client to facilitate communication between the two applications and synchronize camera control and mirror capture. Fig. 4 illustrates an example of the extracted viewports. The resolution of the viewports is 2880 $\times$ 2469 pixels, consistent with the texture generated by HMD for the left- and right-hand displays.

FIGURE 4. A sample of viewports extracted from a 360-degree image, with the image on the left illustrating the left eye's perspective and the image on the right illustrating the right eye's perspective. It is important to note that left and right view in Varjo VR-3 HMD do not completely overlap, making only part of the visual field binocular while the extreme periphery remains monocular.

IV. OBJECTIVE QUALITY ASSESSMENT

One of the primary objectives in creating subjective IQA datasets is to examine the performance of objective IQA metrics. The performance of an IQA metric is measured by its ability to accurately predict subjective opinion scores. Although numerous objective IQA metrics have been developed, only a few are suitable for omnidirectional visual content, and even fewer can handle foveated visual content. These foveated metrics typically consider the decrease in spatial contrast sensitivity in peripheral vision by using a Contrast Sensitivity Function (CSF) that incorporates eccentricity as an input parameter. Certain peripheral contrast sensitivity models [9], [30] use a cortical magnification factor to anticipate changes in sensitivity at higher eccentricities. This method is backed by observations from earlier studies [31], [32], [33], [34], in which peripheral contrast sensitivity was made comparable to foveal vision by adjusting the area and spatial frequency of the target to activate a similar cortical neural population. However, subsequent evidence indicates that simply scaling the foveal spatial contrast sensitivity could be insufficient to accurately describe peripheral spatial contrast sensitivity [35], [36], [37], due to the detectability of aliasing effects in the periphery. Thibos et al. [38] highlighted that the shape of the peripheral CSF for the detection task differs significantly from the foveal CSF, particularly at frequencies beyond the neural Nyquist limit. A recently proposed CSF model aims to accommodate this effect [39].

A. FOVEATED OBJECTIVE IQA METRICS

In this study, we evaluated two full-reference foveated metrics using our dataset: the Foveated Wavelet Image Quality Index (FWQI) [4] and FovVideoVDP [9]. Both metrics involve space-frequency analysis, which helps to accommodate contrast sensitivity and its variations across the visual field.

FWQI operates in the wavelet domain. Wavelet analysis allows the examination of spatial and frequency information concurrently. FWQI considers multiple aspects of the human

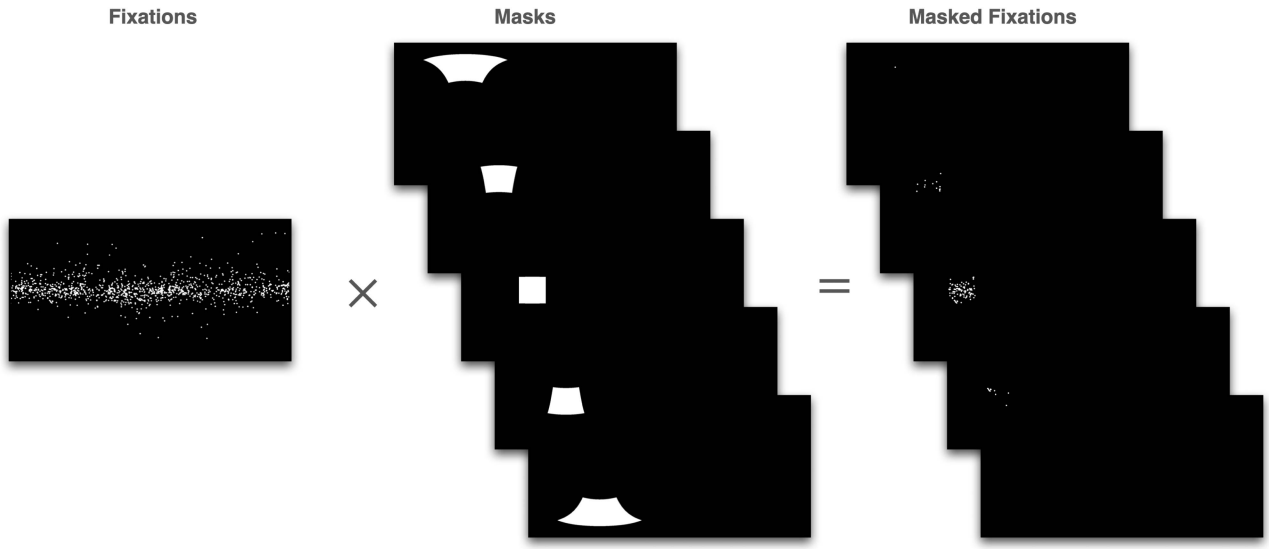

FIGURE 5. A schematic diagram illustrates the gaze-contingent weighting system for objective metrics. On the left column of the plot, there is a fixation map generated from all subjects for a sample scene. The middle column displays binary masks. These binary masks and the fixation map are multiplied together. The masks are crafted to isolate the ROI from the fixation map. Finally, the last column shows the masked fixations, where only the fixations within the ROI are retained, and those outside the ROI are excluded. Note that this plot showcases only a subset of masks. The fixation points have been enlarged for clarity.

visual system, including the variation of spatial sensitivity and cutoff frequency in the periphery, the variation of visual sensitivity across different wavelet subbands, and the influence of viewing distance on display resolution [4].

FovVideoVDP is a visible difference predictor that is designed for immersive video content, but is also applicable to wide-field-of-view images. The FovVideVDP metric relies on the Laplacian pyramid for multiscale decomposition. In addition, this metric incorporates a spatio-temporal CSF model, contrast masking, and cortical magnification to estimate the visibility of distortions across the visual field. It also takes into account various display characteristics, such as peak luminance, size, resolution, and viewing distance [9].

B. EVALUATION FRAMEWORK

When generating viewports, there are countless potential head-directions to sample from a sphere. This creates a challenge in constructing a dataset for omnidirectional visual content. The content-dependent nature of objective quality metrics means that their predictions can vary with the number and location of viewports. Therefore, we should first generate a sufficiently large number of viewports to capture the distribution of all possible objective quality scores and then implement a pooling strategy to derive a descriptor that yields a better correlation with the DMOS. To address this challenge, we rely on two pooling methods: the baseline method and the gaze-contingent method. We extract 60 pairs of viewports for all approaches. A pair of viewports includes one view for the left eye and another view for the right eye. The sampled head-directions are derived from 12 equally spaced longitudes and 5 latitudes. The number of head-directions is kept constant among the approaches to make them comparable. The point of zero eccentricity is assumed to coincide with the point on which a head-direction vector lands on. It is important to note that the center of the visual field does not coincide with the centers of the left and right eye viewports, since the two viewports do not completely overlap (see Fig. 4).

1) BASELINE METHOD

In the baseline method, we simply calculate the mean of predicted quality scores of all extracted viewports. This approach ignores gaze information and presumes that there are no consistent patterns in how individuals explore a 360-degree image.

2) GAZE-CONTINGENT METHOD

In the gaze-contingent method, we implement a weighting system in which viewports containing more fixation points are given a greater weight, while those with fewer or no fixation points are assigned less weight. Fig. 5 illustrates a schematic representation of the weighting system. First, we create 60 equirectangular binary masks with the Region Of Interest (ROI) having the maximum pixel value (255) and pixels outside the ROI set to zero. The ROI in each binary mask corresponds to the region from which the viewports are derived. The ROI size for each mask is fixed at 30×30 visual degrees, roughly matching the focus area size in the foveation algorithm implemented by the HMD. Next, each binary mask is multiplied with a fixation map obtained from gaze data across all subjects, as detailed in Section III-D2. As a result, we generate a vector with 60 components, where each component records the number of fixation points that fall into the ROIs. In the final step, we calculate the ratio of the fixations within each ROI to the total fixations present in the

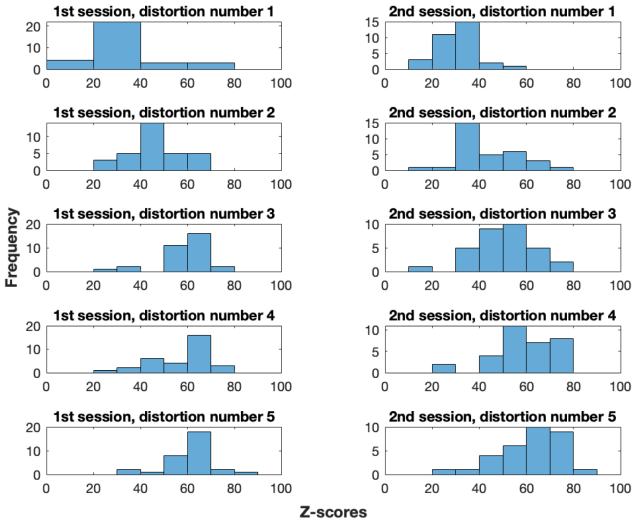

FIGURE 6. The histogram of Z-scores representing standardized opinion scores for a sample scene under five distortion levels in both sessions. Distortion level 1 represents the highest degree of foveation, whereas distortion level 5 is the hiddien reference image.

test image. This new vector is then used to assign weights to the objective scores predicted for each viewport pair.

3) EVALUATION CRITERIA

Four evaluation criteria were used to measure the accuracy of the predicted objective quality scores: the Spearman Rank Correlation Coefficient (SRCC), the Pearson Linear Correlation Coefficient (PLCC), the Root Mean Square Error (RMSE), and Uncertainty Standardised residual sum of squares (USTRESS) [40]. Prior to the computation of these metrics, a non-linear monotonic transformation was applied to the IQA metric predictions. This transformation is intended to mitigate any potential nonlinearity in the output of the objective metrics. We use the following expression [41]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2 (x - \beta_3)\right)} \right) + \beta_4 x + \beta_5, \quad (15)$$

with x representing the output of the IQA metrics and β_1 through β_5 being the parameters to optimize. The objective function used for optimization was the root squared mean error between DMOS and the metric's output. The "fmincon" solver in MATLAB was used to find the minimum of the objective function. We set constraint on the parameters to ensure monotocity of the regression.

V. RESULTS

A. OPINION SCORES

Fig. 6 shows the histograms of the opinion scores for a sample scene from the main experiment. The figure clearly demonstrates that higher foveation levels shift the distribution of opinion scores towards the lower quality range. A similar trend is observed in all scenes; however, we have not included the corresponding histograms for other scenes to maintain brevity. The MOS obtained from the main experiment are illustrated in Fig. 7, with associated error bars. The lengths of these error bars represent the 95% confidence intervals calculated using (12). The figure verifies that each scene experienced a significant variation in image quality. The lowest recorded MOS score was 23.5 and the highest was 64.7.

B. INTRA-SUBJECT CONSISTENCY

In each session, five test images were chosen randomly and repeated to evaluate the consistency of subjects' quality judgements. To assess intra-subject consistency, we use the Intraclass Correlation Coefficient (ICC), which is particularly effective for examining correlations and agreements within a dataset. Following the guidelines provided by [42], we applied the ICC with a 95% confidence interval, considering the mean rating, absolute agreement, and a two-way mixed-effects model as described in [43]. To calculate ICC values, we create a table for each subject in which columns represent the trials of quality judgment (e.g., the first and second judgments of the same test image), and rows represent different repeated images. The ICC values are derived using the following expression:

$$ICC(A-k) = \frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}} \quad (16)$$

where MS_R is the mean square of the rows, MS_E is mean square error (random error, unexplained variation), MS_C is the mean square between columns. ICC values range from -1 to 1, with higher values indicating more consistency. Fig. 8 shows the ICC values for all subjects. The majority of subjects exhibit ICC values exceeding 0.5, indicating consistency in their quality judgements; however, seven subjects have ICC values below 0.5, suggesting inconsistency in their judgements [42]. The observed inconsistency may stem from the subjects' limited interaction with VR headsets. Out of the seven subjects exhibiting poor consistency, five had never used VR devices before, and the other two had limited exposure. It is important to note that the ratings from these subjects were not outliers, as they passed our screening for outliers. Their ratings merely showed poor consistency within each subject.

C. INTER-SUBJECT CONSISTENCY

To evaluate inter-subject consistency, we compute the standard deviation of the opinion scores for each image across all subjects. A lower standard deviation suggests less variation and tighter consensus among subjects, indicating consistent evaluations. The standard deviation of the opinion scores ranged from 7.66 to 17.90, averaging 12.71 in the first session and 12.12 in the second session. In addition, to gauge the agreement among subjects on quality judgments, we calculate ICC values. First, we generate a table where the columns represent subjects and the rows represent test images. The resulting ICC value is 0.9397 which implies good agreement across subjects [42].

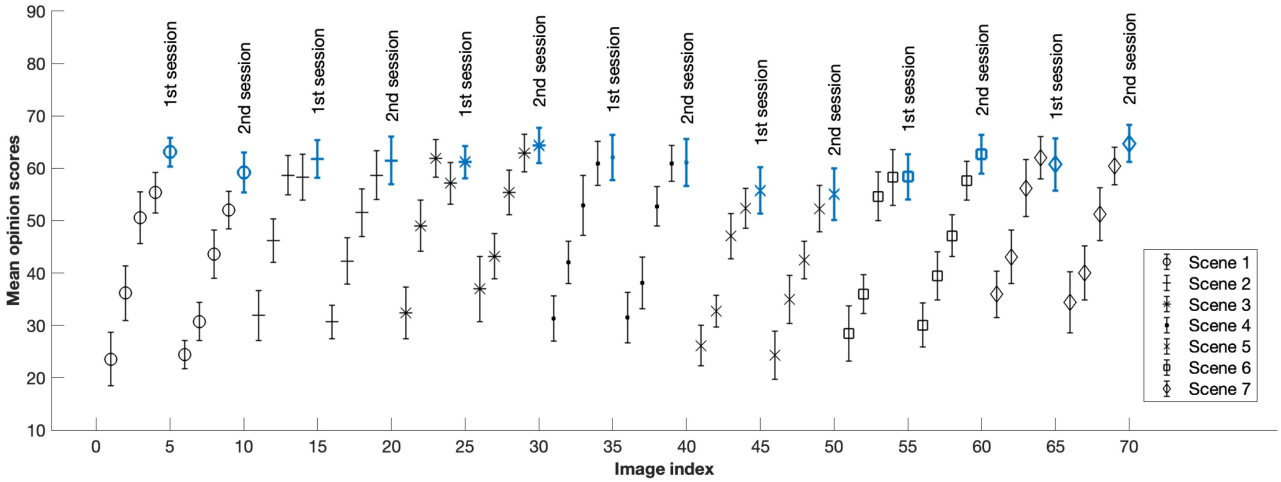

FIGURE 7. MOS from the main experiment for both the first and second sessions. The error bars represent 95% confidence intervals. The image indices on the x-axis correspond to those listed in Table 2. Hidden reference images are shown in blue, and distinct markers differentiate scores for each scene. The vertical labels above the data clusters indicate the first and second sessions for each scene.

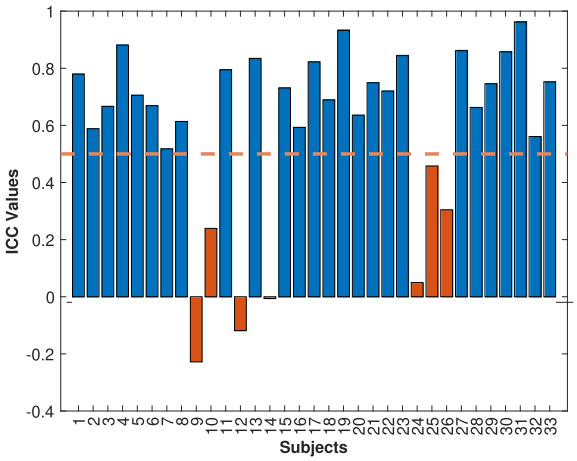

FIGURE 8. The ICC analysis for intra-subject consistency is depicted. The horizontal dashed line represents ICC = 0.5, indicating the threshold below which consistency is considered poor, as per [42]. Blue bars signify ICC values above 0.5, whereas orange bars denote ICC values below 0.5.

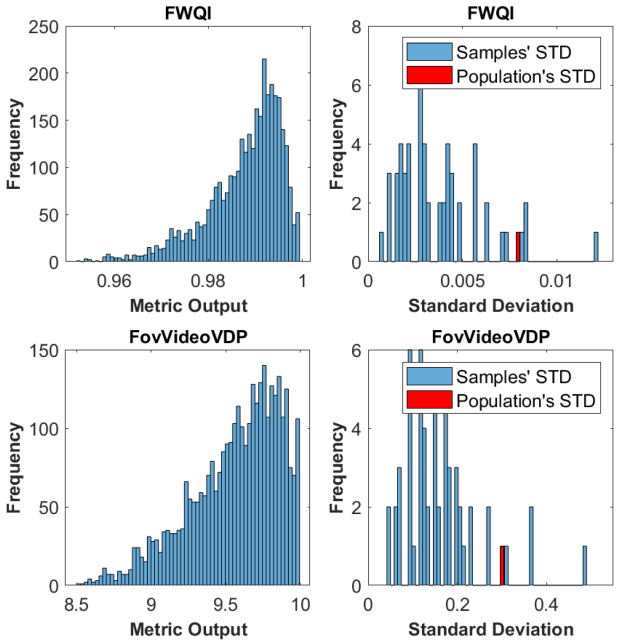

FIGURE 9. The histogram of z-scores representing standardized opinion scores for a sample scene under five distortion levels in both sessions. Distortion level 1 represents the highest degree of foveation, whereas distortion level 5 is the hiddien reference image.

D. THE OBJECTIVE QUALITY PREDICTIONS

The predictions of the objective metric were derived from a total of 6720 input images, with 3360 images per eye. In the case of FWQI, we set the viewing distance to $v = 0.5\,w$ where w is the width of the test image. The viewing distance is adjusted so that the eccentricity map generated by the metric matches what was rendered by the headset for subjects during the experiment. In addition, we used six levels of wavelet decomposition. The possible outputs of FWQI range from 0 to 1, where higher values indicate better quality [4]. In our dataset, FWQI predictions range between 0.9663 and 0.9966. A histogram displaying the predictions of FWQI is located in the top left plot of Fig. 9. The FovVideoVDP metric is measured in Just-Objectionable Differences (JOD), where a JOD of 10 means that there are no noticeable differences between the reference and test images, suggesting optimal quality [9]. For our dataset, FovVideoVDP predictions range from 8.51 to 9.81. A histogram illustrating the predictions of FovVideoVDP is shown in the bottom left plot of Fig. 9. For comparison purposes, we also calculated the predictions of SSIM on the dataset. SSIM is a widely used, non-foveated metric that overlooks the reduction in sensitivity in peripheral vision [44].

As mentioned in Section IV-B, we use two different pooling methods to derive a single objective opinion score from the distribution of all objective predictions obtained from the

TABLE 3. The Numerical Values for SRCC, PRCC, RMSE, and USTRESS Calculated Between Objective Quality Scores and DMOS

Method	Metric	SRCC	PRCC	RMSE	USTRESS
Baseline	SSIM	0.5782	0.5789	8.6950	1.3814
	FWQI	0.7558	0.7617	6.9097	1.3461
	FovVideoVDP	0.6269	0.6450	8.1488	1.2888
Gaze-contingent	SSIM	0.5778	0.5971	8.5539	1.3511
	FWQI	0.7505	0.7673	6.8383	1.3525
	FovVideoVDP	0.6876	0.6913	7.7054	1.2059

Higher numerical values for SRCC and PRCC show higher performance, while lower values for RMSE and USTRESS show higher performance. For comparative purposes, the results of SSIM are also included. Note that SSIM is not a foveated metric.

viewports of a particular test image. Examining these pooling methods is only sensible if there is a substantial variation in the predictions of the objective metric across the viewports within a test image. Therefore, prior to delving into the results of these methods, an analysis of the spread of objective quality predictions is conducted.

There were 56 test images in total, excluding the hidden reference in the subjective experiment. Each of these test images yields 60 objective quality predictions, leading to 3320 predictions in total. The subset of predictions associated with a test image is considered a sample from the overall population of objective scores, and its standard deviation is compared to that of the whole population. The top right and bottom right plots in Fig. 9 show that the standard deviation of the samples can be as large as that of the population, indicating that the predictions of the objective metric for a test image can vary as much as they do across different distortions and scenes. This suggests that a pooling method can play a significant role in deriving a single descriptive opinion score.

The numerical values for SRCC, PRCC, RMSE, and USTRESS are presented for metrics in Table 3. USTRESS [40] is a recently proposed performance measure that takes into account the variability of the subjective scores. Figs. 10 and 11 depict scatter plots of metric predictions versus DMOS, for both the baseline and gaze contingent methods. As shown in the figures, there is a significant flattening of the non-linear transformation in the lower quality score range across metrics and pooling methods. This indicates that while the metrics assign different quality scores to the test images in the lower quality range, the subjects' scores did not vary. Furthermore, as evidenced by both Table 3 and Fig. 11, the predictions of FovVideoVDP are improved in the gaze-contingent pooling method. Ultimately, we carried out a Wilcoxon Signed Rank test to verify that the differences between the baseline and the gaze-contingent pooling methods are significant. The hypothesis test indicated that the pooling method significantly affects the data distribution. The Wilcoxon Signed Rank test is suitable when we cannot assume that the distribution of the data is normal.

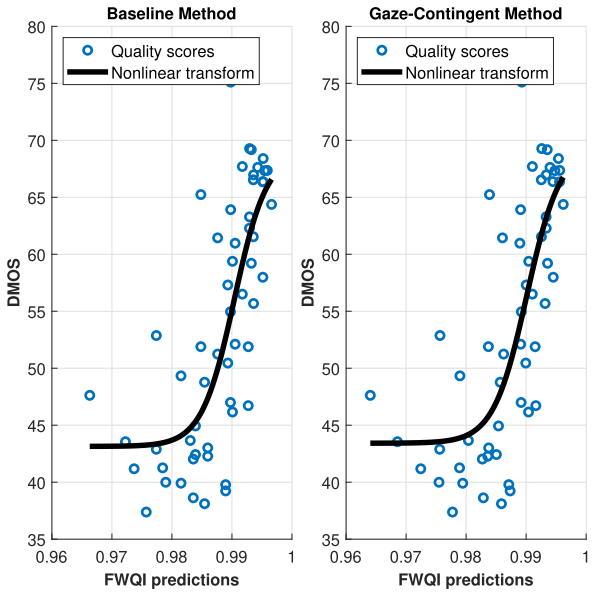

FIGURE 10. The left plot illustrates the comparison of FWQI predictions to DMOS for the Baseline method, whereas the right plot shows this comparison for the Gaze-contingent pooling method. The solid line represents the best fit for the non-linear transformation.

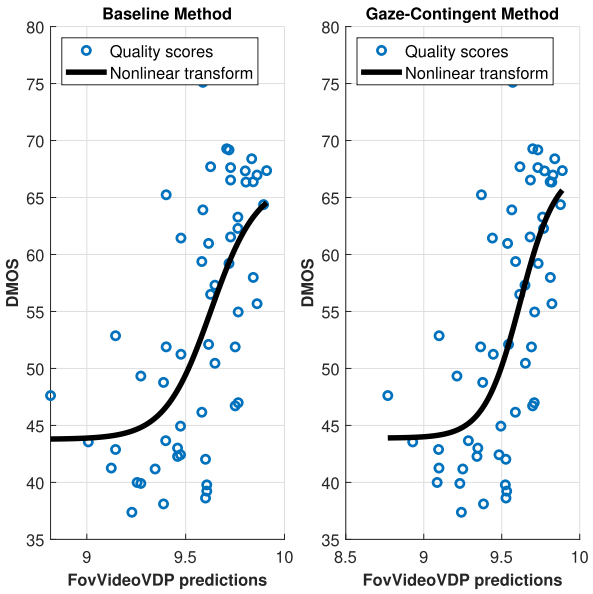

FIGURE 11. The left plot illustrates the comparison of FovVideoVDP predictions to DMOS for the Baseline method, whereas the right plot shows this comparison for the Gaze-contingent pooling method. The solid line represents the best fit for the non-linear transformation.

E. DIFFERENCE MAPS

In addition to a single JOD score, FovVideoVDP can generate a difference map to illustrate the distribution of distortions across the test image. Fig. 12 shows an example of a difference map, represented by a heat map where warmer hues denote a greater visibility of the distortions. These difference maps represent both near-threshold and suprathreshold distortions. Clearly seen on the difference map, there is no visible distortion within a square region. This square region pertains to the high-resolution focus view from the foveated rendering algorithm. As anticipated, with the scaling factors for both the reference and test images set to 1, there is no distortion within

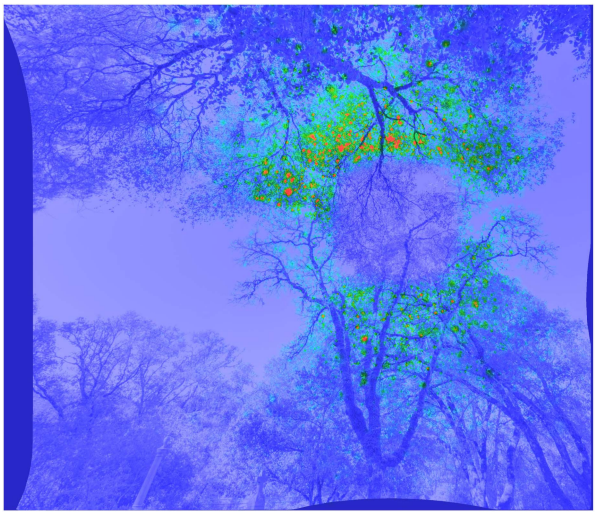

FIGURE 12. An example of a difference map from the first scene generated for the left eye's viewport. The difference map is illustrated as a heatmap where cooler hues signify no visible distortions, and warmer hues signify visible distortions.

the focus area. However, distortions start to emerge at the boundary between the focus view and the context view. This occurs because the context view's scaling factor is reduced for test images, resulting in visible distortions. Nevertheless, distortions are noticeable only over a narrow range and gradually diminish towards the test image's edge in areas of higher eccentricity. This is because the spatial contrast sensitivity significantly reduces in higher eccentricities, rendering distortions unnoticeable. A notable observation in this difference map is that distortions are present mainly in regions with high spatial frequency details (such as tree branches), while areas with lower detail (like the sky) exhibit fewer distortions. This suggests that high-spatial frequency visual content is more susceptible to distortions caused by foveated rendering in our dataset.

VI. CONCLUSION

This study presents a new subjective foveated IQA dataset, useful for evaluating no-reference, reduced-reference, and full-reference objective IQA metrics. This dataset is comprised of 8400 viewports (4200 images per eye). We introduced a new gaze-contingent method for foveated quality metric evaluation, demonstrating its impact on metric prediction distributions and performance enhancement in one tested metric. Alongside reference and test images, we are sharing individual opinion scores, MOS, DMOS, and raw gaze data. This dataset is freely accessible for download in https://doi.org/10.5281/zenodo.14009106. One constraint of our research is that we did not account for the optical transfer function of the Varjo VR-3's optical system in our metric predictions. This is due to the proprietary nature of HMD's optical systems, with manufacturers not fully revealing their optical characterization. Nonetheless, this constraint is common across all current datasets used for subjective image quality evaluation, and it is not exclusive to our work.

REFERENCES

[1] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Comput. Graph. Forum*, vol. 38, no. 2, pp. 493–519, 2019.

[2] S. J. Anderson, K. T. Mullen, and R. F. Hess, "Human peripheral spatial resolution for achromatic and chromatic stimuli: Limits imposed by optical and retinal factors," *J. Physiol.*, vol. 442, no. 1, pp. 47–64, 1991.

[3] A. Floren and A. C Bovik, "Foveated image and video processing and search," in *Academic Press Library in Signal Processing*, J. Trussell, A. Srivastava, A. K. Roy-Chowdhury, A. Srivastava, P. A. Naylor, R. Chellappa, and S. Theodoridis, Eds., vol. 4, Amsterdam, The Netherlands: Elsevier, 2014, pp. 349–401.

[4] Z. Wang, A. C. Bovik, L. Lu, and J. L. Kouloheris, "Foveated wavelet image quality index," *Appl. Digit. Image Process. XXIV*, vol. 4472, pp. 42–52, 2001.

[5] S. R. Drlje, M. Vranješ, and D. Žagar, "Foveated mean squared error - A novel video quality metric," *Multimedia Tools Appl.*, vol. 49, no. 3, pp. 425–445, 2010.

[6] W. J. Tsai and Y. S. Liu, "Foveation-based image quality assessment," in *Proc. IEEE Vis. Commun. Image Process. Conf.*, 2014, pp. 25–28.

[7] Y. Jin, M. Chen, T. Goodall, A. Patney, and A. C. Bovik, "Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality," *IEEE Trans. Image Process.*, vol. 30, pp. 5905–5919, 2021.

[8] Y. Jin, A. Patney, R. Webb, and A. C. Bovik, "FOVQA: Blind foveated video quality assessment," *IEEE Trans. Image Process.*, vol. 31, pp. 4571–4584, 2022.

[9] R. K. Mantiuk et al., "FovVideoVDP: A visible difference predictor for wide field-of-view video," *ACM Trans. Graph.*, vol. 40, no. 4, 2021, Art. no. 49.

[10] A. Bozorgian, M. Pedersen, and J. B. D. Thomas, "The effect of peripheral contrast sensitivity functions on the performance of the foveated wavelet image quality index," in *Proc. London Imag. Meeting*, Society for Imaging Science and Technology, 2022, pp. 6–10.

[11] P. Wang, H. Duan, Z. Xie, X. Min, and G. Zhai, "Subjective and objective quality assessment for augmented reality images," *IEEE Open J. Immersive Displays*, vol. 1, pp. 135–145, 2024.

[12] H. Duan, X. Zhu, Y. Zhu, X. Min, and G. Zhai, "A quick review of human perception in immersive media," *IEEE Open J. Immersive Displays*, vol. 1, pp. 41–50, 2024.

[13] A. S. Kaplanyan, A. Sochenov, T. Leimkühler, M. Okunev, T. Goodall, and G. Rufo, "DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–13, 2019.

[14] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *Proc. Picture Coding Symp.*, 2017, pp. 1–5.

[15] Subjective Video Quality Assessment Methods for Multimedia Applications, ITU-T Recommendation P.910, International Telecommunication Union, Geneva, Switzerland, 2009.

[16] W. Sun et al., "CVIQD: Subjective quality evaluation of compressed virtual reality images," in *Proc. - Int. Conf. Image Process.*, Septe, 2018, pp. 3450–3454.

[17] Methodologies for the Subjective Assessment of the Quality of Television Images, IT.U. Recommendation BT. 500-15, International Telecommunication Union, Geneva, Switzerland, 2023.

[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Conf. Rec. Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.

[19] 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.*, 2018, pp. 1–5.

[20] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, "Foveated 3D graphics," *ACM Trans. Graph.*, vol. 31, pp. 1–10, 2012.

[21] Methodology for the Subjective Assessment of the Quality of Television Pictures, Rec. ITU-R BT.500-11. 211BT.500-11, International Telecommunication Union, Geneva, Switzerland, 2002.

[22] *Objective Perceptual Assessment of Video Quality: Full Reference Television*, ITU-T Telecommunication Standardization Bureau, Standard ITU-T J.144, Geneva, Switzerland, 2004.
[23] E. David et al., "The salient360! toolbox: Processing, visualising and comparing gaze data in 3D," in *Proc. Eye Tracking Res. Appl. Symp.*, 2023, pp. 1–8.
[24] J. Ross, M. C. Morrone, M. E. Goldberg, and D. C. Burr, "Changes in visual perception at the time of saccades," *Trends Neurosciences*, vol. 24, no. 2, pp. 113–121, 2001.
[25] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Eye Tracking Res. Appl. Symp.*, 2000, pp. 71–78.
[26] "Varjo Native SDK," Varjo for developers, Apr. 3, 2025, Accessed: Apr. 6, 2025. [Online]. Available: https://developer.varjo.com/docs/native/varjo-native-sdk
[27] "Multi-app with Native SDK," Varjo for developers, Apr. 3, 2025, Accessed: Apr. 6, 2025. [Online]. Available: https://developer.varjo.com/docs/native/multi-app-native
[28] "Varjo Native SDK: include/Varjo_types_layers.h File Reference," Varjo.com, 2025, Accessed: Apr. 6, 2025. [Online]. Available: https://developer.varjo.com/docs/v3.3.0/apidocs/_varjo__types__layers_8h.html
[29] Microsoft, "ScreenGrab," GitHub, Nov. 5, 2022, Accessed: Apr. 6, 2025. [Online]. Available: https://github.com/microsoft/DirectXTK12/wiki/ScreenGrab
[30] J. Scott Daly, "Visible differences predictor: An algorithm for the assessment of image fidelity," *Proc. SPIE*, vol. 1666, pp. 2–15, 1992.
[31] J. Rovamo, V. Virsu, and R. Näsänen, "Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision," *Nature*, vol. 271, no. 5640, pp. 54–56, 1978.
[32] J. Rovamo and V. Virsu, "An estimation and application of the human cortical magnification factor," *Exp. Brain Res.*, vol. 37, no. 3, pp. 495–510, 1979.
[33] J. J. Koenderink, M. A. Bouman, A. E. B. Mesquita, and S. Slappendel, "Perimetry of contrast detection thresholds of moving spatial sine wave patterns. III. the target extent as a sensitivity controlling parameter," *J. Opt. Soc. Amer.*, vol. 68, no. 6, pp. 854–859, 1978.
[34] D. H. Kelly, "Retinal inhomogeneity I spatiotemporal contrast sensitivity," *J. Opt. Soc. Amer. A*, vol. 1, no. 1, pp. 107–113, 1984.
[35] N. J. Coletta and D. R. Williams, "Psychophysical estimate of extrafoveal cone spacing," *J. Opt. Soc. Amer. A*, vol. 4, no. 8, pp. 1503–1513, 1987.
[36] N. J. Coletta, D. R. Williams, and C. L. M. Tiana, "Consequences of spatial sampling for human motion perception," *Vis. Res.*, vol. 30, no. 11, pp. 1631–1648, 1990.
[37] D. R. Williams and N. J. Coletta, "Cone spacing and the visual resolution limit," *J. Opt. Soc. Amer. A*, vol. 4, no. 8, pp. 1514–1523, 1987.
[38] L. N. Thibos, D. L. Still, and A. Bradley, "Characterization of spatial aliasing and contrast sensitivity in peripheral vision," *Vis. Res.*, vol. 36, no. 2, pp. 249–258, 1996.
[39] A. Bozorgian, M. Pedersen, and J.-B. Thomas, "Modification and evaluation of the peripheral contrast sensitivity function models," *J. Opt. Soc. Amer. A*, vol. 39, no. 9, pp. 1650–1658, 2022.
[40] P. L.Carmona, R. Huertas, M. Pedersen, and S. Morillas, "Proposal of a new fidelity measure between computed image quality and observers quality scores accounting for scores variability," *J. Vis. Commun. Image Representation*, vol. 90, 2023, Art. no. 103704.
[41] A. Bovik, *The Essential Guide to Image Processing*. New York, NY, USA: Academic, 2009.
[42] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, 2016.
[43] O. Cherepkova, A. Amirshahi, and M. Pedersen, "Individual contrast preferences in natural images," *J. Imag.*, 2024, vol. 10, 2024, Art. no. 25.
[44] 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.

ALI BOZORGIAN received the Bachelor of Science degree in engineering and the Master of Science degree in color science from the Amirkabir University of Technology, Tehran, Iran, in 2017 and 2020, respectively. He was with the Ph.D. program with the Computer Science Department, Norwegian University of Science and Technology, Trondheim, Norway. His research interests include human visual perception, image quality, and computer vision.

MARIUS PEDERSEN (Member, IEEE) is currently a Professor in Colour Imaging with the Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway. He is also the Leader for the unit Colourlab and Deputy for Gjøvik for the Head of Department. His research interests include colour imaging, with applications to graphic arts, medical, and security.

JEAN-BAPTISTE THOMAS received the bachelor's in applied physics in 2004, the master's in optics, image, and vision in 2006 from Université Jean Monnet, Saint-Étienne, France,and the Ph.D. degree from the Université de Bourgogne, Dijon, France, in 2009. Since 2010, he has been an Associate Professor with the Université de Bourgogne. Between 2016 and 2019, he was in a sabbatical with NTNU as a Researcher, and from 2019 as an Associate Professor. He was extensively on the development of spectral imaging systems through Spectral Filter Arrays technology. Since 2016, he has been in understanding material appearance and its measure by using imaging systems.

MOHAMED-CHAKER LARABI (Senior Member, IEEE) received the Ph.D. degree in 2002 from the University of Poitiers, Poitiers, France, where he is currently a Full Professor. He has supervised more than 20 Ph.Ds., authored or coauthored 200 papers, and led several national and international projects. His scientific interests include image and video processing including quality assessment, compression, optimization, and enhancement, traditional, and learning-based, for various types of content including immersive media. He is member of JPEG, MPEG, and AOM where he took various roles. He is currently a Senior Associate Editor for IEEE Transactions on Image Processing and has held editorial roles for IEEE, Elsevier, and EURASIP journals. He has organized or participated to the organizing committees of numerous conferences including MMSP, ICIP, ICASSP, and ICME.