

Research Article

A Layered Approach for Quality Assessment of DIBR-Synthesized Images

Rafia Mansoor, Muhammad Shahid Farid , Muhammad Hassan Khan , and Asma Maqsood

Department of Computer Science, University of the Punjab, Lahore 54590, Pakistan

Correspondence should be addressed to Muhammad Shahid Farid; shahid@pucit.edu.pk

Received 9 June 2021; Revised 4 October 2021; Accepted 20 October 2021; Published 13 November 2021

Academic Editor: Salvatore Serrano

Copyright © 2021 Rafia Mansoor et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiview video plus depth (MVD) is a popular video format that supports three-dimensional television (3DTV) and free viewpoint television (FTV). 3DTV and FTV provide depth sensation to the viewer by presenting two views of the same scene but with slightly different angles. In MVD, few views are captured, and each view has the color image and the corresponding depth map which is used in depth image-based rendering (DIBR) to generate views at novel viewpoints. The DIBR can introduce various artifacts in the synthesized view resulting in poor quality. Therefore, evaluating the quality of the synthesized image is crucial to provide an appreciable quality of experience (QoE) to the viewer. In a 3D scene, objects are at a different distance from the camera, characterized by their depth. In this paper, we investigate the effect that objects at a different distance make on the overall QoE. In particular, we find that the quality of the closer objects contributes more to the overall quality as compared to the background objects. Based on this phenomenon, we propose a 3D quality assessment metric to evaluate the quality of the synthesized images. The proposed metric using the depth of the scene divides the image into different layers where each layer represents the objects at a different distance from the camera. The quality of each layer is individually computed, and their scores are pooled together to obtain a single quality score that represents the quality of the synthesized image. The performance of the proposed metric is evaluated on two benchmark DIBR image databases. The results show that the proposed metric is highly accurate and performs better than most existing 2D and 3D quality assessment algorithms.

1. Introduction

In everyday life, humans gain a lot of information through magazines, television, videos, images, etc. along with capturing, viewing, receiving, sending, and utilizing the information [1].

With the already presence of two-dimensional (2D) technologies, different three-dimensional (3D) technologies have also been introduced to the customers since the past few years mainly through cinemas, gaming, 3D televisions (3DTV) [2], and free viewpoint television (FTV) [3]. This huge utilization and demand of visual applications initiated the research of assessing the image quality that a consumer receives. Therefore, for the past few decades, image quality assessment (IQA) of both 2D and more recently 3D technologies has been a major focus of many researchers.

Subjective quality assessment, done through the human visual system (HVS), is considered to be the most accurate form of assessment because here humans are asked to view the images and videos and provide their opinion about the quality.

However, image evaluation done subjectively can be expensive, time-consuming, and in many cases not possible [2]. In some cases, subjective experimentation can get further complicated because of many factors that include subjects' physiological status (vision accuracy, binocular rivalry), psychological status (their emotions, mood), and certain environmental factors (distance from viewing display device, lighting conditions) [4].

Therefore, it has become the necessity of researchers to create certain mathematical models that can be able to predict the image quality for humans. IQA done objectively

includes certain computational algorithms that are usually designed to predict the quality automatically and accurately while maybe using some data from subjective assessment [5]. In order to evaluate still images and videos, many researchers use a direct or modified version of 2D quality metrics. However, 3D image quality assessment is still a new area to explore.

Depending on the 3D video application, different 3D video formats are available, e.g., stereo video, video plus depth (V+D), layered depth video (LDV), multiview video (MVV), and multiview video plus depth (MVD) [6, 7]. In MVD format, multiple color views and their corresponding depth sequences are used to generate virtual views using depth image-based rendering (DIBR) [8]. The 3D technologies and FTV allow the user to control the viewpoint in the scene. It creates a depth perception of any scene with simultaneous display of numerous different views to provide seamless horizontal parallax. Practically, capturing, coding, and transmitting these various numbers of views at the same time are not possible due to certain constraints of hardware, software, and processes not being cost-effective.

Therefore, depth image-based rendering (DIBR) techniques can be used to generate additional views with the presence of limited images. DIBR algorithms help convert 2D monocular images into 3D stereoscopic image [9, 10]. DIBR consists of two stages of the process named warping and rendering. Errors in the rendering process can typically include image edge misalignment or displacements [11], boundary blur, or black hole [12]. Some of the artifacts are listed in the following.

- (i) Motion blur: it is due to low light conditions
- (ii) Ghosting effect: it is due to misalignment of the two views being fused. It can also appear due to the repeated reflection of light from the surface of the lens and is seen in an image as a shadow
- (iii) Binocular rivalry: perception alternates between different images presented to each eye [13]
- (iv) Keystone distortion: it results in a trapezoid shape, and it affects the geometric relation and can break the 3D effect of a stereo video
- (v) Cardboard effect: unnatural flattening of objects occurs in an image and creates inappropriate depth scaling
- (vi) Staircase effect: it affects diagonal edges of an image
- (vii) Crosstalk effect: distortion occurs due to display imperfections

The quality of views which are synthesized using DIBR algorithms is mainly determined by depth of the image and quality of the texture [14].

2. Related Work

In literature, many computational algorithms based on stereoscopic and synthesized IQA have been documented

[15]. Most of the early 3D-IQA algorithms are extended from 2D-IQA algorithms. For example, Ref. [16] proposed an algorithm which is known as View Synthesis Quality Assessment (VSQA) metric.

It starts with a 2D image quality metric to find the distortion or similarity signals between DIBR-synthesized views and reference views. Later, three weighting maps are formed including image contrast, gradient orientations, and textural complexity.

Based on PSNR, Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) [12] helped find and estimate the structural geometric distortion present in the DIBR-synthesized image in order to find out the image's quality. MSSIM [17] is the extended algorithm of a 2D-IQA known as structural similarity index (SSIM) [18]. By using this algorithm, structural information of an image can be obtained by separating the effect of illumination. The process is completed in three steps. The first step includes luminance comparison, the second step involves contrast comparison, and lastly structural comparison of original and synthesized view is done. Based on DIBR, with the help of 3D Video Quality Measure (3VQM) [19], the first distortion-free depth estimation method was established. The quality of the video was determined by temporal and spatial variations which were estimated by comparison between the given depth map with the ideal depth map.

Li et al. [14] proposed a method to assess the quality of synthesized views. This method worked on the preliminary idea that distortions in depth images create changes in the edge regions of the image. Three steps involved in the method are the generation of similarity map, weighting map, and finally edge-guided pooling is formed. Another full reference metric was formed to assess the quality of the image synthesized by DIBR [20]. It proposed to compare the original view with the edges of the synthesized view. Restricted to the structural distortion only, this algorithm ignores the color distortion while evaluating the quality of an image. The quality metric presented in [21] estimates the geometric distortions and sharpness in the synthesized image to assess its quality. The geometric distortions are estimated by analyzing local similarities in the disoccluded regions. The sharpness is measured globally using the synthesized image and its downsampled version. Battisti et al. [22] researched on 3D Synthesized view Image quality Metric (3DSwIM). This metric compares certain statistical features of wavelet subbands of reference views and DIBR-synthesized views. This algorithm is working on the assumption that humans are prone to make distortions when they are surrounded by other humans. The 3D-IQA algorithm presented in [23] captures the textural and structural distortions in the DIBR-synthesized image to estimate its quality.

There are quality metrics designed on the fact that there is a binocular asymmetry between a human's right and left eyes which disable humans to form a single binocular image. Critical Binocular Asymmetry (CBA) metric [24] used DIBR techniques to assess the quality of the image objectively. This method detects the critical areas where excessive binocular asymmetry is induced. Then, structural similarity is calculated in those critical areas. A 3D no-reference objective

metric was proposed [25] to assess the quality of virtual views. With the help of cyclopean eye theory, to measure the quality of the synthesized image, the proposed metric compares the characteristics of a cyclopean image with the produced image. In [26], an NR algorithm was presented to assess the quality of synthesized images. With the help of this method, synthesis distortion is calculated by using left and right views. It takes an original image and generates a virtual image which is used to assess the distortion that is created during the process of DIBR. A similar research [27] proposed an algorithm which was a combination of two metrics; one is used to assess the quality of synthesized images, and the second metric is responsible to assess the quality of depth maps. The proposed SIQM metric utilizes the phenomenon of cyclopean eye theory.

A no-reference 3D quality assessment metric presented in [28] proposed a natural scene statistics (NSS) model that captures the geometry distortions in the synthesized image to predict its quality. A metric proposed in [29] is known as the No-reference Image Quality assessment of Synthesized Views (NIQSV). Based on morphological operations, the algorithm integrates the distortions in saturation, contrast, and luminance.

Then, all these distortions are integrated into a single color weight factor. Another no-reference image quality assessment method for DIBR-synthesized images is presented in [30]. This method known as NIQSV+ estimates the DIBR-introduced distortions such as blurry regions, black holes, and stretching to predict the quality of the image. Tsai et al. [31] proposed an IQA based on DIBR techniques. The model was used to analyze the quality of the synthesized image made by distorting the depth map. Gaussian noise, quantization, and offset distortions were used. Consistent pixel shifts were eliminated inside the image and then rendering of an image was done.

In the existing 2D and 3D-IQA algorithms, quality computed at pixel or region level contributes equally to get an overall quality score. However, studies have shown that some regions or objects in an image attract more attention of the viewer than others; they are referred to as salient regions [32, 33]. In this paper, we investigate the impact of saliency on the overall quality of the synthesized image. Motivated by the findings of this investigation, we propose a DIBR-synthesized image quality assessment metric that finds the salient regions in the image with the help of the corresponding depth map, and based on the saliency, each pixel or region contributes differently to the final quality of the image.

The rest of the paper is arranged as follows: the proposed metric is presented in Section 3, the experimental evaluations and results are discussed in Section 4, and the conclusions of this research are presented in Section 5.

3. Proposed Method

Estimating the quality of the synthesized images is of paramount importance to provide a better viewing experience to the 3DTV and FTV viewers. Most existing quality metrics estimate the quality of individual pixels or groups of pixels

and combine these estimates to obtain a single quality score of the synthesized image. However, various studies, e.g., [32–34], have shown that each pixel or region is of different importance in visual perception. The objects which are closer to the viewer are more focused upon and attract more attention of the viewer as compared to the far objects [35]. In Figure-Ground Principle [36], the figure is considered the positive space, and the ground is considered as negative space or the background, i.e., the surrounding area on which the figure is placed upon. According to these studies, one can conclude that the foreground objects are crisper and more eye-catching as compared to the objects in the background. We exploit this phenomenon to design a quality metric that segments the image into multiple regions, each with different visual importance. These regions are termed layers in the rest of the text. The quality of each layer is computed independently, and they are merged in such a way that each layer based upon its visual importance contributes differently in this merge. The layers with high visual importance should contribute more than the low saliency layers. The working of the proposed algorithm can be divided into two steps.

First, the synthesized image is divided into layers, and second, the quality of each layer is computed and pooled to obtain a single quality indicator. These steps are described in the following sections.

3.1. Image Layering. Numerous techniques have been proposed to segment the image into visually important regions and rank them accordingly, e.g., [32–34, 37]. In our case, the depth map of the image is also available which contains the geometrical information of the scene. Specifically, a depth map is a grayscale image whose values range between 0 and 255 [9]. These values are inversely coded, that is, the farthest object has depth 0 and the closest has 255. Figure 1 shows a sample synthesized image (Figure 1(a)) and its corresponding depth map (Figure 1(b)). Note that the two persons in this scene are closest to the camera and have depth values close to 255; the depth values of the rest of the scene are quite low and fall in the lower end of the depth range. Each object in an image is at a certain distance from the camera, and therefore, all its pixels have the same depth values or the variation in their depth is rather limited. We exploit this fact to divide the image into layers, where each layer corresponds to the pixels having similar depth. For example, the two persons in Figure 1(b) have similar depth values and therefore can be put into the same layer.

We propose a histogram-based algorithm to compute the image layers. The depth values of an object in an image are in a limited range, concentrated around its mean depth value, and thus form a peak in the histogram. Such a peak appears for each object or a set of objects having similar depth values. So we can identify the layers of the image by computing the histogram of the depth map and finding the peaks or the regions between the valleys in the histograms. Figure 2 shows the histogram of the depth map shown in Figure 1(b). The histogram shows two peaks, first from 0 to m_1 and second from m_1 to n . All the image pixels with depth values in $[0, m_1)$ form a layer, and the pixels with



FIGURE 1: A sample image and its corresponding depth map from LoveBird1 MVD sequence: (a) synthesized image, view 8 frame 112, (b) corresponding depth image.

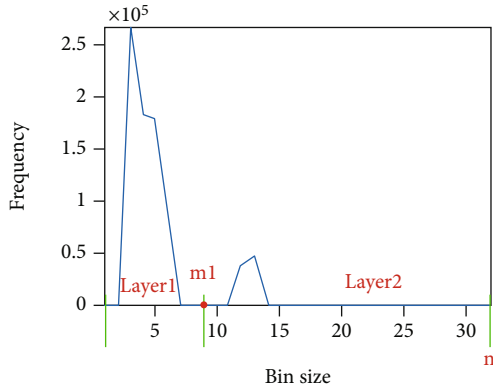


FIGURE 2: Histogram of the depth map shown in Figure 1(b).

depth values in $[m_1, 255]$ form the second layer. Figure 3 shows these two layers. Figure 3(a) is the sample input image (Figure 1(a)), Figure 3(b) shows the first layer, and Figure 3(c) shows the second layer of the image obtained through the proposed layering strategy.

Let I_w be a synthesized color image of size $M \times N$ and I_d be its depth map of the same size. To divide the image I_c , a histogram of I_d is computed with b bin size. To find the layers, the local minimas in the histogram are found. Let $m_1, m_2, m_3, \dots, m_n$ be the n local minimas. Then, the image pixels with depth values in 0 to m_1 are marked as layer 1 and those with depth values between m_1 and m_2 are marked as layer 2 and so on. Specifically, we define a layer map L as

$$L(i, j) = \begin{cases} 1, & \text{if } 0 \leq I_d(i, j) < m_1, \\ \vdots & \\ k, & \text{if } m_{k-1} \leq I_d(i, j) < 256, \\ \vdots & \\ n+1, & \text{if } m_n \leq I_d(i, j) < 256. \end{cases} \quad (1)$$

This means the image pixel $I_w(i, j)$ belongs to layer $L(i, j)$. The pixels of a specific layer k can be obtained as $\{I_w(i, j) \mid L(i, j) = k\}$. Figure 4 shows another example layer estimation, where five layers are detected. Figure 4(a) shows a sample synthesized image of Book_Arrival sequence, Figure 4(b) shows the corresponding depth map, and

Figure 4(c) shows the histogram and the detected layers in the image.

3.2. Estimating the Image Quality. After segmenting the synthesized image I_w into n layers, the next step is to compute the quality of I_w . To this end, the quality of each layer is estimated by comparing the layer image pixels with the corresponding pixels in the reference image I_r . Any existing image quality metric can be used for this purpose. However, from experiments, we found that WSNR [38] performs the best with the proposed framework; a detailed discussion is presented in Section 4.6. Let $Q_1, Q_2, Q_3, \dots, Q_n$ be the quality scores for n layers. We aggregate these scores in a weighted manner in order to get the overall quality score Q of the synthesized image I_w .

$$Q = c_1 Q_1 + c_2 Q_2 + \dots + c_n Q_n, \quad (2)$$

where c_1, c_2, \dots, c_n are the weights assigned to each layer. We have performed different experiments with the above described methodology and also with a fixed number of layers.

That is instead of automatically detecting the layers from the input image, we divide the image into a fixed number of layers. From experiments, we found that almost the same quality estimation accuracy can be achieved by using two layers which significantly simplifies the method and also makes it computationally efficient. When two layers are used, the layering process divides the image into foreground and background images (as shown in Figure 3, layer 1 represented the background, and layer 2 contains the foreground). The quality scores of the two layers are computed and combined in a weighted manner to obtain the quality of the image I_w .

$$Q = c Q_b + (1 - c) Q_f, \quad (3)$$

where Q_b is the quality score of background (layer 1), Q_f is the quality score of foreground (layer 2), and c is the parameter that controls the relative importance of foreground and background. Its value is empirically estimated and set to $c = 0.4$; a detailed discussion on this is presented in Section 4.3. This means the quality score Q is largely dominated by the foreground layer which is in line with the previous

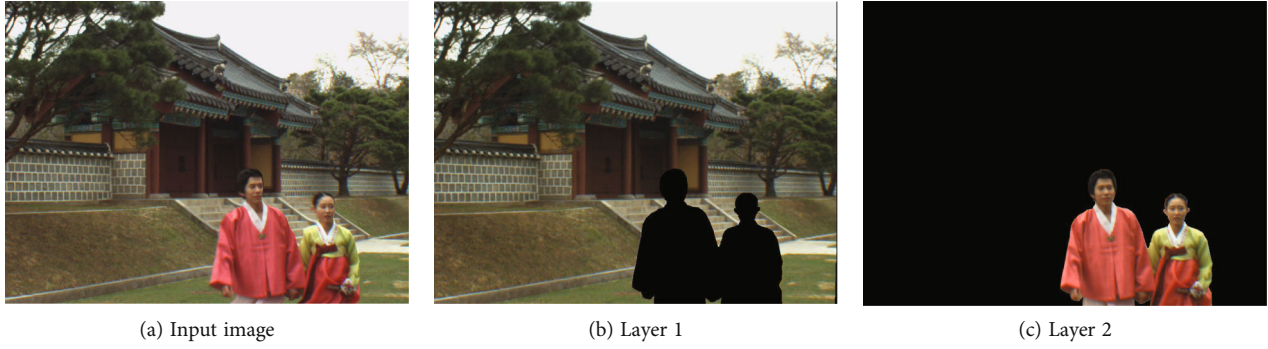


FIGURE 3: Layers of LoveBird1 sequence based on the histogram of depth map shown in Figure 2.

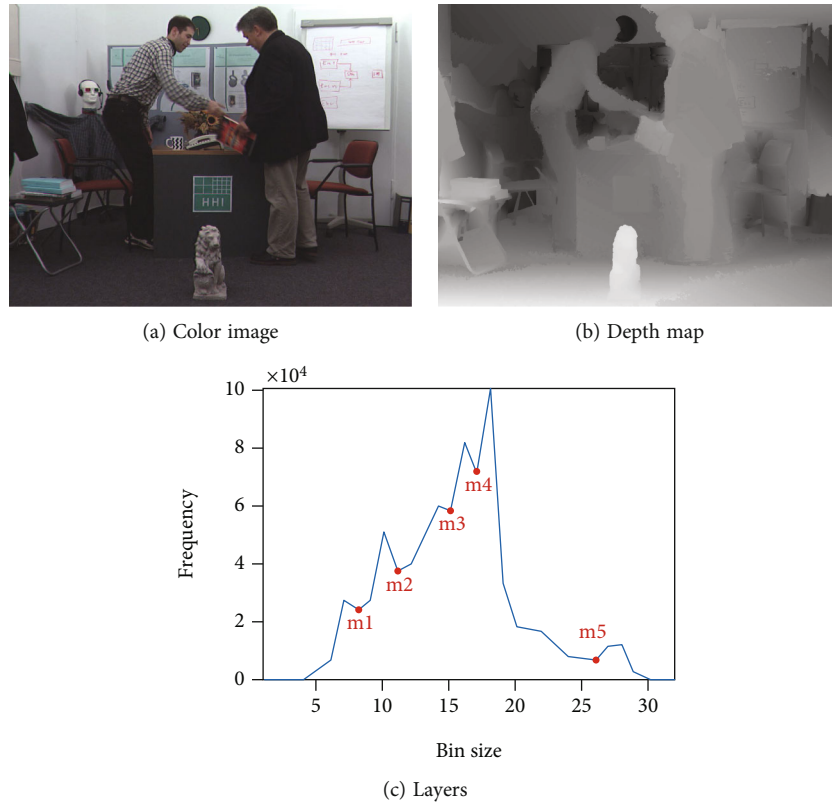


FIGURE 4: A sample synthesized image of Book_Arrival MVD sequence (a), its depth maps (b), and the histogram showing the five layers (c).

studies stating that the closer objects in a scene are visually more important than the objects at a far distance.

4. Experiments and Results

In this section, we perform different experiments to evaluate the performance of the proposed 3D-Layered Quality Metric (3D-LQM). Various statistical tools are used in this evaluation; we compare the performance of our method with the existing 3D quality assessment algorithms.

4.1. Evaluation Datasets. In the experimental evaluation, we have used two benchmark DIBR datasets, IRCCyN/IVC DIBR image database [39] and IETR DIBR image database [40]. Each dataset is a collection of DIBR-synthesized images

generated with different DIBR algorithms and MVD sequences. Subjective evaluations are available to test the performance of objective quality assessment metrics. Each dataset is briefly introduced in the following sections.

4.1.1. IRCCyN/IVC DIBR Image Database. This database contains DIBR images generated from three multiview video plus depth sequences including Book_Arrival, LoveBird1, and Newspaper. Four new viewpoints are generated from these three sequences using seven DIBR algorithms referred to as A1 to A7 which are introduced in the following.

- (i) A1: in [8], the holes on the borders are not filled so the border is cropped, and image is interpolated to its original size

- (ii) A2: the holes on the border are inpainted using image inpainting technique presented in [41]
- (iii) A3: Tanimoto et al. [42] is a view generating system with 3D warping; it uses inpainting to fill missing parts in virtual image
- (iv) A4: in the method proposed by Müller et al. [43], the depth information is used to fill holes in virtual image
- (v) A5: Ndjiki-Nya et al. [44] proposed a patch-based texture synthesis approach to fill the holes
- (vi) A6: Köppel et al. [45] extends the A5 approach by background sprite and uses temporal information in a video sequence to fill the holes
- (vii) A7: these are the DIBR-synthesized images with unfilled holes
- (iii) VSRS1 (View Synthesis Reference Software): it is single-view version of VSRS2 [42]
- (iv) Criminisi's method: the input view is warped to the target viewpoint and the holes are estimated using Criminisi's inpainting method [47]
- (v) LDI (Layered Depth Image) [48]: it is an object-based warping method that utilizes the inpainting method proposed in [49] to fill the holes
- (vi) HHF (Hierarchical Hole-Filling) method [50]: the disocclusions in the DIBR-synthesized view are estimated using a pyramid-based hierarchical approach
- (vii) Luo's method [51]: this method proposed a background reconstruction algorithm to estimate the holes in the DIBR images
- (viii) Ahn's method [52]: the holes in Ahn's DIBR generated image are filled with the patch-based texture synthesis

In Figure 5, a sample original image of LoveBird1 sequence is shown in Figure 5(a), and the synthesized images using the A1 to A7 DIBR approaches are presented in Figures 5(b)–5(h). Thus, in total, 84 view sequences were generated and rated by 48 subjects using the absolute categorical rating (ACR) and the mean opinion scores (MOS) were calculated. The reference images of the synthesized views were also rated by the subjects and were used to calculate the differential mean opinion score (DMOS).

4.1.2. IETR DIBR Image Database. IETR DIBR image database contains 140 images generated from 10 MVD sequences using 8 DIBR algorithms. The ten MVD sequences used in this database include Balloons, Book_Arrival, Kendo, LoveBird1, Newspaper, Poznan Street, Poznan Hall, Undo Dancer, Shark, and GT Fly. For each MVD sequence, two input views with their corresponding depth maps are used to generate a novel intermediate image using selected eight DIBR approaches. Two of these methods generate a single synthesized image by warping the input views of an MVD sequence to the virtual viewpoint and fusing the resultant images to obtain the target view. The rest of the six DIBR algorithms generate two synthesized images for each MVD sequence by warping the input views to the virtual viewpoint and recovering the holes using different strategies. Thus, for each MVD sequence, 14 DIBR images are obtained. These DIBR algorithms are introduced in the following.

- (i) Zhu's method [46]: the method does not use inpainting techniques to estimate the holes in the synthesized view; instead, it uses the occluded information to recover the holes
- (ii) VSRS2 (View Synthesis Reference Software) [40]: it is the reference DIBR algorithm adopted by the MPEG 3D video group. The method handles the depth-related artifacts by performing different filters and then uses it to obtain the virtual view. The holes are inpainted using Telea method [41]

The subjective evaluation was carried out with the help of 42 naive observers. Their ratings were used to obtain the differential mean opinion score (DMOS) which is scaled to (0, 1). Sample images from the IETR DIBR image database are presented in Figure 6. The IETR DIBR image database only provides the DIBR-synthesized color images; the corresponding depth maps are not available. Since the proposed quality metric requires depth information for layer segmentation, the original depth map of the virtual viewpoint for each sequence is used to make the database compatible with the proposed algorithm.

4.2. Objective Evaluation Parameters. We have used different statistical measures to evaluate the performance of the proposed metric. These include Pearson's linear correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SROCC), Kendall's rank order correlation coefficient (KROCC), and root mean square error (RMSE). PLCC is used to test the prediction accuracy of the metric, computed as

$$PLCC = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (4)$$

where x and y are the subjective rating and the objective metric score of image, respectively, and \bar{x} and \bar{y} are the mean of x and y , respectively.

SROCC measures the accuracy of an image quality assessment IQA metric using monotonic function. It is calculated as

$$SROCC = 1 - \frac{6 \sum_i (x_i - y_i)^2}{n(n^2 - 1)}, \quad (5)$$

where n is the number of observations.



FIGURE 5: Sample DIBR image from IRCCyN/IVC DIBR image database. Images of LoveBird1 sequence, original image and the DIBR images produced by the seven algorithms (A1 to A7).

KROCC is a nonparametric measure and represents the relationship between two variables.

$$\text{KROCC} = \frac{n_c - n_d}{(1/2)n(n-1)}, \quad (6)$$

where n_c and n_d are the number of concordant and the number of discordant in the dataset, respectively. RMSE is used

to measure the difference between predicted values and observed values.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - y_{\wedge 1})^2}{n}}. \quad (7)$$

According to the video quality expert group (VQEG) recommendation [53], the objective scores are mapped to

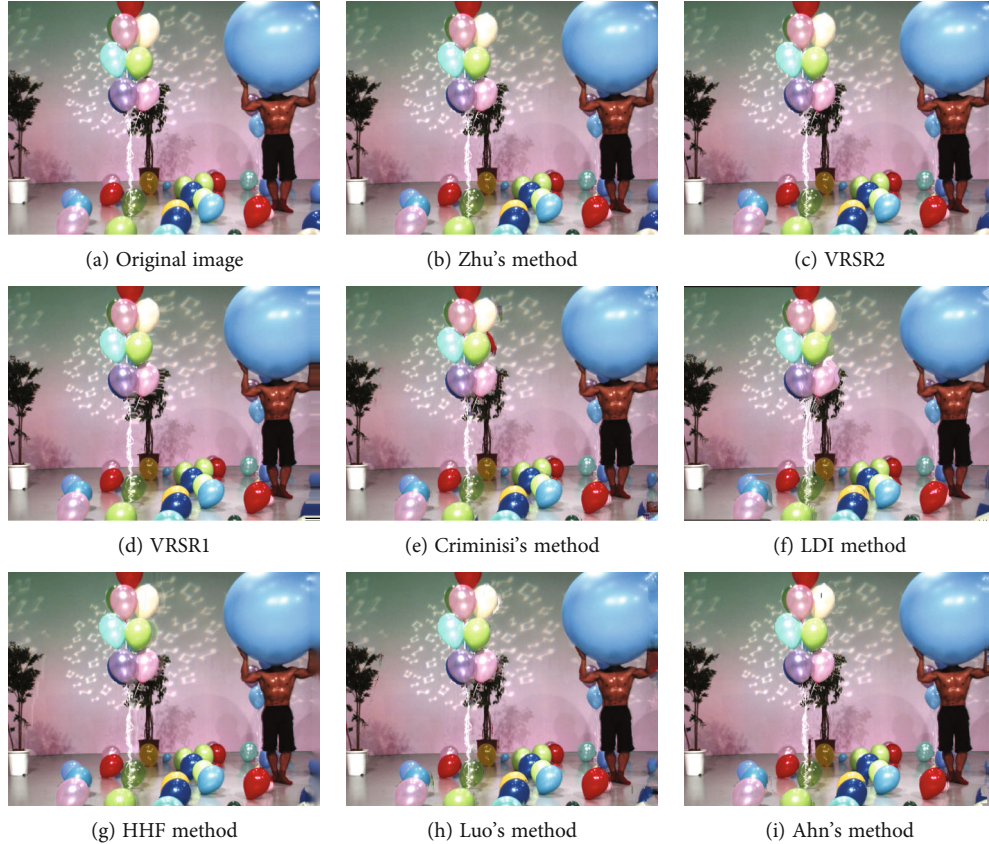


FIGURE 6: Sample images from IETR DIBR image database. Images of Balloons sequence, original reference image and the DIBR images produced by the eight DIBR algorithms.

the subjective differential mean opinion score (DMOS) using a nonlinear logistic mapping function. For this purpose, we have used the logistic function outlined in [9].

$$\hat{y}(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp \beta_2(x - \beta_3)} \right) + \beta_4(x) + \beta_5, \quad (8)$$

where $\hat{y}(x)$ is the predicted value, x is the score obtained by the objective quality metric, and β_1, \dots, β_5 are the parameters of the logistic function.

4.3. Parameter Settings. We recall that the final quality score of a synthesized image is calculated by combining the quality scores of the foreground and the background layers of the image (Equation (3)). The parameter c in this pooling controls the relative importance of the two scores, Q_b and Q_f . In the first experiment, we test the proposed algorithm with different values of c to find the optimal settings. Specifically, the value of c is varied between 0 and 1, and the performance parameters PLCC, SROCC, and KROCC on the results of the proposed method are computed. The results are presented in the Kiviat diagram in Figure 7. Each label at the axis denotes two values separated with $-$, the first value (c) represents the weight given to the Q_b score, and the second value ($1 - c$) represents the weight given to the Q_f score. The diagram shows that the proposed quality metric

achieves the best correlation with the subjective ratings when Q_b contributes 40% and Q_f 60%, that is, $c = 0.4$. At this setting, the proposed method achieves the highest 0.6859 PLCC, 0.6277 SROCC, and 0.6277 KROCC. Kindly note that this analysis is performed only on the IRCCyN/IVC DIBR image database and the same settings are applied in performance evaluation for the IETR DIBR image database.

4.4. Performance Comparison with 2D-IQA Metrics. In the next set of experiments, we compare the performance of the proposed quality metric with the existing 2D-IQA algorithms. For this evaluation, we have selected the widely used and well-known 2D-IQA algorithms including peak signal-to-noise ratio (PSNR). It is the ratio between the maximum power of a signal and the power of distorting noise that affects the quality of its representation. Structural similarity index (SSIM) [18] is used to measure the similarity between two images, original and synthesized image using luminance, contrast, and structural comparison. Multiscale structural similarity index (MSSIM) [17] is the modified form of SSIM. It is the mean of all the three components of SSIM. Visual signal-to-noise ratio (VSNR) [54] is based on the contrast threshold of distortion detection and visual distortion detection. Weighted signal-to-noise ratio (WSNR) [38] considers the weighting function of the human visual system with signal-to-noise ratio. Visual information fidelity (VIF) [55] is a full-reference image quality metric that uses both

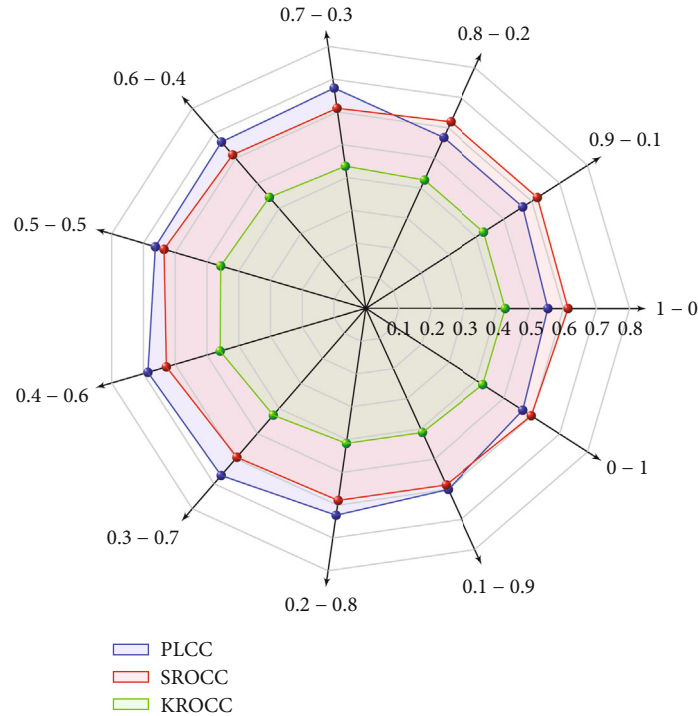


FIGURE 7: Performance of the proposed method with different contributions of background and foreground layers.

the distortion model and HVS model to evaluate an image. VIFP [55] is the pixel-based version of visual information fidelity. Information Fidelity Criterion (IFC) [56] uses the natural scene statistics to assess an image. Universal Quality Index (UQI) [57] uses the loss of correlation, luminance distortion, and contrast distortion in order to model image distortion.

The compared methods are executed on each test dataset, and all performance parameters are computed similar to the proposed algorithm. In the evaluation, the implementation of the compared methods provided by the authors or third-party libraries is used. The evaluation results on the IRCCyN/IVC DIBR image database are presented in Table 1 and on the IETR DIBR image database in Table 2. The results reveal that the proposed 3D-LQM algorithm outperforms all the compared methods on both databases with significant margins. The proposed metric achieves the highest correlation coefficients and minimum RMSE.

4.5. Performance Comparison with 3D-IQA Metrics. We also compare the performance of the proposed algorithm with existing different 3D-IQA algorithms. The compared methods include Bosc [58], VSQA [16], MW_PSNR [12], RMW_PSNR [12], MP_PSNR [59], RMP_PSNR [60], 3DSwIM [22], ST_SIAQ [61], NIQSV [29], NIQSV+ [30], and SIQE [25]. We computed the scores of these methods on IRCCyN/IVC DIBR image database and applied the same nonlinear regression function on these data scores before computing the performance parameters. The results are presented in Table 3.

The results show that the proposed metric outperforms all compared methods in PLCC and achieves more than a

TABLE 1: Performance comparison of the proposed and the compared 2D-IQA methods on IRCCyN/IVC DIBR image database. The best results are marked in bold.

Method	PLCC	SROCC	KROCC	RMSE
PSNR	0.4283	0.4616	0.3422	0.6017
SSIM	0.5715	0.4805	0.3283	0.5464
MSSIM	0.5489	0.5324	0.3801	0.5566
VSNR	0.3851	0.3982	0.2806	0.6147
WSNR	0.4174	0.4133	0.2962	0.6051
VIF	0.3085	0.1173	0.0730	0.6333
VIFP	0.2932	0.2337	0.1585	0.6366
IFC	0.3164	0.2539	0.1757	0.6316
UQI	0.3036	0.2961	0.2029	0.6344
3D-LQM	0.6859	0.6277	0.4584	0.4845

0.68 score. In terms of SROCC, RMP_PSNR performs marginally better than our method; however, in the other two measures, KROCC and RMSE, the proposed algorithm performs better than all compared methods.

The performance of the proposed method on the IETR DIBR image database is also computed and compared with the existing 3D-IQA metrics. The performance of the compared 3D-IQA algorithms is evaluated using the same regression function used for the proposed method. The results of the evaluation are presented in Table 4. The results show that the proposed method performs the best amongst all compared methods. It achieves the highest PLCC and SROCC of more than 0.60 and KROCC of more than 0.43 and minimum RMSE around 0.19.

TABLE 2: Performance comparison of the proposed and the compared 2D-IQA methods on IETR DIBR image database. The best results are marked in bold.

Method	PLCC	SROCC	KROCC	RMSE
PSNR	0.6067	0.5440	0.3801	0.1971
SSIM	0.3590	0.2547	0.1710	0.2314
MSSIM	0.4329	0.4096	0.2773	0.2251
VSNR	0.5241	0.4141	0.2742	0.2111
WSNR	0.6290	0.5696	0.4093	0.1986
VIF	0.2863	0.2640	0.1776	0.2375
VIFP	0.3229	0.2190	0.1496	0.2348
IFC	0.2829	0.2153	0.1363	0.2378
UQI	0.1983	0.1493	0.0956	0.2430
3D-LQM	0.6437	0.6000	0.4234	0.1897

TABLE 3: Performance comparison of the proposed and the compared 3D-IQA methods on IRCCyN/IVC DIBR image database. The best results are marked in bold.

Method	PLCC	SROCC	KROCC	RMSE
Bosc	0.5843	0.4905	0.3414	0.5403
VSQA	0.5742	0.5233	0.3673	0.5451
MW_PSNR	0.5622	0.5757	0.4378	0.5506
RMW_PSNR	0.5744	0.6245	0.4960	0.5450
MP_PSNR	0.6174	0.6227	0.4833	0.5238
RMP_PSNR	0.6772	0.6634	0.5382	0.4899
3DSwIM	0.6584	0.6154	0.4496	0.5011
ST_SIAQ	0.2277	0.1911	0.1203	0.6483
SIQE	0.5824	0.4492	0.3269	0.5653
NIQSV	0.6438	0.4248	0.2968	0.5095
NIQSV+	0.6519	0.5201	0.3830	0.5049
3D-LQM	0.6859	0.6277	0.4584	0.4845

TABLE 4: Performance comparison of the proposed and the compared 3D-IQA methods on IETR DIBR image database. The best results are marked in bold.

Method	PLCC	SROCC	KROCC	RMSE
Bosc	0.4164	0.3402	0.2282	0.2254
VSQA	0.5390	0.4740	0.3880	0.2476
MW_PSNR	0.5249	0.4875	0.3394	0.2110
RMW_PSN	0.5317	0.4953	0.3449	0.2100
MP_PSNR	0.5683	0.5488	0.3852	0.2040
RMP_PSNR	0.5981	0.5870	0.4134	0.1987
3DSwIM	0.1230	0.0668	0.0485	0.2460
ST_SIAQ	0.3000	0.2776	0.1934	0.2365
SIQE	0.2282	0.2333	0.1605	0.2414
NIQSV	0.1799	0.1545	0.1083	0.2439
NIQSV+	0.1805	0.2304	0.1568	0.2438
3D-LQM	0.6437	0.6000	0.4234	0.1897

From the experimental evaluation results presented in Tables 1–4, it is evident that the proposed method is accurate and achieves high correlations with the subjective ratings on both testing databases. We observe that the superior performance of the proposed algorithm is due to the segmentation of the image into the foreground and background layers and giving different importance to each layer. This helps the proposed method to find the salient regions in the image which contributes more than the other regions towards the total quality of the synthesized image. Moreover, unlike most existing saliency detection algorithms, we proposed a simple strategy that exploits the depth information of the scene to separate the visually important regions, foreground, from the visually less important regions, background. The proposed image layering method is accurate and computationally efficient.

4.6. Performance Analysis of 2D-IQA Metrics Coupled with the Proposed Framework. We recall that in the proposed quality assessment algorithm, after segmenting the synthesized image into layers with the help of the depth map, the quality of each layer is computed using available 2D quality metrics. In the next set of experiments, we used various 2D-IQA metrics with the proposed strategy to evaluate their performance using the IRCCyN/IVC DIBR image database. We executed them for the whole test dataset and computed the four performance parameters, PLCC, SROCC, KROCC, and RMSE. All the 2D-IQA metrics used in performance comparison in Section 4.4 are evaluated here with the proposed strategy. The performance of these metrics with the proposed strategy and without the proposed strategy (their standard implementation) is compared to capture the change. The results of these experiments are reported in graphs shown in Figure 8.

Figure 8(a) compares the prediction accuracy, Pearson’s linear correlation coefficient, of the quality metrics with working with the proposed strategy and without the proposed strategy.

The graph shows that the PLCC of each quality metric significantly improves when it is coupled with the proposed scheme where the image is segmented into layers and quality of each layer is assessed dependently and their scores are combined in a weighted manner to obtain a single quality score. For example, the PLCC of PSNR increased from 0.42 to 0.59 when used with the proposed strategy, an increase of more than 0.17 (38%). The graph shows that the PLCC of all quality assessment metrics witness a significant increase of more than 0.23, 40% on average when implemented with the proposed scheme.

Figure 8(b) compares the performance of the quality assessment metrics with and without the proposed strategy in terms of SROCC. The graph shows a significant increase in SROCC when implemented with the proposed scheme. For example, the SROCC of PSNR, IFC, and VIF increases more than 0.11, 0.32, and 0.47, respectively. Similar improvements in KROCC can be seen from Figure 8(c). The final comparison is performed on RMSE, presented in Figure 8(d). Similar to the other three performance parameters, RMSE also shows a significant improvement in the

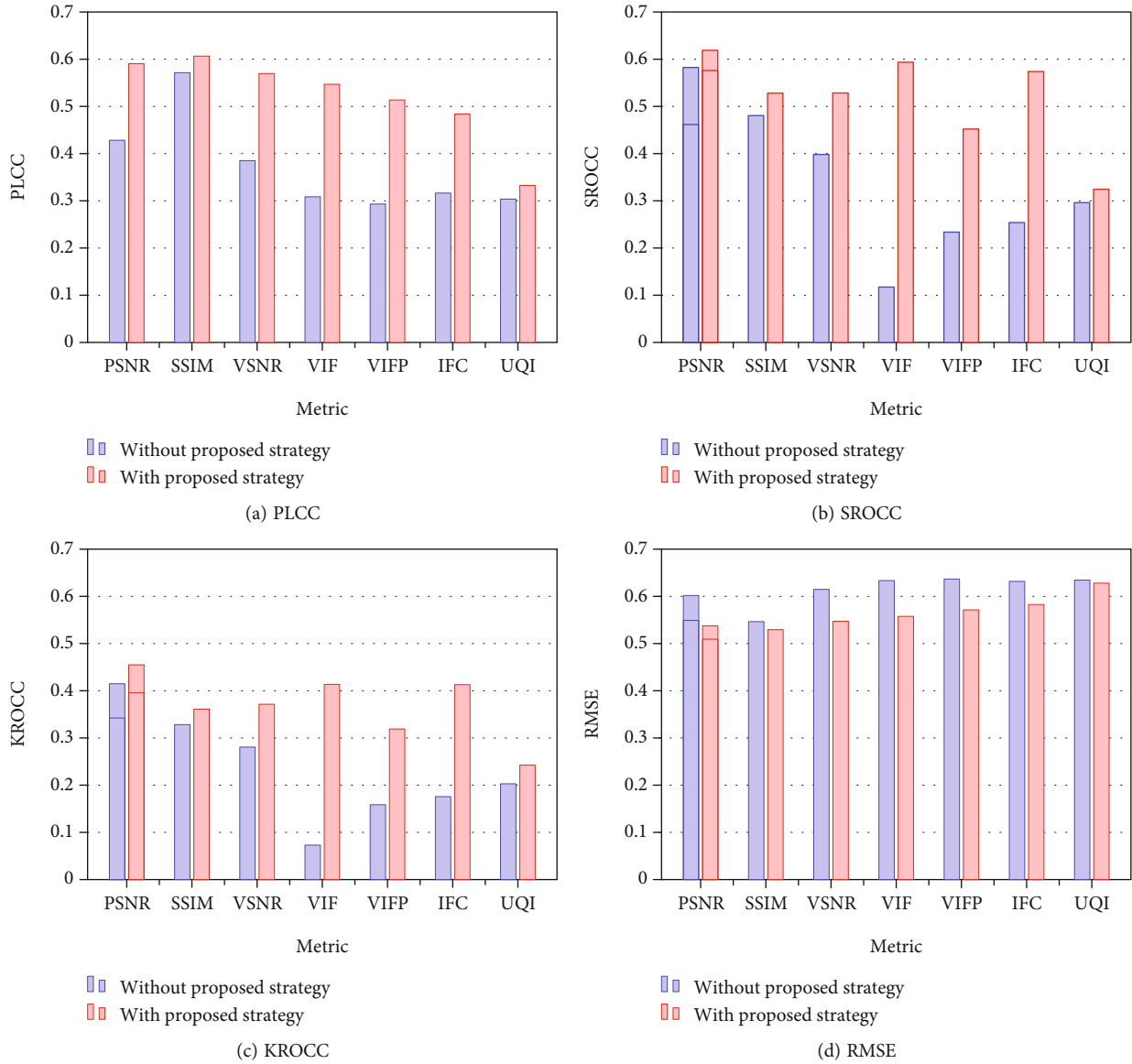


FIGURE 8: Performance evaluation of the quality assessment metrics working with and without proposed strategy in terms of (a) PLCC, (b) SORCC, (c) KROCC, and (d) RMSE.

performance of all compared methods when implemented with the proposed strategy. The statistics reveal an average improvement of more than 29% in RMSE. From the statistics presented graphs shown in Figure 8, we can conclude that the performance of the quality assessment techniques significantly improves when implemented with the proposed strategy. These conclusions are based on the experiments performed on the IRCCyN/IVC DIBR image database.

5. Conclusions and Future Research Directions

In this paper, we proposed a novel image quality assessment algorithm for 3D synthesized images. It is a layer-based algorithm where each layer contains objects at a certain distance from the viewing eye. In particular, the DIBR-synthesized images are divided into two layers, foreground layer and background layer. The former layer contains the objects

close to the observing eye and attracts most of the user’s attention. The background layer, on the other hand, contains the regions in the image that are unimportant and inconspicuous and thus are less likely to have viewer attention. The quality of each layer is computed individually, and the results are combined in a weighted manner. Since the foreground layer is salient, it is weighted more than the background layer. The performance of the proposed method is evaluated on the benchmark IRCCyN/IVC DIBR image database and IETR DIBR image database, and the results are compared with the existing 2D-IQA and 3D-IQA algorithms. The results reveal the effectiveness of the proposed quality assessment metric. A software release of the proposed metric is made publicly available on the project website: <http://faculty.pucit.edu.pk/~farid/Research/LQM.html>.

At the end of this research, we uncovered two potential research directions for future work that were out of the

scope of this research. In the proposed metric, we use the depth information of the image to segment it into the so-called layers; the color or texture information of the image is not used in this process. In addition to exploiting the depth map for layering, using the color information to improve the segmentation would be an interesting research direction. The proposed framework is implemented with 2D quality assessment metrics and it showed appreciatingly good results. However, investigating its performance when coupled with the 3D quality metrics would be another interesting study.

Data Availability

Previously reported IRCCyN/IVC DIBR image database and IETR DIBR image database were used to support this study and are available at https://qualinet.github.io/databases/image/irccynivc_dibr_images_database/ and https://vaader-data.insa-rennes.fr/data/stian/ieeetom/IETR_DIBR_Database.zip, respectively. These prior studies (and datasets) are cited at relevant places within the text as references [39, 40].

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was partially supported by the Higher Education Commission, Pakistan, under project “3DViCoQa” grant number NRPU-7389.

References

- [1] D. M. Chandler, “Seven challenges in image quality assessment: past, present, and future research,” *ISRN Signal Processing*, vol. 2013, Article ID 905685, 53 pages, 2013.
- [2] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, “Video quality assessment: from 2d to 3D — challenges and future trends,” in *2010 IEEE International Conference on Image Processing*, pp. 4025–4028, Hong Kong, China, 2010.
- [3] M. Tanimoto, “FTV: free-viewpoint television,” *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 555–570, 2012.
- [4] Y. Lin and J. Wu, “Quality assessment of stereoscopic 3D image compression by binocular integration behaviors,” *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1527–1542, 2014.
- [5] G. Zhai and X. Min, “Perceptual image quality assessment: a survey,” *Science China Information Sciences*, vol. 63, no. 11, p. 11, 2020.
- [6] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Multi-view video plus depth representation and coding,” in *2007 IEEE International Conference on Image Processing*, vol. 1, pp. 1–201–I–204, San Antonio, TX, USA, 2007.
- [7] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Panorama view with spatiotemporal occlusion compensation for 3D video coding,” *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 205–219, 2015.
- [8] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV,” in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 93–104, International Society for Optics and Photonics, 2004.
- [9] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Depth image based rendering with inverse mapping,” in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 135–140, Pula, Italy, 2013.
- [10] C. Zhu, L. Y. Yin Zhao, and M. Tanimoto, *3DTV System with Depth-Image-Based Rendering*, Springer, New York, 2013.
- [11] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Edge enhancement of depth based rendered images,” in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5452–5456, Paris, France, 2014.
- [12] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, “DIBR synthesized image quality assessment based on morphological wavelets,” in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, Pilos, Greece, 2015.
- [13] R. Blake and N. K. Logothetis, “Visual competition,” *Nature Reviews Neuroscience*, vol. 3, no. 1, pp. 13–21, 2002.
- [14] L. Li, X. Chen, Z. Yu, J. Wu, and G. Shi, “Depth image quality assessment for view synthesis based on weighted edge similarity,” in *CVPR Workshops*, pp. 17–25, Long Beach, California, USA, 2019.
- [15] S. Tian, L. Zhang, W. Zou et al., “Quality assessment of DIBR-synthesized views: an overview,” *Neurocomputing*, vol. 423, pp. 158–178, 2021.
- [16] P.-H. Conze, P. Robert, and L. Morin, “Objective view synthesis quality assessment,” in *Stereoscopic Displays and Applications XXIII*, vol. 8288, Burlingame, California, USA, 2012.
- [17] Z. Xiao, “A multi-scale structure similarity metric for image fusion quality assessment,” in *2011 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 69–72, Guilin, China, 2011.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [19] P. Merkle, Y. Morvan, A. Smolic et al., “The effects of multi-view depth video compression on multiview rendering,” *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 73–88, 2009.
- [20] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, “An edge-based structural distortion indicator for the quality assessment of 3D synthesized views,” in *2012 Picture Coding Symposium*, pp. 249–252, Krakow, Poland, 2012.
- [21] G. Yue, C. Hou, G. Ke, T. Zhou, and G.-t. Zhai, “Combining local and global measures for DIBR-synthesized image quality evaluation,” *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2019.
- [22] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, “Objective image quality assessment of 3D synthesized views,” *Signal Process.-Image Commun.*, vol. 30, pp. 78–88, 2015.
- [23] H. M. U. H. Alvi, M. S. Farid, M. H. Khan, and M. Grzegorzek, “Quality assessment of 3D synthesized images based on textural and structural distortion estimation,” *Applied Sciences*, vol. 11, no. 6, p. 2666, 2021.
- [24] Y. J. Jung, H. G. Kim, and Y. M. Ro, “Critical binocular asymmetry measure for the perceptual quality assessment of

- synthesized stereo 3D images in view synthesis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1201–1214, 2016.
- [25] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Objective quality metric for 3D virtual views,” in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3720–3724, Quebec City, QC, Canada, 2015.
- [26] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Perceptual quality assessment of 3D synthesized images,” in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 505–510, Hong Kong, 2017.
- [27] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Evaluating virtual image quality using the side-views information fusion and depth maps,” *Information Fusion*, vol. 43, pp. 47–56, 2018.
- [28] K. Gu, V. Jakhethiya, J.-F. Qiao, X. Li, W. Lin, and D. Thalmann, “Model-based referenceless quality metric of 3D synthesized images using local image description,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 394–405, 2018.
- [29] S. Tian, L. Zhang, L. Morin, and O. Deforges, “NIQSV: a no reference image quality assessment metric for 3D synthesized views,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1248–1252, New Orleans, LA, USA, 2017.
- [30] S. Tian, Z. Lu, L. Morin, and O. Déforges, “NIQSV+: a no-reference synthesized view quality assessment metric,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1652–1664, 2018.
- [31] C. Tsai and H. Hang, “Quality assessment of 3D synthesized views with depth map distortion,” in *2013 Visual Communications and Image Processing (VCIP)*, pp. 1–6, Kuching, Malaysia, 2013.
- [32] Y. Dong, M. T. Pourazad, and P. Nasiopoulos, “Human visual system-based saliency detection for high dynamic range content,” *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 549–562, 2016.
- [33] J. Wu, G. Han, P. Liu, H. Yang, H. Luo, and Q. Li, “Saliency detection with bilateral absorbing Markov chain guided by depth information,” *Sensors*, vol. 21, no. 3, p. 838, 2021.
- [34] D. Cheng, Y. Xu, F. Nie, and D. Tao, “Saliency detection via a multiple self-weighted graph-based manifold ranking,” *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 885–896, 2020.
- [35] D. Kim, S. Ryu, and K. Sohn, “Depth perception and motion cue based 3D video quality assessment,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1–4, Seoul, Korea (South), 2012.
- [36] J. Wagemans, J. Elder, M. Kubovy et al., “A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization,” *Psychological Bulletin*, vol. 138, no. 6, pp. 1172–1217, 2012.
- [37] Y. Fang, W. Lin, B.-S. Lee, C.-T. Lau, Z. Chen, and C.-W. Lin, “Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum,” *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 187–198, 2012.
- [38] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.
- [39] E. Bosc, R. Pepion, P. Le Callet et al., “Towards a new quality metric for 3-d synthesized view assessment,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.
- [40] S. Tian, Z. Lu, L. Morin, and O. Déforges, “A benchmark of DIBR synthesized view quality assessment metrics on a new database for immersive media applications,” *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1235–1247, 2019.
- [41] A. Telea, “An image inpainting technique based on the fast marching method,” *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [42] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “View generation with 3D warping using depth information for FTV,” *Signal Processing: Image Communication*, vol. 24, no. 1–2, pp. 65–72, 2008.
- [43] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “View synthesis for advanced 3D video systems,” *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 438148, 11 pages, 2008.
- [44] P. Ndjiki-Nya, M. Koppel, D. Doshkov et al., “Depth image-based rendering with advanced texture synthesis for 3-d video,” *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 453–465, 2011.
- [45] M. Köppel, P. Ndjiki-Nya, D. Doshkov et al., “Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering,” in *2010 IEEE International Conference on Image Processing*, pp. 1809–1812, Hong Kong, China, 2010.
- [46] C. Zhu and S. Li, “Depth image based view synthesis: new insights and perspectives on hole generation and filling,” *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 82–93, 2016.
- [47] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [48] V. Jantet, C. Guillemot, and L. Morin, “Object-based layered depth images for improved virtual view synthesis in rate-constrained context,” in *2011 18th IEEE International Conference on Image Processing*, pp. 125–128, Brussels, Belgium, 2011.
- [49] O. Le Meur, J. Gautier, and C. Guillemot, “Exemplar-based inpainting based on local geometry,” in *2011 18th IEEE International Conference on Image Processing*, pp. 3401–3404, Brussels, Belgium, 2011.
- [50] M. Solh and G. AlRegib, “Hierarchical hole-filling for depth-based view synthesis in ftv and 3D video,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 495–504, 2012.
- [51] G. Luo, Y. Zhu, Z. Li, and L. Zhang, “A hole filling approach based on background reconstruction for view synthesis in 3D video,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1781–1789, Las Vegas, Nevada, USA, June 2016.
- [52] I. Ahn and C. Kim, “A novel depth-based virtual view synthesis method for free viewpoint video,” *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 614–626, 2013.
- [53] Video Quality Expert Group (VQEG), *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase ii (fr-tv2)*, Technical report, ITU, 2003.
- [54] D. M. Chandler and S. S. Hemami, “VSNR: a wavelet-based visual signal-to-noise ratio for natural images,” *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [55] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

- [56] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [57] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [58] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Transactions on Broadcasting*, vol. 62, no. 2, pp. 446–456, 2016.
- [59] D. Sandic-Stankovic, D. Kukulj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, Lisbon, Portugal, 2015.
- [60] D. Sandic-Stankovic, D. Kukulj, and L. Callet, "Multi-scale synthesized view assessment based on morphological pyramids," *Journal of Electronic Imaging*, vol. 67, no. 1, pp. 3–11, 2016.
- [61] S. Ling and P. Le Callet, "Image quality assessment for free viewpoint video based on mid-level contours feature," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 79–84, Hong Kong, China, 2017.