

---

# Visual Quality Assessment of Point Clouds Compared to Natural Reference Images

---

Aram Baek<sup>1</sup>, Minseop Kim<sup>1</sup>, Sohee Son<sup>1</sup>, Sangwoo Ahn<sup>2</sup>,  
Jeongil Seo<sup>3</sup>, Hui Yong Kim<sup>4,\*</sup> and Haechul Choi<sup>1,\*</sup>

<sup>1</sup>*Department of Intelligence Media Engineering, Hanbat National University, Daejeon, Republic of Korea*

<sup>2</sup>*Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea*

<sup>3</sup>*Department of Computer Engineering, Dong-A University, Busan, Republic of Korea*

<sup>4</sup>*School of Computing, Kyung Hee University, Yongin 17104, Republic of Korea*  
*E-mail: hykim.v@khu.ac.kr; choihc@hanbat.ac.kr*

*\*Corresponding Author*

Received 25 October 2022; Accepted 18 February 2023;  
Publication 04 July 2023

## Abstract

This paper proposes a point cloud (PC) visual quality assessment (VQA) framework that reflects the human visual system (HVS). The proposed framework compares natural images acquired using a digital camera and PC images generated via 2D projection in terms of appropriate objective quality evaluation metrics. Humans primarily consume natural images; thus, human knowledge is typically formed from natural images. Thus, natural images can be more reliable reference data than PC data. The proposed framework performs an image alignment process based on feature matching and image warping to use the natural images as a reference which enhances the similarities of the acquired natural and corresponding PC images. The framework facilitates identifying which objective VQA metrics can be used to reflect the HVS effectively. We constructed a database of natural images

*Journal of Web Engineering, Vol. 22\_3, 405–432.*

doi: 10.13052/jwe1540-9589.2232

© 2023 River Publishers

and three PC image qualities, and objective and subjective VQAs were conducted. The experimental result demonstrates that the acceptable consistency among different PC qualities appears in the metrics that compare the global structural similarity of images. We found that the SSIM, MAD, and GMSD achieved remarkable Spearman rank-order correlation coefficient scores of 0.882, 0.871, and 0.930, respectively. Thus, the proposed framework can reflect the HVS by comparing the global structural similarity between PC and natural reference images.

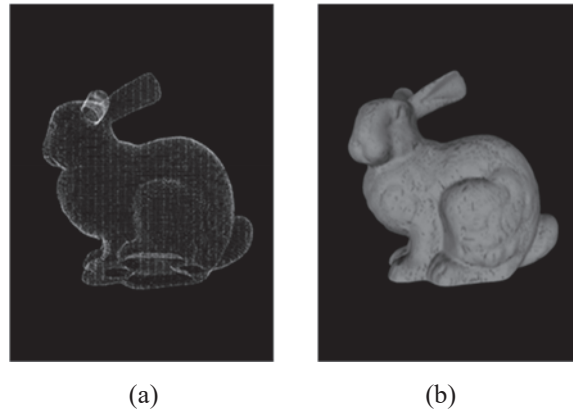
**Keywords:** Point cloud, quality evaluation, visual quality assessment.

## 1 Introduction

In recent years, rich content expression technologies are required to increase user participation with the development of immersive applications, e.g., augmented reality, virtual reality, and car or robot navigation systems. Such applications provide experiences that stimulate the user's senses in increasingly realistic ways. The content used for such applications requires capturing and processing 3D scene and object shape data [1]. One of the most primitive forms of such data is the point cloud (PC), a set of points in a given coordinate space [2]. The rapid development of 3D scanning technologies, e.g., the Microsoft Kinect and FARO Laser Scanner, has made it easier to acquire PC data [3–5]. Figure 1 shows an example of a PC “bunny,” generated via surface reconstruction using the zippering method [6–8].

A PC contains a large amount of data, including points, colors, location information, and attributes. These data often require compression technology while retaining data quality [10–14]. PC data generation could be subject to significant error due to inaccuracy in the depth acquisition or 3D reconstruction techniques when the data are obtained directly from depth sensors or from image information [15, 16]. Thus, a framework is required to evaluate various PC processing techniques, e.g., data acquisition, error removal, compression, and streaming, which would provide useful guidelines for PC-based system design [17]. Accordingly, with the development of PC technologies to provide users with immersive experiences, research into visual quality assessment (VQA) of PC data is becoming increasingly important.

VQA has been studied in the image and video processing field. VQA is generally interpreted as fidelity or similarity to a reference image in some spaces. One of the purposes of VQA is signal fidelity measurements that reflect physiological and psychological visual features of the human



**Figure 1** Bunny point cloud [9]: (a) raw points, and (b) surface reconstruction.

visual system (HVS), so it is important that VQA exploits the properties of HVS [18]. A typical point-to-point error evaluation metric is the peak signal to noise ratio (PSNR); however, this metric is unlikely to reflect HVS perception accurately [19, 20]. For 2D image VQA, the structural similarity index (SSIM) was developed to utilize HVS characteristics better [21]. VQA applications reflecting the HVS are being conducted in various fields, such as panoramic video, video conferencing, stereoscopic video, video codecs, and composite image databases [18].

PC quality can be evaluated quantitatively according to geometry errors, e.g., point-to-point, point-to-plane, and point-to-surface errors between the target PC data and reference PC data in virtual 3D space [22–24]. However, points are scattered in 3D space, and there is no clear connection structure; thus, it isn't easy to find corresponding reference PC data. Like the PSNR, geometric-based PC VQA techniques do not reflect the HVS [25–38]. Thus, an assessment technique that can reflect and verify the HVS is required for PC data.

Therefore, to reflect the HVS in PC VQA, this paper proposes a VQA method that utilizes natural images as a reference. Natural images captured by cameras are ubiquitous in modern society; thus, they significantly affect prior human knowledge. Such images can serve as reliable references based on human experiences, whereas PC data are used for reference in conventional PC VQAs. Thus, the proposed framework includes a database for VQA comprising natural images captured by digital cameras and PC images projected from PC data. Feature-based matching is first performed to warp and align

the common area of the images to compare the natural and PC image. The aligned images are then evaluated using objective VQA techniques from the image and video processing fields. The proposed framework was verified by observing the consistency among different PC qualities in objective VQA experiments and the correlation between the results of relevant objective metrics and subjective VQA techniques.

The remainder of this paper is organized as follows. Section 2 introduces studies related to the VQA of PCs. Section 3 introduces the proposed PC VQA method, and Section 4 describes our experimental evaluations and corresponding results. Finally, the paper is concluded in Section 5.

## 2 Related Work

Machine-based objective and human-based subjective VQAs can be performed according to the evaluation subject.

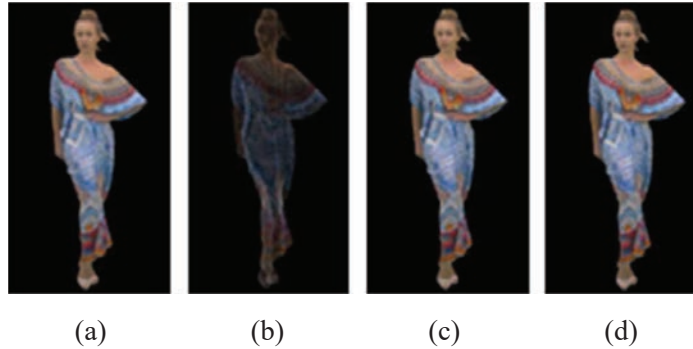
An objective assessment can obtain consistent results through experiments using the same test data without being affected by time, space, and people. Objective assessments are frequently used to facilitate the design of video communication systems and optimize video processing systems. Objective VQA techniques can be classified as full-reference (FR), reduced-reference (RR), and nonreference methods [39]. FR methods require undistorted or original images, and RR methods require information derived from the original images. For objective VQA, the FR method is the most widely used, and the mean squared error (MSE) and PSNR are the most common evaluation metrics used in FR methods. However, these point-to-point evaluation metrics correlate poorly with subjective VQA results [40–45].

As stated previously, the objective quality of a PC can be evaluated based on geometric-based metrics, i.e., point-to-point, point-to-plane, and point-to-surface comparisons between the target and reference PC data. Point-to-point metrics evaluate the average distance between points in the target PC and the corresponding points in the reference PC. The average distance is obtained using the root mean square or Hausdorff distance. Point-to-plane metrics identify corresponding points between the target PC and the reference PC. Here, normal vectors are taken if available; otherwise, they are estimated using the nearest neighbor points. Point-to-plane distortion is computed as the root mean square of the magnitudes of all projected error vectors. The point-to-surface metrics consider the distance between points in the target PC and the surface of the reference PC. Note that point-to-surface metrics require reference PCs with surface reconstruction using 3D mesh, and its

distortion is obtained like the point-to-plane metric [22]. Another approach for objective VQA of a PC is projection-based methods, which obtain a set of projections of the target PC and reference PC. Projection-based methods also use standard 2D image quality evaluation metrics, e.g., PSNR and SSIM, to compare the set [46]. Deep learning-based VQA approaches have recently begun to be studied [47]. Alessandro et al. [48] propose a model for learning visual metric features to handle high computational complexity caused by exploiting the objective metrics based on HVS. On the other hand, Rafal et al. [49] propose a visual metric, based on a visual model for all luminance conditions, for predicting visibility rather than quality. This approach could handle a wide range of luminance, unlike current metrics that see narrow intensity ranges, but there is a limitation that it does not consider color information.

Humans are the ultimate receivers of visual signals. The VQA of an image should consider human cognitive performance. Objective quality evaluation metrics cannot fully reflect the HVS of a 3D model. Thus, subjective VQA techniques are frequently used to verify the accuracy of objective VQA methods. The subjective VQA of an image and video is a combination of psychophysical responses to color, movement, texture, and context. These responses are directly evaluated and observed by human evaluators. A typical subjective quality evaluation method is the mean opinion score (MOS), expressed as a single rational number in the range 1–5, where 1 and 5 represent the lowest and highest perceived quality, respectively. ITU-R BT.500-13 is a well-known subjective VQA protocol [50] that defines several test methods. These test methods are categorized into single and double stimulation methods depending on how the evaluated content is presented to the observer. Such testing methods include the single-stimulus (SS), double-stimulus continuous quality scale, stimulus-compare, and single-stimulus continuous quality evaluation techniques. The ITU-R BT.2021 recommendation [51], another assessment protocol extended from ITU-R BT.500-13, includes the subjective assessment of stereoscopic 3DTV systems. According to the document, the subjective quality of 3D images was evaluated based on depth quality and visual comfort in addition to general image assessment techniques. In that study, the evaluators' color perception, visual acuity, and stereoscopic vision were also assessed.

Note that the subjective VQA of a PC is not standardized; however, it can be performed using existing recommendations, e.g., ITU-R BT.500-13. A subjective VQA can be performed by comparing the target PC data and the reference PC data. As shown in Figure 2, various PCs generated according



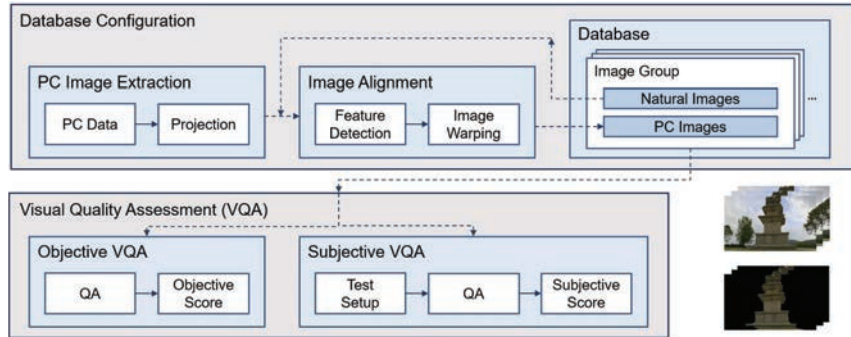
**Figure 2** Different rendering methods for “longdress” PC [9]: (a) original PC, (b) same PC with screen resolution multiplied, (c) PC with screen resolution multiplied three times and larger point size, and (d) PC after screened Poisson surface reconstruction.

to the resolution of the projected image, rendering methods, surface reconstruction methods, and scaling of the number of points can be evaluated [9]. ISO/IEC JTC1 SC29 WG1 joint photographic experts group introduces subjective VQA results according to the various levels of remaining points on several PC models [23]. Here, PCs of various quality levels are evaluated on 3D and 2D displays using the projection method. However, such subjective assessments are complex, expensive, elaborative, and incapable of evaluating real-time automatic systems. Thus, the goal of PC VQA research is the design of an objective quality evaluation metric consistent with the results of subjective human assessments.

### 3 Proposed PC VQA Framework

Generally, PC VQA evaluates the geometric distortion between the target PC and the reference PC. Geometric evaluation metrics focus on the difference in the corresponding data; thus, it is difficult to reflect the HVS in objective VQA techniques. Thus, this paper proposes an objective VQA method that uses natural images as a reference. Natural images are ubiquitous in modern society, which is likely to increase the demand for quality PC data. Thus, natural images are expected to work well as reference data for PC VQA.

Figure 3 illustrates the proposed PC VQA framework. As shown, PC data input to the framework are converted to a 2D image via a projection process in a virtual 3D space. The database comprises pairs of natural images and projected PC images. Then, objective and subjective VQA techniques use this



**Figure 3** Proposed PC VQA framework.

database to perform their evaluations. The database construction, objective VQA methods, and subjective VQA methods are described in the following sections.

### 3.1 Database

A PC comprises a set of individual 3D points. In addition to having a 3D  $(x, y, z)$  position, i.e., the spatial attribute, each point may also contain several other attributes, e.g., color, reflectance, and surface normal information. The proposed framework evaluates PC data relative to natural images, where the natural images play the role of an anchor in determining the degree of PC data quality in both the objective and subjective VQA techniques. The PC data must be represented as a 2D image with a similar screen view as the natural image to compare the 3D PC data with a 2D natural image. The target PC data can be projected into 2D space, resulting in a 2D image, referred to as a PC image in this paper.

To generate the PC image, the PC data are rotated and scaled in virtual 3D space to fit the rendering boundary. Then, the surface of the PC data is rendered using one of the surface reconstructions [16]. Note that the virtual illumination source and camera must be positioned sufficiently distant from the PC data to illuminate all the PC data uniformly. By controlling the virtual camera's parameters, e.g., position, pose, rotation, and viewing angle, a PC image with a similar screen view as the corresponding natural image can be obtained.

Note that 2D image quality evaluation metrics, e.g., PSNR and SSIM, are applied to the obtained PC images.

The VQA database comprises groups of images that are sets of PC images and corresponding natural images. The PC images can be produced by the PC data projection process and have various quality levels depending on the resolution, number of points, and surface reconstruction method. PC images of different quality levels may belong to an image group with a corresponding natural image. Assessment for various PC quality levels can provide useful guidance in optimizing PC-based systems.

For the composition of the image group, the natural image with a view similar to the PC images is required. For PC images, freedom of view is supported by controlling the virtual camera's parameters. However, there are limitations related to capturing a natural image with the desired view. For example, the view and color of a natural image may vary significantly according to capturing environmental conditions, e.g., location, time, and weather. Considering such limitations, it is recommended to acquire natural images together with PC data to match the natural and PC images as closely as possible. The difference between the generated PC and the natural images can be moderated through post-processing techniques, e.g., color calibration.

### **3.2 Objective Quality Assessment**

The processing sequence of the framework for objective VQA goes as follows: (1) generation of PC images from PC data; (2) post-processing by image alignment process; (3) evaluation in terms of objective VQA metrics. To this end, first, PC images are created through the projection of PC data in virtual 3D space. The screen view similarity between the PC image and the corresponding natural image can be increased by adjusting the virtual camera's parameters during the projection process. In the PC image projected through the virtual camera parameter adjustment it is difficult to accurately reflect the various camera information of the natural image view. In particular, the natural image view may include a wider area, e.g., the background, than the PC image view. The difference in co-located information caused by the discrepancy between the two images can significantly reduce the accuracy of the VQA. However, postprocessing techniques, e.g., image alignment and color calibration, can reduce the differences between images. For post-processing, the proposed objective PC VQA framework includes an image alignment process (Figure 3).

The image alignment process, which involves both feature detection and image warping, attempts to increase the matching accuracy of the region of interest (ROI) in the PC image and the corresponding region in the natural



image. First, an ROI in the PC image is set. Here, a feature detection algorithm, e.g., scale-invariant feature transform (SIFT) and speed-up robust features (SURF), is employed to increase the detection accuracy of the ROI and a corresponding region [52, 53]. This algorithm is also applied to the entire region of the natural image. Then, the corresponding regions of the two images are determined by matching the features. The corresponding regions may still have geometric differences; thus, the PC image is warped to minimize the mean square difference, which can realize well-matched areas.

When the two images are aligned, the proposed framework measures the quality of the matching images. Image VQA metrics generally interpret image quality as fidelity or similarity with a “reference” or “perfect” image in some perceptual space. Modeling the HVS has been considered the most suitable paradigm to achieve better quality evaluation. A visual signal has different characteristics, e.g., brightness, contrast, frequency components, and interactions between different signal components. The basic premise is that the visual system has different sensitivities depending on which aspects of the signal it perceives. When considering different sensitivities, it is sensible to calculate the strength of the error between the test and reference signals. The FR VQA methods are suitable for achieving consistency in quality evaluations by modeling the salient physiological and visual psychological characteristics of the HVS or measuring signal fidelity [18]. By reflecting the characteristics of the FR VQA, the proposed framework evaluates PC data in virtual space using a natural image captured by a camera, i.e., the reference image.

Objective VQA metrics reflect human perceptual behavior when computationally evaluating image quality. The MSE and PSNR are simple fidelity measures that are widely used. The MSE is the  $L_2$  norm of the arithmetic difference between the reference and test signals, and the PSNR is a transform of the MSE ( $PSNR = 10 * \log(max^2/MSE)$ ), where  $max$  is the maximum pixel value in the image. These metrics’ simplicity and mathematical convenience make them suitable for measuring image quality [54]. However, the correlation between the fidelity measures and the HVS is low for most applications. Over the past few decades, VQA research has been conducted to improve the PSNR. The SSIM is derived by hypothetically considering what constitutes a loss in a signal structure. It is hypothesized that distortions in an image come from variations in lighting, e.g., contrast or brightness changes, are nonstructural distortions. These distortions should be treated differently from structural distortions that image quality can be captured relative to three complementary aspects of information loss, i.e., correlation distortion,

contrast distortion, and luminance distortion [21]. In other words, the SSIM measures the similarity in luminance, contrast, and structural content. The multiscale structural similarity (MS-SSIM) and information content weighting structural similarity (IW-SSIM), which are extended from the SSIM, are estimated on multiple scales and contain a more advanced pooling strategy of the local SSIM map by considering local information content, respectively [55, 56]. The feature similarity index (FSIM) considers low-level features, e.g., phase congruency and gradient magnitude [57]. In this study, PC data are expressed as a 2D image through projection; thus, the above measures can be utilized effectively to evaluate quality.

### 3.3 Subjective Quality Assessment

In most subjective VQA techniques, test subjects observe an original image as the reference image and a degraded image as the test image. The proposed framework uses natural images as the reference image and PC images as the test image to evaluate the quality of a PC image using the MOS. The proposed subjective VQA is performed as shown in Figure 3. In the subjective VQA framework, the test configuration is a stage in which the subject, subjective VQA method, evaluation equipment, and training session are set according to the purpose and test data. The subjective VQA is performed in the test environment with the database.

For subjective VQA, several standards that define subjective experimental conditions for multimedia content, e.g., images and videos, can be applied. There is currently no specific recommendation for subjective VQA of PC data. PC data are represented in 3D virtual space; thus, subjective VQA could be conducted on actual 3D display devices. The ITU-R BT.2021 Recommendation describes the subjective VQA of stereoscopic 3DTV systems. However, 2D VQA differs from 3D VQA, where depth quality and visual comfort are considered important factors. The proposed framework attempts to measure the visual quality of the PC rather than realism in a 3D environment. Thus, to compare the natural reference images and projected PC images, we employ the subjective assessment methodology of video quality (SAMVIQ) for 2D images described in the ITU-R BT.1788 Recommendation [58].

SAMVIQ can use both explicit and hidden references, and the image can be displayed randomly or multiple times. Note that playing and scoring any sequence in any order is possible. In addition, each sequence can be replayed and scored again. The maximum viewing time of each sequence is 10–15 s.

All scenes must be evaluated to complete the experiment. The SAMVIQ method provides PC images randomly at three different quality levels after displaying a natural reference image in the proposed framework. The goal of the subjective VQA is to verify an objective evaluation method consistent with the subjective evaluation of PC images.

## **4 Experimental Results of the Database**

In this section, we introduce the database prepared to evaluate the proposed framework experimentally. Then, the experimental results of various objective VQA metrics are observed to find an evaluation method suitable for PC VQA using natural images as a reference. Finally, the correlations between the evaluated objective results and subjective VQA scores are compared to verify whether the proposed framework reflects the HVS. Note that we developed custom software to support PC viewing, objective and subjective VQAs, and PC data projection functionalities for our experiments.

### **4.1 Database**

This study captured PC data and natural image pairs using a FARO scanner and a digital camera to facilitate our experimental evaluation. The PC images were rendered from the acquired PC data using the 2D projection-based method. However, due to the surface properties (e.g., low or specular reflectance), occlusions, and sensor limitations, PC data obtained from acquisition devices, e.g., FARO scanners, inevitably suffers from noise contamination. The PC data also contains holes and outliers in the resulting models. In particular, areas without data in the PC image created from PC data comprising only 3D objects without background information can be identified easily. These problems seriously affect the generated PC image; thus, the similarity between a PC image and a natural image can decrease. Our database comprises various scenes with and without a background to study how these issues affect the VQA. In addition, one of three deep learning-based segmentation algorithms, i.e., the boundary-aware salient object detection, U2Net, or U2Netp algorithms, with the best visual completion is selected and applied to the natural images to separate the background and target object [59, 60]. As shown in Figure 4, the database contains 10 image groups comprising 5 image groups of the nonsegmented scene and five image groups of the segmented scene. Each image group contains three types of PC images and the corresponding natural image. The resolution of the natural and PC images

is  $1920 \times 1080$  pixels considering a balance between the influence of aliasing and image alignment complexity.

In this study, the PC images were generated as follows. The quality of PC data can be affected by the number of points, point size, and surface reconstruction algorithms. To observe the consistency between the objective and subjective VQAs, we produced three types of PC data, i.e., (1) PC data with a texture created using additional camera-captured images (Texture PC), (2) the original PC data (Original PC), and (3) degraded PC data that randomly eliminates 70–90% of the number of points (Degraded PC), as shown in Figure 4.

Note that the projected 2D PC image cannot be structurally identical to the corresponding natural image; thus, the projected 2D PC images are aligned via the image alignment process using corresponding natural images to reduce structural differences. The image alignment process is performed based on the matching features of two input images. In most experiments, the image alignment process improves the structural similarity between the two images. However, in some experiments not included in the database, an appropriate number of features for image alignment were not detected. In particular, for the images of a segmented scene, feature matching accuracy reduced significantly when aligning objects that exhibit repetitive patterns.

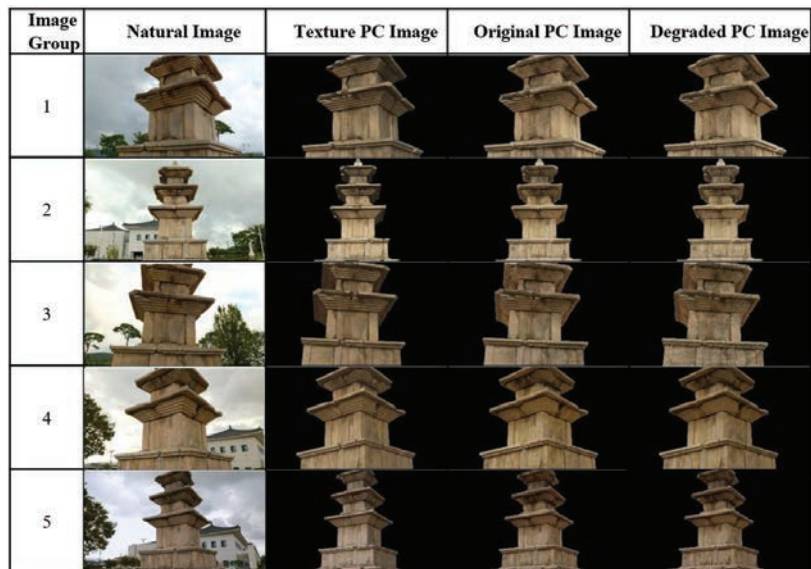
In contrast, nonsegmented scene images exhibited high feature-matching accuracy because many easily detectable features are included in the objects and background. Each level of PC data quality also slightly affects feature matching accuracy. The image alignment process was performed for many PC images, and then the PC images with an appropriate degree of completion for VQA were selected for the database.

## 4.2 Objective Quality Evaluation

In our experiments, 10 FR metrics were employed to perform objective VQA on our database, i.e., PSNR [54], SSIM [21], MS-SSIM [55], FSIM [57], IWSIM [56], VIF [17], MAD [61], GMSD [62], VSI [63], and NLPD [64]. The 10 FR metrics can be classified as follows. PSNR performs the point-to-point operation, and FSIM and VSI include the chrominance channels. SSIM, MS-SSIM, MAD, and GMSD compare the two images' global structural similarity or statistical difference, and IW-SSIM, VIF, and NLPD observe low-level features. This classification helps us understand which aspects the PC VQA should consider relative to reflecting the HVS. The scores for each



(a) Nonsegmented scenes



(b) Segmented scenes

**Figure 4** Database constructed for the proposed framework. Each image group comprises texture, original, and degraded PC images and the corresponding natural image.

**Table 1** Comparison of FR VQA scores observed on different PC qualities in nonsegmented scenes

Image Group	PC Quality	PSNR $\uparrow$	SSIM $\uparrow$	MS-SSIM $\uparrow$	FSIM $\uparrow$	IW-SSIM $\uparrow$	VIF $\uparrow$	MAD $\downarrow$	GMSD $\downarrow$	VSI $\uparrow$	NLPD $\downarrow$
1	Tex.	23.768	0.852	0.900	0.498	0.749	0.773	1068.141	0.281	0.870	2.732
	Orig.	20.621	0.748	0.816	0.435	0.653	0.789	1200.608	0.304	0.681	2.993
	Deg.	20.464	0.686	0.793	0.432	0.634	0.784	1511.303	0.313	0.608	3.101
2	Tex.	16.135	0.711	0.692	0.340	0.399	0.788	1329.691	0.302	0.783	6.149
	Orig.	13.915	0.582	0.509	0.301	0.180	0.782	1829.728	0.313	0.746	5.624
	Deg.	13.625	0.488	0.451	0.292	0.178	0.782	2199.199	0.322	0.717	4.338
3	Tex.	20.537	0.728	0.732	0.375	0.512	0.835	1261.521	0.298	0.842	4.988
	Orig.	20.066	0.675	0.704	0.389	0.509	0.807	1448.856	0.307	0.810	4.691
	Deg.	19.316	0.610	0.670	0.371	0.487	0.808	1804.755	0.315	0.794	5.351
4	Tex.	20.831	0.797	0.833	0.439	0.633	0.802	1349.037	0.287	0.726	2.783
	Orig.	18.196	0.656	0.718	0.344	0.472	0.804	1619.741	0.306	0.651	3.393
	Deg.	17.646	0.550	0.653	0.318	0.402	0.799	1934.721	0.321	0.699	3.494
5	Tex.	24.680	0.834	0.901	0.518	0.756	0.910	1070.609	0.270	0.831	3.036
	Orig.	20.680	0.721	0.812	0.441	0.644	0.855	1247.066	0.295	0.810	3.107
	Deg.	20.014	0.574	0.718	0.360	0.494	0.857	1747.097	0.309	0.736	3.839

**Table 2** Comparison of FR VQA scores observed on different PC qualities in segmented scenes

Image Group	PC Quality	PSNR $\uparrow$	SSIM $\uparrow$	MS-SSIM $\uparrow$	FSIM $\uparrow$	IW-SSIM $\uparrow$	VIF $\uparrow$	MAD $\downarrow$	GMSD $\downarrow$	VSI $\uparrow$	NLPD $\downarrow$
1	Tex.	19.627	0.685	0.697	0.413	0.331	0.777	1331.674	0.264	0.842	2.427
	Orig.	19.223	0.705	0.753	0.434	0.429	0.791	1283.216	0.261	0.836	2.109
	Deg.	19.018	0.685	0.745	0.423	0.419	0.785	1401.756	0.266	0.840	2.126
2	Tex.	22.989	0.875	0.923	0.574	0.681	0.838	1286.507	0.198	0.852	1.206
	Orig.	24.618	0.886	0.930	0.599	0.717	0.848	1194.368	0.192	0.852	1.186
	Deg.	23.668	0.859	0.912	0.545	0.643	0.840	1266.635	0.204	0.838	1.269
3	Tex.	21.202	0.724	0.754	0.401	0.417	0.787	1282.059	0.294	0.722	2.142
	Orig.	19.951	0.689	0.711	0.389	0.338	0.783	1342.390	0.286	0.744	2.316
	Deg.	19.574	0.663	0.713	0.403	0.365	0.781	1433.090	0.293	0.721	2.300
4	Tex.	20.587	0.711	0.713	0.420	0.368	0.837	1314.905	0.270	0.836	2.332
	Orig.	20.603	0.722	0.722	0.425	0.383	0.842	1247.568	0.266	0.845	2.307
	Deg.	20.661	0.723	0.735	0.433	0.403	0.835	1237.756	0.261	0.850	2.247
5	Tex.	21.814	0.815	0.856	0.502	0.562	0.798	1497.728	0.244	0.828	1.874
	Orig.	22.653	0.803	0.841	0.475	0.454	0.806	1305.927	0.227	0.863	2.012
	Deg.	23.597	0.806	0.855	0.484	0.486	0.803	1209.351	0.216	0.863	1.732

PC image type were collected and compared to the corresponding natural image.

Tables 1 and 2 show all scores obtained for the objective VQA. As shown in Tables 1 and 2, in most image groups of the nonsegmented scene the scores are distributed consistently in the order of texture PC, original PC, and degraded PC. The results demonstrate consistency in the objective VQA of the three PC image types. However, most image groups exhibit irregular

score distribution in the segmented scene. For example, most of the original PC images obtained the highest scores in image groups 1 and 2, and most degraded PC images obtained the best scores in image groups 4 and 5. The reasons for these inconsistent results relative to the segmented scene include the possibility of lower image alignment accuracy compared to the images of the nonsegmented scene. According to the features of different types of PC data generated in this study, the texture PC score may be lower than the original PC score depending on the texture quality or the performance of the surface reconstruction algorithm. Nevertheless, the degraded PC score should not be greater than the original PC image score because the degraded PC is obtained from the original PC by eliminating 70–90% of the points in the PC. Considering that the results exhibit low consistency in the segmented scene, and the degraded PC images occasionally achieved higher scores than other PC images, the scores for the segmented scene are excluded in the following analyses because the outcomes may confuse the remaining experimental results.

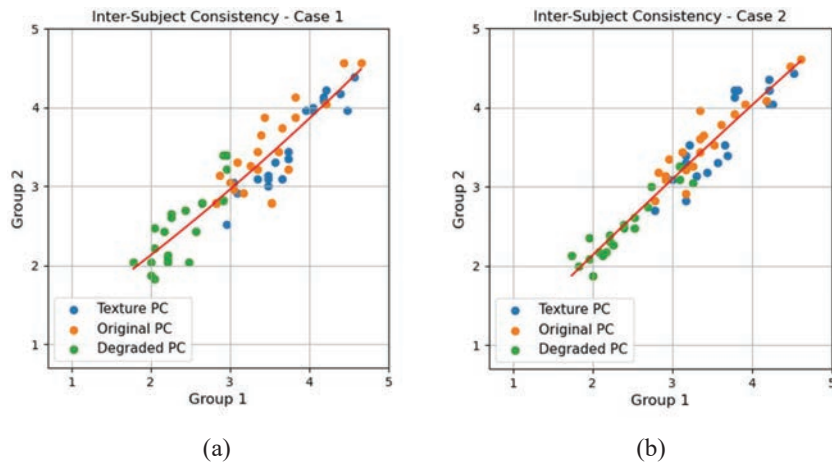
According to the experimental results of the nonsegmented scene, we can observe the properties of the various metrics employed in the proposed framework. As shown in the tables, SSIM, MS-SSIM, MAD, and GMSD, which evaluate the global structural or statistical difference within the luminance channel between two images, exhibit discernable score differences among the three PC image types and prove consistency. VIF and NLPD show similar scores overall and inconsistent trends according to the PC image quality. Regarding the PSNR results, the scores between the texture PC and original PC images are discriminative; however, the scores for the original PC images are similar to those of the degraded PC images. The PSNR results reveal that the difference in texture impacts the PSNR significantly. The remaining metrics demonstrated no meaningful change in the scores. Based on these observations and the characteristics of the 10 FR metrics, the metrics that measure global structural similarity are considered the application to the evaluation of PC quality using natural images as a reference. In the following, the objective and subjective VQA results are discussed to verify whether these metrics are valid in the HVS.

### **4.3 Subjective Quality Evaluation**

In our subjective VQA experiments, we recruited 46 participants (undergraduate students aged 20–30 years). All participants had normal or corrected-to-normal visual acuity and were unfamiliar with video quality assessment and

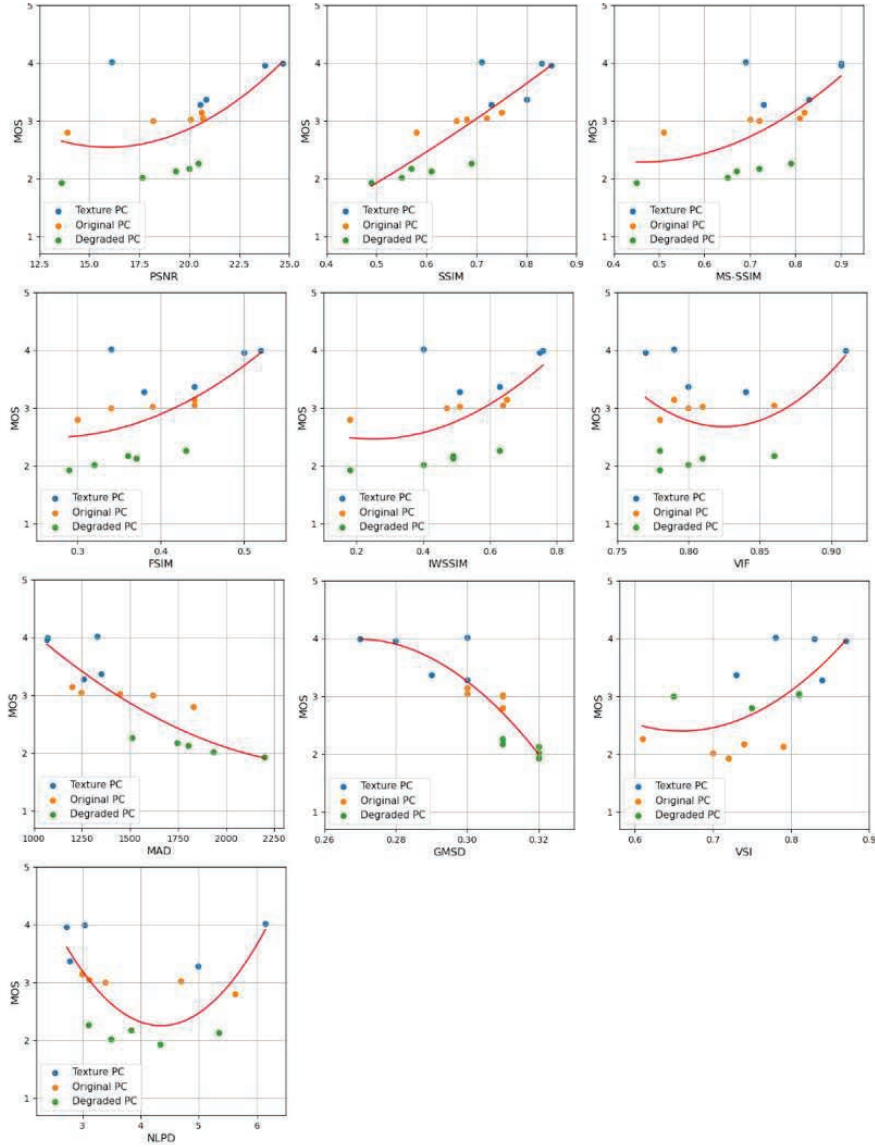
distortion. Before the evaluation, the participants were briefed on the subject and purpose of the experiment. Then, a training session was conducted to familiarize the participants with the subjective testing. Here, a 75-inch TV was used to display contents. The subjects were given scoring sheets to enter their quality scores, and the distance from the display device was determined autonomously within a range of 1–3 m. We evaluated the subjects' recorded ratings in terms of MOS. As Yize et al. [65] validate the subjective and objective quality assessment of 2D by evaluating the consistency between objective and subjective quality to evaluate 3D foveated video compression in virtual reality, this paper also proves our experiments with validation. Hence, the Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) were used to measure further the correlation between the MOS results and the correlation between individual objectives and the VQA metric against MOS [66].

Intersubject consistency experiments were conducted to guarantee the reliability of the subjective VQA results, as shown in Figure 5. The consistencies were measured twice by randomly dividing the results of the 46 participants into two subgroups. Figure 5 plots the MOS results, showing subgroups' average scores regarding PC images as points in the joint MOS–MOS space. We found that the average scores are adequate to indicate normality for 26 outcomes. Here, the subjective scores for the texture,



**Figure 5** Scatter plots of MOS from two nonoverlapping, equal-sized groups of subjects. (a) Intersubject consistency of the subjective quality assessment, SROCC = 0.910. (b) Intersubject consistency of the subjective quality assessment, SROCC = 0.950.





**Figure 6** Scatter plots of all objective FR VQA scores vs. MOSs for all the PC images with the nonsegmented scene. The red curve indicates the best-fitting polynomial function and subjective VQA scores.

**Table 3** Performance comparison of FR models on the database

Metric	Nonsegmented Scene		Segmented Scene	
	PLCC $\uparrow$	SROCC $\uparrow$	PLCC $\uparrow$	SROCC $\uparrow$
PSNR	0.479	0.614	0.125	0.016
SSIM	0.853	0.882	0.057	0.060
MS-SSIM	0.609	0.687	0.177	0.143
FSIM	0.594	0.635	0.149	0.064
IW-SSIM	0.516	0.576	0.212	0.148
VIF	0.191	0.099	0.077	0.047
MAD	0.854	0.871	0.030	0.120
GMSD	0.861	0.930	0.160	0.084
VSI	0.579	0.590	0.082	0.033
NLPD	0.080	0.282	0.274	0.236

original, and degraded PC images are shown using blue, orange, and green dots, respectively. For the two experiments illustrated in Figure 5, the SROCC scores measured using the MOS average are 0.91 and 0.95, respectively, demonstrating reasonably high consistency. The MOS scores for the degraded PC images are the lowest in Figure 5, similar to the objective VQA trend observed with the nonsegmented scene. However, the subjective VQA results do not exhibit much difference in quality between the texture and original PC images. In addition, the subjects rated the quality of several original PCs higher than the corresponding texture PCs, which was analyzed as an effect of the visual satisfaction of texturing. According to the consistency and similarity of the observation, we have verified the correlation between the proposed framework and the HVS by comparing both objectives.

To demonstrate the correlation between the observed MOS and the results of the 10 FR metrics, Figure 6 plots the average MOS and objective VQA scores to the corresponding points for all PC images of the nonsegmented scene. In addition, Table 3 shows the PLCC and SROCC scores for the correlation between each FR metric and the MOS given in Figure 6. As can be seen, the SSIM, MAD, and GMSD scores, which demonstrated the highest consistency, are considered the most appropriate metrics for the proposed PC VQA framework. The SROCC score of MS-SSIM was lower than the appropriate metrics; however, it was the highest among the other FR metrics. The remaining FR metrics indicate insufficient consistency in evaluating quality. In particular, the VIF and NLPD scores exhibited very low consistencies.

The most appropriate metrics compare the overall structural similarity or statistical differences within the luminance channel of the images. These metrics benefit PC quality evaluation using natural images as references, even in the objective VQA results. Thus, the proposed framework has verified that the global structural similarity between 2D images should be considered to evaluate PC quality reflecting the HVS. In addition, we found that minimizing structural limitations, e.g., empty space in PC data is advantageous for PC VQA.

## **5 Conclusions**

This paper has proposed a PC VQA framework using natural images as a reference. We constructed a database with three types of PC images and corresponding natural images to study the consistency of the PC quality assessment. The developed software produced the PC images using 2D-based projection, image alignment, and object-background segmentation to enhance structural similarity with the natural images. The proposed framework utilizes both natural and PC images; thus, the PC images were evaluated in terms of various FR metrics in objective VQA. In the objective VQA results, the proposed framework exhibited high consistency in the evaluation methods when comparing the global structural similarity of images with more accurate features, including the background. However, PC images representing 3D objects without background demonstrated remarkably low similarity to the natural images. In other words, there is a limitation in that low accuracy of the image alignment affects inconsistent results; thus, further research is required to improve the image alignment. In the subjective VQA results obtained from 46 participants, the scores of each metric and the average MOS scores were compared to confirm how effectively the objective VQA reflected the HVS. The experimental results exhibited strong consistency and were verified in the objective VQA, demonstrating a remarkably high correlation with the subjective VQA. Thus, we believe the proposed framework reflects the HVS through highly accurate image alignment and structural similarity between the PC and natural reference images.

Recently, since the emerging high dynamic range (HDR) point cloud [67] can provide us with a dominantly improved viewing experience, the proposed framework needs further study from a perceptual and HDR perspective. The performance of the proposed method could also be further analyzed by building various databases, including PC data and corresponding natural images.

## Acknowledgments

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [22ZH1200, the research of the fundamental media contents technologies for hyper realistic media space] and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No.2020-0-00452, Development of Adaptive Viewer-centric Point cloud AR/VR(AVPA) Streaming Platform].

## References

- [1] D. J. Brady, W. Pang, H. Li, et al., “Parallel cameras,” *Optica*, vol. 5, no. 2, pp. 127–137, Feb 2018.
- [2] S. Schwarz, M. Preda, V. Baroncini, et al., “Emerging mpeg standards for point cloud compression,” *IEEE J. Emerg. Sel. Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, Dec 2018.
- [3] K. Khoshelham, and S. O. Elberink, “Accuracy and resolution of kinect depth data for indoor mapping applications,” *Sensors (Basel)*, vol. 12, no. 2, pp. 1437–1454, 2012.
- [4] H.-G. Maas, A. Bienert, S. Scheller, E. Keane, “Automatic forest inventory parameter determination from terrestrial laser scanner data,” *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1579–1593, 2008.
- [5] D. D. Lichti, and M. G. Licht, “Experiences with terrestrial laser scanner modelling and accuracy assessment,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 36, no. 5, pp. 155–160, 2006.
- [6] M. Berger, A. Tagliasacchi, L. M. Seversky, et al., “A survey of surface reconstruction from point clouds,” *Comput. Graph. Forum*, vol. 36, no. 1, 301–329, 2017.
- [7] Available at: <http://graphics.stanford.edu/data/3Dscanrep/>.
- [8] G. Turk, and M. Levoy, “Zippered polygon meshes from range images,” *Proc. SIGGRAPH '94, Computer Graphics Proceedings, Annual Conference Series*, 1994, pp. 311–318.
- [9] E. Dunic, C. R. Duarte and L. A. da Silva Cruz, “Subjective evaluation and objective measures for point clouds – State of the art,” *2018 First International Colloquium on Smart Grid Metrology (SmaGriMet)*, 2018, pp. 1–5.
- [10] K. Mammou, P. A. Chou, D. Flynn, et al., *G-PCC codec description v2 ISO/IEC JTC1/SC29/WG11*, vol. N18189, 2019.

- [11] M. Committee, “V-PCC Codec Description JTC1/SC29/WG11 w19526,” Italy: Virtual, Document ISO/IEC, Sep. 2020.
- [12] *JPEG Pleno Point Cloud Coding Common Test Conditions ISO/IEC JTC1/SC29/WG11*, vol. N18474, 2019.
- [13] “Final call for evidence on JPEG pleno point cloud coding” in ISO/IEC JTC1/SC29/WG1 JPEG Output Document, vol. N88014, Jul. 2020.
- [14] S. Park, J. Kim, Y. Hwang, et al., “Enhancement of 3D point cloud contents using 2D image super resolution network,” *Journal of Web Engineering*, vol. 21, pp. 425–442, 2022.
- [15] K. Wolff et al., “Point cloud noise and outlier removal for image-based 3D reconstruction,” *IEEE Fourth Int. Conf. on 3D Vis., Stanford, CA, USA, Oct. 2016*.
- [16] M. Berger, A. Tagliasacchi, L. Seversky, et al., “A survey of surface reconstruction from point clouds,” *Comput. Graph. Forum*, vol. 36, no. 1, Mar 2016.
- [17] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “Subjective and objective quality evaluation of 3d point cloud denoising algorithms,” *ISO/IEC JTC m75024. Sydney, Australia, Mar. 2017*.
- [18] H. R. Sheikh, and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 4–2298, Sep 2007.
- [19] Z. Wang, and A. C. Bovik, “Mean squared error: Love it or leave it? – A new look at signal fidelity measures,” *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan 2009.
- [20] E. Alexiou, “On subjective and objective quality evaluation of point cloud geometry,” in *9th Int. Conf. Qual. of Multimedia Experience (QoMEX’17)*. IEEE, pp. 1–3, 2017.
- [21] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [22] D. Tiang, H. Ochimizu, C. Feng, et al., “Geometric distortion metrics for point cloud compression,” *IEEE Intl Conf. on Image Process., Sep. 2017*.
- [23] S. Perry, A. Pinheiro, E. Dunic, et al., “Study of subjective and objective quality evaluation of 3D point cloud data by the JPEG committee,” *Electronic Imaging*, vol. 31, no. 10, pp. 312–316, 2019.
- [24] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, “Subjective and objective quality evaluation of 3D point cloud denoising algorithms,” *IEEE Intl Conf. on Multimedia Expo Workshops (ICMEW)*, pp. 1–6, Jul. 2017.

- [25] Q. Yang, Z. Ma, Y. Xu, et al., “Inferring point cloud quality via graph similarity,” *arXiv Preprint ArXiv:2006.00497*, 2020.
- [26] S. Chen, D. Tian, C. Feng, et al., “Fast resampling of three-dimensional point clouds via graphs,” *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 666–681, 2017.
- [27] S. Yan, Y. Peng, G. Wang, et al., “Weakly supported plane surface reconstruction via plane segmentation guided point cloud enhancement,” *IEEE Access*, vol. 8, 60491–60504, 2019.
- [28] Y. Regaya, F. Fadli, A. Amira, “3d point cloud enhancement using unsupervised anomaly detection,” *Int. Symp. on Syst. Eng. (ISSE’19), vol. 2019*. IEEE, 2019, pp. 1–6.
- [29] T. Zheng, C. Chen, J. Yuan, et al., “Pointcloud saliency maps,” *Proc. IEEE Int. Conf. Computer Vision (ICCV’19)*, 2019, pp. 1598–1606.
- [30] T. Hackel, N. Savinov, L. Ladicky, et al., “Semantic3d. net: A new large-scale point cloud classification benchmark,” *arXiv Preprint ArXiv:1704.03847*, 2017.
- [31] Z. Zhang, L. Zhang, X. Tong, et al., “A multilevel point-cluster-based discriminative feature for als point cloud classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3309–3321, 2016.
- [32] T. Rabbani, F. Van Den Heuvel, G. Vosselmann, “Segmentation of point clouds using smoothness constraint,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 36, no. 5, pp. 248–253, 2006.
- [33] R. Schnabel and R. Klein, “Octree-based point-cloud compression,” *SPBG*, vol. 6, pp. 111–120, 2006.
- [34] Y. Shao, Q. Zhang, G. Li, et al., “Hybrid point cloud attribute compression using slice-based layered structure and block-based intra prediction,” *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 1199–1207.
- [35] S. Gumhold, Z. Kami, M. Isenburg, H.-P. Seidel, “Predictive point-cloud compression,” *ACM SIGGRAPH, 2005 Sketches*, pp. 137–es, 2005.
- [36] L. Li, Z. Li, S. Liu, H. Li, “Occupancy-map-based rate distortion optimization and partition for video-based point cloud compression,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, 326–338, 2020.
- [37] E. Alexiou and T. Ebrahimi, “Towards a point cloud structural similarity metric,” *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, pp. 1–6, 2020.
- [38] B. An and Y. Kim, “Image link through adaptive encoding data base and optimized GPU algorithm for real-time image processing of artificial intelligence,” *Journal of Web Engineering*, 2022, vol. 21, pp. 459–496.

- [39] W. Zhou, Q. Yang, Q. Jiang, et al., “Blind quality assessment of 3D dense point clouds with structure guided resampling,” *arXiv* 2022, arXiv:2208.14603.
- [40] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Trans. Broadcast.*, vol. 3, pp. 312–322, Sep. 2004.
- [41] A. K. Moorthy, C. Su, A. Mittal, and A. C. Bovik, “Subjective evaluation of stereoscopic image quality,” *Signal Processing: Image Communication*, vol. 28, no. 8, 870–883, 2013.
- [42] E. Dumic and S. Grgic, “Reduced video quality measure based on 3D steerable wavelet transform and modified structural similarity index,” *Proc. of the 55th Intl Symp. ELMAR-2013*, 2013, pp. 65–69.
- [43] R. Soundararajan and A. C. Bovik, “Video quality assessment by reduced reference spatio-temporal entropic differencing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.
- [44] F. Battisti, E. Bosc, M. Carli, et al., “Objective image quality assessment of 3D synthesized views,” *Signal Process. Image Commun.*, vol. 30, pp. 78–88, Jan 2015.
- [45] E. Dumić, S. Grgić, K. Šakic, P. M. R. Rocha, L. A. da Silva Cruz, “3D video subjective quality: A new database and grade comparison study,” *Multimedia Tool. Appl.*, vol. 76, no. 2, pp. 2087–2109, Jan. 2017.
- [46] E. M. Torlig, E. Alexiou, T. A. Fonseca, et al., “A novel methodology for quality assessment of voxelized point clouds,” *SPIE vol. 10752, Applications of Digital Image Processing XLI. San Diego, CA, USA, 2018*.
- [47] Q. Liu, H. Yuan, H. Su, et al., “PQA-Net: Deep no reference point cloud quality assessment via multi-view projection,” *IEEE Transactions on Circuits and Systems for Video Technology* 2021, 31, 4645–4660.
- [48] A. Artusi, F. Banterle, F. Carra, A. Moreno, “Efficient evaluation of image quality via deep-learning approximation of perceptual metrics,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1843–1855, 2019.
- [49] R. Mantiuk, K. Kim, A. G. Rempel, W. Heidrich, “HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions,” *ACM Transactions on Graphics*, vol. 30, pp. 1–14, 2011.
- [50] ITU-R BT.500-13 “Methodology for the subjective assessment of the quality of television pictures,” *International Telecommunication Union. ITU Radiocommunication Sector*, 2012.

- [51] ITU-R BT.2021, “Subjective methods for the assessment of stereoscopic 3DTV systems,” *International Telecommunication Union*. ITU Radiocommunication Sector, 2012.
- [52] D. G. Lowe, “Distinctive image features from scale-invariant Keypoints,” *Intl. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [53] H. Bay, T. Tuytelaars, L. Van Gool, “Surf: Speeded up robust features,” *Eur. Conf. on Comput. Vis.*, May 2006.
- [54] G. Bjøntegaard, “Calculation of average PSNR differences between RD curves,” *ITU-T*, vol. Q6, no. 16, Apr, document VCEG-M33, 2001.
- [55] Z. Wang, E. P. Simoncelli, A. C. Bovik, “Multiscale structural similarity for image quality assessment,” *Proc. IEEE Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov 2003, vol. 2, pp.1398-1402.
- [56] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, Nov 2011.
- [57] L. Zhang, L. Zhang, X. Mou, D. Zhang, “FSIM: A feature similarity index for image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.
- [58] ITU-R Recommendation BT. 1788, “Methodology for the subjective assessment of video quality in multimedia application,” 2007.
- [59] X. Qin, Z. Zhang, C. Huang, et al., “BASNet: Boundary-aware salient object detection,” *Proc. IEEE/CVF Conf. Comput. Vis., Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7479–7489.
- [60] X. Qin, Z. Zhang, C. Huang, et al., “U2-net: Going deeper with nested U-structure for salient object detection,” *Pattern Recognit.*, vol. 106, no. Oct, Art. no. 107404, 2020.
- [61] E. C. Larson and D. M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” *J. Electron. Imag.*, vol. 19, pp. 011006:1-011006:21, Jan 2010.
- [62] W. Xue, L. Zhang, X. Mou, A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
- [63] L. Zhang, Y. Shen, H. Li, “VSI: A visual saliency-induced index for perceptual image quality assessment,” *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug 2014.
- [64] V. Laparra, J. Ballé, A. Berardino, E. P. Simoncelli, “Perceptual image quality assessment using a normalized Laplacian pyramid,” *Electronic Imaging*, vol. 28, no. 16, pp. 1–6, Feb 2016.



- [65] Y. Jin, M. Chen, T. Goodall, et al., “Subjective and objective quality assessment of 2d and 3d foveated video compression in virtual reality,” *IEEE Transactions on Image Processing*, vol. 30, pp. 5905–5919, 2021.
- [66] ITU-T P.1401, “Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models,” *International Telecommunication Union*, Jul 2012.
- [67] C. Loscos, P. Souchet, T. Barrios, et al., “From capture to immersive viewing of 3D HDR point clouds,” *Eurographics, Reims, France*, 25-29 April 2022.

## Biographies



**Aram Baek** received B.Sc., M.Sc., and Ph.D. degrees in multimedia engineering in 2012, 2014, and 2019, respectively, from Hanbat National University, Daejeon, Republic of Korea. From 2019 to 2020, he was a Senior Researcher for Pixtree, Seoul, Republic of Korea. From 2020 to 2022, he was a postdoctoral fellow at Hanbat National University, Daejeon, Republic of Korea. He is currently a Senior Researcher for Intekmedi, Sejong, Republic of Korea. His research interests include image processing, video coding, and computer vision.



**Minseop Kim** received a B.Sc. degree from the Department of Information and Communication Engineering from Hanbat National University, Daejeon,

Republic of Korea, in 2018, where he also received a M.Sc. degree in the Department of Multimedia Engineering. He is currently working toward a Ph.D. degree in the Department of Multimedia Engineering at Hanbat National University. His research interests include computer vision, machine learning, and parallel processing.



**Sohee Son** received B.Sc. and M.Sc. degrees in the department of multimedia engineering from Hanbat National University, Daejeon, Korea, in 2015 and 2017, respectively. She is currently pursuing a Ph.D. degree in the department of multimedia engineering at Hanbat National University. Her research interests include image processing and computer vision.



**Sangwoo Ahn** received his B.Sc. and M.Sc. degrees in Electronics Engineering from Kyunghee University, Suwon, Republic of Korea, in 1997 and 1999, respectively. He studied for his Ph.D. degree at Chungnam National University, Daejeon, Republic of Korea in 2011. Since 1999, he has worked as a Senior Member of the Research Staff at the Immersive Media Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. His research interests include image processing, video processing, and realistic broadcasting media service technology.



**Jeongil Seo** was born in Goryoung, Republic of Korea, in 1971. He received a Ph.D. degree in Electronics from Kyungpook National University, Daegu, Republic of Korea, in 2005 for his work on audio signal processing systems. He worked as a member of the engineering staff at the Laboratory of Semiconductor, LG-semicon, Cheongju, Republic of Korea, from 1998 until 2000. Since 2000, he has worked as a Director at the Immersive Media Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea,. His research activities include image and video processing, audio processing, and realistic broadcasting and media service systems.



**Hui Yong Kim** received his BS, MS, and PhD degrees from KAIST in 1994, 1998, and 2004, respectively. From 2003 to 2005, he worked for AddPac Technology Co. Ltd., Seoul, Rep. of Korea as the Leader of the Multimedia Research Team. From 2005 to 2019, he joined ETRI, Daejeon, Rep. of Korea and served as the Managing Director of Realistic Audio and Video Research Group. Since 2020, he has been with the Department of Computer Science and Engineering in Kyung Hee University, Yongin, Rep. of Korea, as an Associate Professor. He has been an active technology contributor, editor, and ad-hoc group chair in developing several international standards

including MPEG Multimedia Application Format, ITU-T/ISO/IEC JCT-VC High Efficiency Video Coding and JVET Versatile Video Coding. His current research interests include image and video signal processing and compression for realistic media applications such as UHD, 3D, VR, HDR, and digital holograms.



**Haechul Choi** received his B.S. in electronics engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1997, and his M.S. and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 1999 and 2004, respectively. He is a professor in information and communication engineering at Hanbat National University, Daejeon, Korea. From 2004 to 2010, he was a Senior Member of the Research Staff in the Broadcasting Media Research Group of the Electronics and Telecommunications Research Institute (ETRI). His current research interests include image processing, video coding, and computer vision.