frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 November 2021
doi: 10.3389/fnins.2021.761610

DIBR-Synthesized Image Quality Assessment With Texture and Depth Information

Guangcheng Wang[1], *Quan Shi*[1]*, *Yeqin Shao*[1] *and Lijuan Tang*[2]

[1] School of Transportation and Civil Engineering, Nantong University, Nantong, China, [2] School of Electronics and Information, Jiangsu Vocational College of Business, Nantong, China

Accurately predicting the quality of depth-image-based-rendering (DIBR) synthesized images is of great significance in promoting DIBR techniques. Recently, many DIBR-synthesized image quality assessment (IQA) algorithms have been proposed to quantify the distortion that existed in texture images. However, these methods ignore the damage of DIBR algorithms on the depth structure of DIBR-synthesized images and thus fail to accurately evaluate the visual quality of DIBR-synthesized images. To this end, this paper presents a DIBR-synthesized image quality assessment metric with Texture and Depth Information, dubbed as TDI. TDI predicts the quality of DIBR-synthesized images by jointly measuring the synthesized image's colorfulness, texture structure, and depth structure. The design principle of our TDI includes two points: (1) DIBR technologies bring color deviation to DIBR-synthesized images, and so measuring colorfulness can effectively predict the quality of DIBR-synthesized images. (2) In the hole-filling process, DIBR technologies introduce the local geometric distortion, which destroys the texture structure of DIBR-synthesized images and affects the relationship between the foreground and background of DIBR-synthesized images. Thus, we can accurately evaluate DIBR-synthesized image quality through a joint representation of texture and depth structures. Experiments show that our TDI outperforms the competing state-of-the-art algorithms in predicting the visual quality of DIBR-synthesized images.

Keywords: depth-image-based-rendering, image quality assessment, colorfulness, texture structure, depth structure

OPEN ACCESS

Edited by:
Ke Gu,
Beijing University of Technology, China

Reviewed by:
Guanghui Yue,
Shenzhen University, China
Suiyi Ling,
Xlaomi Inc., France
Wei Sun,
Shanghai Jiao Tong University, China

***Correspondence:**
Quan Shi
sq@ntu.edu.cn

Specialty section:
This article was submitted to Perception SciencePerception Science, a section of the journal Frontiers in Neuroscience

Received: *20 August 2021*
Accepted: *11 October 2021*
Published: *03 November 2021*

Citation:
Wang G, Shi Q, Shao Y and Tang L (2021) DIBR-Synthesized Image Quality Assessment With Texture and Depth Information. Front. Neurosci. 15:761610. doi: 10.3389/fnins.2021.761610

1. INTRODUCTION

With the advent of the 5G era and the advancement of 3-dimensional display technology, video technology moves from "seeing clearly" to the ultra-high definition and immersive virtual reality era of "seeing the reality." Free-viewpoint videos (FVVs) have broad applications in entertainment, education, medical treatment, military applications for its ability to provide users with visual information of integrity, immersion, and interactivity (Selzer et al., 2019; Yildirim, 2019). Thus, FVV is also regarded as the vital research direction of next-generation video technologies (Tanimoto et al., 2011). Due to hardware conditions, cost, and bandwidth constraints, it is feasible to collect a certain number of viewpoint images in realistic environments. Still, it is often impractical to collect a full range of 360-degree viewpoint images. Therefore, it is necessary to synthesize virtual viewpoint images from existing reference viewpoint images by relying on virtual viewpoint synthesis techniques (Wang et al., 2020, 2021; Li et al., 2021a; Ling et al., 2021).

Because depth-image-based-rendering (DIBR) technologies only require a texture image and its corresponding depth map to generate the image at any viewpoint, it becomes the most popular virtual viewpoint synthesis technique (Luo et al., 2020). Unfortunately, because the performance of existing DIBR algorithms is not perfect, some distortions are often introduced during the warping and rendering processes, as shown in **Figure 1**. The quality of DIBR-synthesized images directly influences the visual experience in FVV-related applications, determining whether these applications can be successfully put into use. Hence, studying the quality evaluation methods for virtual viewpoint synthesis has important practical significance.

Image quality assessment (IQA) has been a crucial frontier research direction in image processing in recent decades. Massive IQA algorithms for natural images have been proposed, divided into full-reference, reduced-reference, and no-reference according to the required full, partial, and no information of the reference image. For instance, Wang et al. (2004) proposed a full-reference IQA metric based on comparing the structural information between the reference and distorted images, namely Structural SIMilarity (SSIM). Zhai et al. (2012) quantify psychovisual quality of images based on free-energy interpretation of cognition in brain theory. Min et al. (2018) proposed a pseudo-reference image (PRI) based IQA framework, which is different from the traditional full-reference IQA framework. The standard full-reference IQA framework assumes that the reference image is a high visual quality image. In contrast, the framework proposed by Min et al. assumes that the reference image suffers the most severe distortion in related applications. Based on the PRI-based IQA framework, Min et al. measures the similarity between the distorted image's and the PRI's structures to estimate blockiness, sharpness, and noisiness.

In recent years, researchers have realized that IQA algorithms for natural images have difficulty in estimating the geometric distortion prevalent in DIBR-synthesized images. For this problem, Bosc et al. (2011) calculated the difference map between the synthesized image and the reference image based on SSIM and adopted a threshold strategy to detect the disoccluded area in the synthesized image. Then, the quality score of a synthesized image is obtained by measuring the average structural similarity of the disoccluded region. Conze et al. (2012) used SSIM to generate a similarity map between the reference image and the synthesized image and further extracted the texture, gradient direction, and image contrast weighting maps based on the obtained similarity map to predict the synthesized image quality score. Stankovic et al. designed the Morphological Wavelet Peak signal-to-noise ratio (MW-PSNR) for assessing the synthesized image quality (Dragana et al., 2015b). Meanwhile, the authors proposed a simplified version of MW-PSNR called MW-PSNR-reduce (Dragana et al., 2015b), which only uses the PSNR value of the higher-level scale image to predict the synthesized image quality. For better performance, Stankovic et al. adopted morphological pyramid decomposition to replace the morphological wavelet decomposition in the above-mentioned MW-PSNR (Dragana et al., 2015b) and MW-PSNR-reduce (Dragana et al., 2015b), which successively produce MP-PSNR (Dragana et al., 2015a) and MP-PSNR-reduce (Dragana et al., 2016). Although these methods for the synthesized images have better performance than the IQA algorithms devised for natural images, their performance still misses the actual requirements.

Over the past few years, researchers have been aware of a close relationship between quantifying the local geometric distortion and the quality assessment of DIBR-synthesized images and the screen content images (Gu et al., 2017b). Gu et al. (2018a), Li et al. (2018b), Jakhetiya et al. (2019), and Yue et al. (2019) have arranged the idea in the design of DIBR-synthesized IQA methods, respectively. In literature (Gu et al., 2018a), Gu et al. adopted an autoregression (AR)-based local description operator to estimate the local geometric distortion. Specifically, the authors measure the local geometric distortion by calculating the reconstruction error between the synthesized image and its AR-based prediction. In literature (Jakhetiya et al., 2019), assumed that the geometric distortion behavior is similar to the outliers and further proved this hypothesis using ROR statistics based on the three-Sigma rule. Based on this view, the authors highlight the local geometric distortion through a median filter and further fuse these prominent distortions to assess the synthesized image quality.

Moreover, based on the local geometric distortion measurement, Yue et al. (2019)'s and Li et al. (2018b)'s methods introduce global sharpness estimation to predict the synthesized image quality. Yue et al. (2019) considered three major DIBR-related distortions, including the disoccluded

FIGURE 1 | Examples of the local geometric distortion and the color deviation distortion in the synthesized images. **(A)** is the ground-truth image. **(B,C)** are the synthesized images, which includes the local geometric distortion and the color deviation distortion compared to the ground-truth image.

region, the stretching region, and global sharpness. The authors first detect disoccluded regions by analyzing the local similarity. Then, the stretching regions are determined by combining the local similarity analysis and a threshold solution. Finally, the authors measure inter-scale self-similarity to estimate global sharpness. Li et al. (2018b) designed a SIFT-flow warping based disoccluded region detection algorithm. Then, the geometric distortion is measured by combining with the size and distortion intensity of local disoccluded areas. Moreover, a reblurring-based solution is developed to capture blur distortion. We find two critical problems from the above-mentioned DIBR-synthesized IQA methods. First, these methods ignore the influence of color deviation distortion on the visual quality of DIBR-synthesized images. Second, These methods only focus on estimating the geometric distortion and blur distortion from textured images without considering the local geometric distortion's adverse effects on the synthesized image's depth structure.

Inspired by these findings, we present a newly synthesized image quality assessment metric that combines Texture and Depth Information, namely TDI. Specifically, we adopt the colorfulness module proposed by Hasler and Suesstrunk (2003) to extract the color features of a synthesized image and its reference image (i.e., the ground-truth image) and then calculate the feature error to estimate the color deviation distortion. We perform discrete wavelet transform on the texture information of the synthesized image and its reference image and further calculate the similarity of the high-frequency subbands of a pair of synthesized and reference images. The similarity result is used to estimate the local geometric distortion and global sharpness. Meanwhile, we use SSIM to compute the structural similarity between the depth maps of a pair of synthesized and reference images to represent the effects of the local geometric distortion and blur distortion on the depth of field of the synthesized image. In addition, TDI develops a linear weighting scheme to fuse the obtained features. We verify the performance of our TDI metric on the public IRCCyN/IVC DIBR-synthesized image database Bosc et al. (2011), and the experimental results prove that our TDI metric performs better than the competing state-of-the-art (SOTA) IQA algorithms. Compared with the existing works, the highlights of the proposed algorithm mainly include two aspects: (1) we integrate the color deviation distortion caused by DIBR algorithms into the development of DIBR-synthesized view quality perception model; (2) This paper estimates the quality degradation brought by the local geometric distortion and blur distortion from the texture and depth information of the synthesized view.

The remaining chapters of this paper are organized as follows. Section 2 introduces the proposed TDI in detail. Section 3 compares our TDI with SOTA IQA metrics for natural and DIBR-synthesized images. Section IV summarizes the whole research.

2. PROPOSED METHOD

The design philosophy of our TDI is based on quantifying the local geometric distortion, global sharpness, and color deviation distortion. After extracting the corresponding features, a linear weighting strategy fuses the above features to infer the final quality score. **Figure 2** shows the framework of the proposed TDI.

2.1. Color Deviation Distortion Estimation

The human visual system (HVS) is susceptible to color, so the measurement of color deviation distortion has a direct impact on the visual experience (Gu et al., 2017a; Liao et al., 2019). As shown in **Figure 1**, compared to the high-quality reference image, the synthesized image has the color deviation distortion. However, since it is not the main distortion in the synthesized image, most existing DIBR-synthesized IQA algorithms ignore the impact of the color deviation distortion on the visual experience. To more accurately evaluate the synthesized image quality, this paper takes the measurement of color deviation distortion into account in the proposed TDI metric. In the literature (Hasler and Suesstrunk, 2003), Hasler and Suesstrunk devised a highly HVS-related image colorfulness estimation based on psychophysical category scale experiments. The image colorfulness estimation model is specifically defined as follows:

$$C = (\sigma_{rg}^2 + \sigma_{yb}^2)^{\frac{1}{2}} + 0.3 \cdot (\mu_{rg}^2 + \mu_{yb}^2)^{\frac{1}{2}}, \tag{1}$$

where σ_{rg}, σ_{yb}, μ_{rg} and μ_{yb} are the variance and mean of the rg and yb channels, respectively. The calculation method of rg and yb channels is shown in formula 2.

$$rg = R - G, yb = \frac{1}{2}(R + G) - B \tag{2}$$

Then, we calculate the absolute value of the colorfulness difference between a synthesized image and its associated reference image (i.e., formula 5) as the quantized result of the color deviation distortion that existed in the synthesized image.

$$Q_1 = |C_{syn} - C_{ref}|, \tag{3}$$

where C_{syn} and C_{ref} represent the colorfulness of the synthesized image and its reference image, respectively.

2.2. Local Geometric Distortion and Global Sharpness Measurement

The proposed TDI extracts structural features from the texture image and its corresponding depth image and designs a linear pooling strategy for information fusion to achieve a more accurate measurement of the local geometric distortion and global sharpness. This part explains in detail how TDI extracts structure features from texture and depth images.

2.2.1. Structure Feature Extracting From Texture Domain

We first use the Cohen-Daubechies-Fauraue 9/7 filter (Cohen et al., 1992) to perform discrete wavelet transform on the synthesized and reference images. **Figure 3** shows some examples of high-frequency wavelet subbands (i.e., HL, LH, and HH subbands) of two synthesized images and their reference image. From **Figure 3**, we observe that the geometric distortion regions (such as the red box area) of the synthesized and reference

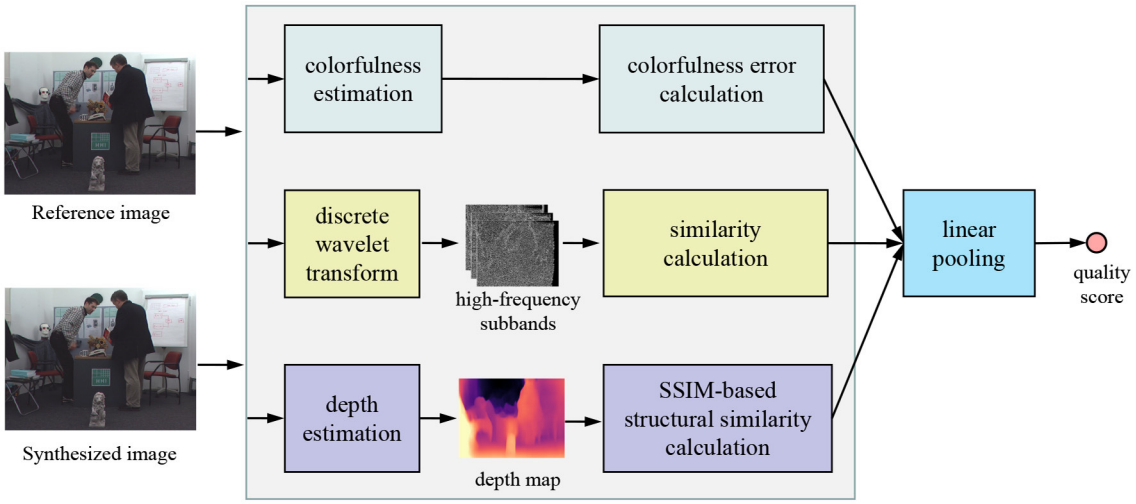

FIGURE 2 | Framework of the proposed TDI metric for predicting the quality of DIBR-synthesized images.

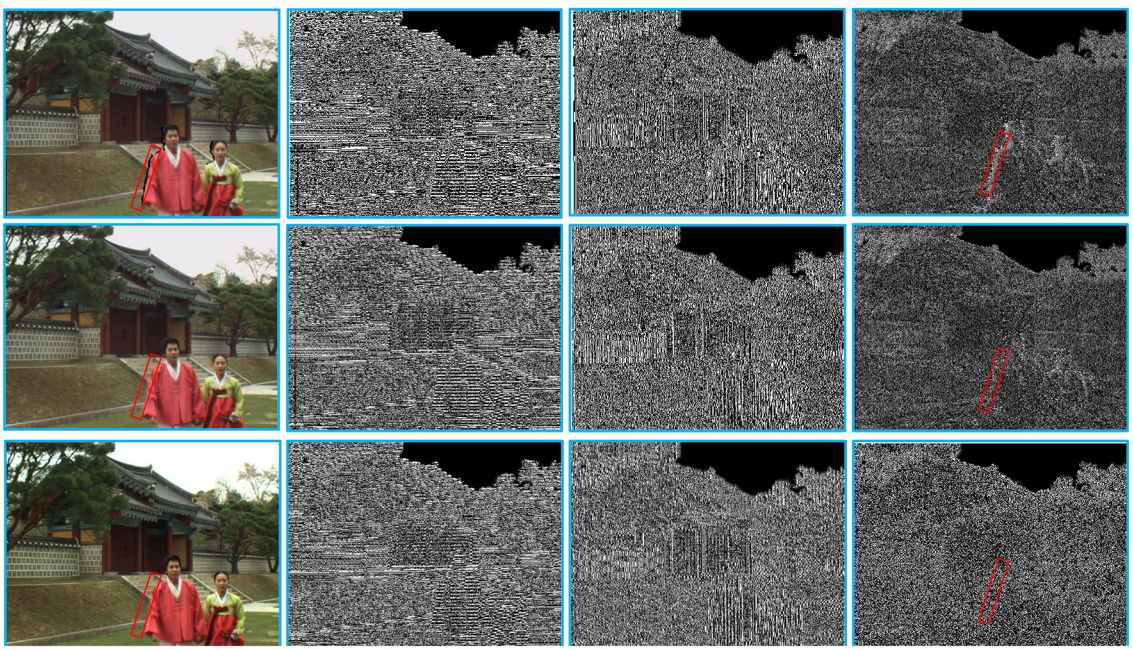

FIGURE 3 | Examples of the high-frequency wavelet subbands (i.e., HL, LH, and HH subbands) of two synthesized images and their reference image. From left to right, the images in each row are a synthesized/reference image and its corresponding HL, LH, and HH wavelet subbands. Note that the synthesized image of the first row has only the warping process.

images in the HH subbands differ significantly. Motivated by this, we measure the local geometric distortion by computing the similarity between the HH subbands of a pair of synthesized and reference images, which is defined as follows:

$$Q_2 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{2\cdot HH_{syn}(i)\cdot HH_{ref}(i)+\epsilon}{HH_{syn}(i)+HH_{ref}(i)+\epsilon}\right], \tag{4}$$

where HH_{syn} and HH_{ref} represent the HH subbands of a synthesized image and its corresponding reference image. i and N are the pixel index and the number of pixels of a given image, respectively. A small constant ϵ avoids the risk of zero denominator. Moreover, since blur distortion usually causes loss of high-frequency information in images, the energy of high-frequency wavelet subbands has been widely used for no-reference image sharpness estimation (Vu and Chandler, 2012; Wang et al., 2020). Therefore, the developed similarity between

FIGURE 4 | Examples of the depth maps of two synthesized images and their reference image. From top to bottom, the images in each column are a synthesized/reference image and its corresponding depth map. Note that the synthesized image of the first column has only the warping process.

the HH subbands of the synthesized image and its reference image can also effectively estimate the global sharpness of the DIBR-synthesized image.

2.2.2. Structure Feature Extracting From Depth Domian

Considering that local geometric distortion and global sharpness damage the structural information of the synthesized view in the texture domain and affect the depth structure of the synthesized view. Thus, we measure the structural similarity between the depth maps of a pair of synthesized and reference views in the depth domain to estimate the depth degradation introduced by the local geometric distortion and blur distortion. The depth map prediction algorithm computes the depth map at the virtual viewpoint. At present, massive deep learning-based depth image estimation algorithms have been proposed (Atapour-Abarghouei and Breckon, 2018; Li et al., 2018a; Zhang et al., 2018; Godard et al., 2019). In our TDI, we employ Clément Godard's depth prediction network for estimating the depth maps of the DIBR-synthesized image and its reference image. **Figure 4** shows some examples of the depth maps of two synthesized images and their ground-truth image estimated by Clément Godard's method. From the green box area in **Figure 4**, it can be easily observed that the local geometric distortion is very destructive to the depth structure of the synthesized image. So the geometric distortion contained in a synthesized image can be effectively estimated by measuring the structural similarity between the depth maps of a pair of synthesized and reference images. In particular, the structural similarity between the depth maps of a synthesized image and its reference image is computed as follows:

$$Q_3 = \frac{1}{N}\sum_{i=1}^{N}(SSIM(D_{syn}(i), D_{ref}(i))), \tag{5}$$

where D_{syn} and D_{ref} represent the depth maps of a synthesized image and its reference image predicted by Clément Godard's algorithm. SSIM is an image quality evaluation index based on the structural similarity between the reference and distorted images (Wang et al., 2004; Jang et al., 2019).

2.3. Linear Pooling Scheme

To evaluate the visual quality of DIBR-synthesized views more efficiently, this paper extracts three features from the texture and depth domains to estimate the color deviation distortion, the local geometric distortion, and global sharpness. Since the features Q_1, Q_2, and Q_3 are complementary, we propose a novel linear pooling scheme to fuse the texture and depth information to form the final TDI model. A smaller Q_1 value shows the difference between the colorfulness of the synthesized image and its reference image is smaller. That is, the quality of the synthesized image is higher. The Q_2 and Q_3 are the texture and depth structure similarity between a pair of synthesized and reference images, respectively. The values of Q_2 and Q_3 are higher, indicating that the quality of a pair of synthesized and reference views is more similar. That is, the quality of the synthesized image is better. With this fact, a linear pooling scheme is developed to fuse the obtained features, which is defined as follows:

$$S = -\frac{\alpha}{1+\alpha+\beta}\cdot Q_1 + \frac{1}{1+\alpha+\beta}\cdot Q_2 + \frac{\beta}{1+\alpha+\beta}\cdot Q_3, \tag{6}$$

where the parameters α and β are used to adjust the contribution of Q_1, Q_2, and Q_3. In section 3, we detail the selection of parameters α and β.

TABLE 1 | Performance comparison of 21 SOTA IQA measures on the IRCCyN/IVC database (Bosc et al., 2011).

Metric	Type	SRCC	PLCC	RMSE
PSNR	Natural Images	0.3095	0.3976	0.6109
SSIM (Wang et al., 2004)	Natural Images	0.4368	0.4850	0.5823
IW-SSIM (Wang and Li, 2011)	Natural Images	0.4053	0.5831	0.5409
ADD-SSIM (Gu et al., 2016)	Natural Images	0.4672	0.5512	0.5556
PSIM (Gu et al., 2017a)	Natural Images	0.4576	0.5315	0.5640
NIQE (Mittal et al., 2013)	Natural Images	0.3739	0.4374	0.5987
IL-NIQE (Zhang et al., 2015)	Natural Images	0.5348	0.4998	0.5767
ARISM (Gu et al., 2015)	Natural Images	0.3728	0.3994	0.6104
BIQME (Gu et al., 2018b)	Natural Images	**0.6770**	**0.7271**	**0.4571**
MW-PSNR (Dragana et al., 2015b)	DIBR-synthesized Images	0.5757	0.5622	0.5506
MP-PSNR (Dragana et al., 2015a)	DIBR-synthesized Images	0.6227	0.6174	0.5238
MP-PSNR-reduce (Dragana et al., 2016)	DIBR-synthesized Images	0.6634	0.6772	0.4899
NIQSV+ (Tian et al., 2018)	DIBR-synthesized Images	0.6668	0.7114	0.4679
APT (Gu et al., 2018a)	DIBR-synthesized Images	0.7157	0.7307	0.4546
CLGM (Yue et al., 2019)	DIBR-synthesized Images	0.6528	0.6750	0.4620
STD (Wang et al., 2021)	DIBR-synthesized Images	0.7729	0.7901	0.4082
LMS (Zhou et al., 2019)	DIBR-synthesized Images	0.8050	0.7690	0.3940
IDEA (Li et al., 2021b)	DIBR-synthesized Images	—	0.7796	—
GANs-NRM (Ling et al., 2020)	DIBR-synthesized Images	**0.8070**	**0.8260**	**0.3860**
OUT (Jakhetiya et al., 2019)	DIBR-synthesized Images	0.7036	0.7678	0.4266
TDI (Pro.)	DIBR-synthesized Images	**0.7905**	**0.7992**	**0.4002**

The best performance in each type is highlighted in bold.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this part, we construct experiments on the IRCCyN/IVC database to test the performance of the proposed TDI method and other SOTA IQA algorithms.

3.1. Experimental Setup

3.1.1. Competing IQA Metrics

In this paper, we collect twenty SOTA IQA algorithms for natural images and DIBR-synthesized images as competing algorithms. The competing IQA metrics designed for natural images include PSNR, SSIM (Wang et al., 2004), IW-SSIM (Wang and Li, 2011), ADD-SSIM (Gu et al., 2016), PSIM (Gu et al., 2017a), NIQE (Mittal et al., 2013), ILNIQE (Zhang et al., 2015), ARISM (Gu et al., 2015), and BIQME (Gu et al., 2018b). The competing IQA methods devised for DIBR-synthesized images consist of MW-PSNR (Dragana et al., 2015b), MP-PSNR (Dragana et al., 2015a), MP-PSNR-reduce (Dragana et al., 2016), NIQSV+ (Tian et al., 2018), APT (Gu et al., 2018a), CLGM (Yue et al., 2019), STD (Wang et al., 2021), LMS (Zhou et al., 2019), IDEA (Li et al., 2021b), GANs-NRM (Ling et al., 2020), and OUT (Jakhetiya et al., 2019).

3.1.2. Testing Dataset

In this paper, we test the performance of the proposed TDI metric and twenty SOTA IQA algorithms on the public IRCCyN/IVC database (Bosc et al., 2011). The IRCCyN/IVC DIBR-synthesized image database contains 12 reference images and its corresponding 84 synthesized images generated via seven DIBR algorithms. In the subjective experiment, the authors adopt the absolute category rating-hidden reference method to mark DIBR-synthesized images. The images in the IRCCyN/IVC image dataset are from three free-view sequences (i.e., "Book Arrival," "Lovebird," and "Newspaper") with a resolution of 1,024 × 768.

TABLE 2 | Ablation experiments about the proposed components.

Metric	SRCC	PLCC	RMSE
Q1	0.4412	0.4971	0.5777
Q2	0.6126	0.6133	0.5259
Q3	0.4470	0.5346	0.5627
TDI (overall model)	0.7905	0.7992	0.4002

3.1.3. Performance Benchmarking

In this paper, three commonly used indicators, including Spearman Rank-order Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE), are used to evaluate the performance of the proposed TDI metric and other competing IQA algorithms devised for natural images and DIBR-synthesized images. The SRCC index evaluates the monotonic consistency between subjective scores and objective scores predicted by IQA metrics. The PLCC and RMSE indicators evaluate the accuracy of the scores predicted by IQA algorithms. The larger values of SRCC and PLCC, and the smaller value of RMSE, indicate the

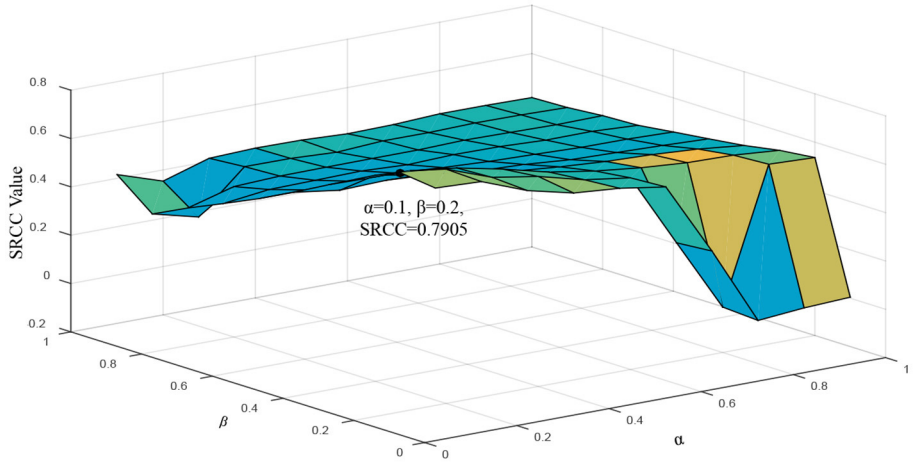

FIGURE 5 | The impact of the parameters α and β on the robustness of the proposed TDI metric.

performance of the corresponding IQA metric is better. The PLCC is defined as follows:

$$PLCC = \frac{\sum_i (a_i - \bar{a})(l_i - \bar{l})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (l_i - \bar{l})^2}}, \tag{7}$$

where a_i and $\bar{a}$ are the estimated quality score of the i-th synthesized image and the average value of all a_i, respectively. l_i and $\bar{l}$ are the subjective quality label of the i-th synthesized image and the average value of all l_i, respectively. The SRCC is computed as follows:

$$SRCC = 1 - \frac{6 \sum_{q=1}^{Q} d_q^2}{Q(Q^2 - 1)}, \tag{8}$$

where Q is the number of pairs of predicted quality scores and subjective quality labels. d_q represents the ranking difference between the predicted quality scores and the subjective quality labels in each group. Before calculating the above indicators, we need to map the quality scores of all IQA methods to the same range through a non-linear logistic function (Min et al., 2020a,b), which is defined as follows:

$$f(x) = \tau_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2 (x - \tau_3)}}\right) + \tau_4 x + \tau_5, \tag{9}$$

where τ_1, τ_2, τ_3, τ_4, and τ_5 are the fitting parameters. x and $f(x)$ are the quality scores predicted by IQA algorithms and their corresponding non-linear mapping results, respectively.

3.2. Performance Comparisons With SOTA IQA Metrics

As shown in **Table 1**, our TDI metric achieves SRCC value of 0.7905, PLCC value of 0.7992, and RMSE value of 0.4002 on the IRCCyN/IVC dataset, which outperforms most competing IQA metrics designed for natural images and DIBR-synthesized images. In terms of SRCC, the performance of our proposed method is very close to that of the best-performing GANs-NRM. From **Table 1**, we observe two important conclusions:

1. The performance of the IQA algorithms for natural images on IRCCyN/IVC is far inferior to the IQA methods designed for DIBR-synthesized images. The SRCC, PLCC, and RMSE values of the best BIQME (Gu et al., 2018b) on the IRCCyN/IVC dataset (Bosc et al., 2011) are 0.6770, 0.7271, and 0.4571, respectively, and its SRCC value still does not reach 0.7. Regarding SRCC, PLCC and RMSE, the proposed TDI metrics are 16.77, 9.92, and 12.45% higher than the top BIQME methods, respectively.
2. The APT (Gu et al., 2018a) and OUT (Jakhetiya et al., 2019) metrics, existing best performing IQA algorithms on the IRCCyN/IVC (Bosc et al., 2011) database based on geometric distortion quantization, achieve SRCC value of 0.7157, PLCC value of 0.7678, and RMSE value of 0.4266, respectively. Our proposed TDI metric increases the values of SRCC, PLCC, and RMSE by 10.45, 4.09, and 6.19% on this result. Experiments show that the proposed TDI metric, combining colorfulness, texture structure, and depth structure, can efficiently predict DIBR-synthesized image quality.

3.3. Ablation Study

In this part, we conduct some ablation experiments to verify the contributions of the proposed key components (i.e., Q_1, Q_2, and Q_3). **Table 2** shows the test results of the components Q_1, Q_2, Q_3, and the overall module on the public IRCCyN/IVC data set. From the results, we observe the performance of the overall TDI model is far superior to each component, which shows that the proposed sub-modules can complementally evaluate the quality of the synthesized view. That is, the fusion of texture and depth information is of great significance to the view synthesis quality perception. Moreover, we further analyze the influence of the parameters α and β in equation (6) on the robustness of the proposed TDI metric, and the experimental results are shown in **Figure 5**. Obviously,

when the parameters α and β are smaller, the performance of the proposed TDI metric is better, that is, compared to the components Q_1 and Q_3, the component Q_2 is more important, which is also in line with the test results in **Table 2**. According to the robustness analysis, the parameters α and β are set to 0.1 and 0.2, respectively, to optimize the proposed TDI module.

3.4. Applications in Other Fields

With the rapid development of computer vision, the three-dimensional-related technologies can be implemented in numerous practical applications. The first application is abnormality detection in industry, especially the smoke detection in industrial scenarios which has received an amount of attention from researchers in recent years (Gu et al., 2020b, 2021b; Liu et al., 2021). The process of abnormality detection relies on images, therefore combining three-dimensional technology with this can make the image acquisition equipment obtain a more accurate, intuitive and realistic image information, so as to enable the staff to monitor the abnormal situation in time and then avoid bad things from happening. The second application is atmospheric pollution monitoring and early warning (Gu et al., 2020a, 2021a; Sun et al., 2021). The three-dimensional visualized images contain more detailed information, thus enabling efficient and accurate air pollution monitoring. The third application field is three-dimensional vision and display technologies (Gao et al., 2020; Ye et al., 2020). Compared with the ordinary two-dimensional screen display, three-dimensional technology can make the image is no longer confined to the plane of the screen (Sugita et al., 2019), as if it can come out of the screen, so that the audience has a feeling of immersion. The fourth application is road traffic monitoring (Ke et al., 2019). Three-dimensional technology can monitor the traffic flow information of major intersections in an all-round and intuitive way. All in all, there are several advantages of DIBR technology, so it is necessary to extend this technology to different fields.

4. CONCLUSION

This paper presents a novel DIBR-synthesized image quality assessment algorithm based on texture and depth information fusion, dubbed as TDI. First, in the texture domain, we evaluate the visual quality of the synthesized images by extracting the differences in colorfulness and HH wavelet subband between the synthesized image and its reference image. Then, in the depth domain, we estimate the impact of the local geometric distortion on the quality of the synthesized views by calculating the structural similarity between the depth maps of a pair of synthesized and reference views. Finally, a linear pooling model is developed to fuse the above features to predict DIBR-synthesized image quality. Experiments on the IRCCyN/IVC database show that the proposed TDI algorithm outperforms each sub-module and most competing SOTA image quality assessment methods designed for natural and DIBR-synthesized images.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

QS and YS designed and instruct the experiments. GW wrote the code for the experiments. GW, QS, and LT carried out the experiments and wrote the manuscript. YS and LT collected and analyzed the experiment data. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was funded by the National Natural Science Foundation of China, grant no. 61771265.

REFERENCES

Atapour-Abarghouei, A., and Breckon, T. P. (2018). "Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 2800–2810.

Bosc, E., Pepion, R., Le Callet, P., Koppel, M., Ndjiki-Nya, P., Pressigout, M., et al. (2011). Towards a new quality metric for 3-d synthesized view assessment. *IEEE J. Sel. Top. Signal. Process.* 5, 1332–1343. doi: 10.1109/JSTSP.2011.2166245

Cohen, A., Daubechies, I., and Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* 45, 485–560. doi: 10.1002/cpa.3160450502

Conze, P.-H., Robert, P., and Morin, L. (2012). Objective view synthesis quality assessment. *Int. Soc. Opt. Eng.* 8288:8256–8288. doi: 10.1117/12.908762

Dragana, S.-S., Dragan, K., and Patrick, L. C. (2015a). "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)* (Lisbon), 1–4.

Dragana, S.-S., Dragan, K., and Patrick, L. C. (2015b). "Dibr synthesized image quality assessment based on morphological wavelets," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)* (Costa Navarino), 1–6.

Dragana, S.-S., Dragan, K., and Patrick, L. C. (2016). Multi-scale synthesized view assessment based on morphological pyramids. *J. Electr. Eng.* 67, 3–11. doi: 10.1515/jee-2016-0001

Gao, Z., Zhai, G., Deng, H., and Yang, X. (2020). Extended geometric models for stereoscopic 3d with vertical screen disparity. *Displays* 65:101972. doi: 10.1016/j.displa.2020.101972

Godard, C., Aodha, O. M., Firman, M., and Brostow, G. (2019). "Digging into self-supervised monocular depth estimation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: IEEE), 3827–3837.

Gu, K., Jakhetiya, V., Qiao, J.-F., Li, X., Lin, W., and Thalmann, D. (2018a). Model-based referenceless quality metric of 3d synthesized images using local image description. *IEEE Trans. Image Process.* 27, 394–405. doi: 10.1109/TIP.2017.2733164

Gu, K., Li, L., Lu, H., Min, X., and Lin, W. (2017a). A fast reliable image quality predictor by fusing micro- and macro-structures. *IEEE Trans. Ind. Electr.* 64, 3903–3912. doi: 10.1109/TIE.2017.2652339

Gu, K., Liu, H., Xia, Z., Qiao, J., Lin, W., and Daniel, T. (2021a). Pm2.5 monitoring: use information abundance measurement and wide and deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4278–4290. doi: 10.1109/TNNLS.2021.3105394

Gu, K., Tao, D., Qiao, J.-F., and Lin, W. (2018b). Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 1301–1313. doi: 10.1109/TNNLS.2017.2649101

Gu, K., Wang, S., Zhai, G., Lin, W., Yang, X., and Zhang, W. (2016). Analysis of distortion distribution for pooling in image quality prediction. *IEEE Trans. Broadcast.* 62, 446–456. doi: 10.1109/TBC.2015.2511624

Gu, K., Xia, Z., and Junfei, Q. (2020a). Stacked selective ensemble for pm2.5 forecast. *IEEE Trans. Instrum. Meas.* 69, 660–671. doi: 10.1109/TIM.2019.2905904

Gu, K., Xia, Z., Qiao, J., and Lin, W. (2020b). Deep dual-channel neural network for image-based smoke detection. *IEEE Trans. Multimedia* 22, 311–323. doi: 10.1109/TMM.2019.2929009

Gu, K., Zhai, G., Lin, W., Yang, X., and Zhang, W. (2015). No-reference image sharpness assessment in autoregressive parameter space. *IEEE Trans. Image Process.* 24, 3218–3231. doi: 10.1109/TIP.2015.2439035

Gu, K., Zhang, Y., and Qiao, J. (2021b). Ensemble meta-learning for few-shot soot density recognition. *IEEE Trans. Ind. Inform.* 17, 2261–2270. doi: 10.1109/TII.2020.2991208

Gu, K., Zhou, J., Qiao, J.-F., Zhai, G., Lin, W., and Bovik, A. C. (2017b). No-reference quality assessment of screen content pictures. *IEEE Trans. Image Process.* 26, 4005–4018. doi: 10.1109/TIP.2017.2711279

Hasler, D., and Suesstrunk, S. E. (2003). "Measuring colorfulness in natural images," in *Human Vision and Electronic Imaging VIII, Vol. 5007*, eds B. E. Rogowitz and T. N. Pappas (Santa Clara, CA: SPIE), 87–95.

Jakhetiya, V., Gu, K., Singhal, T., Guntuku, S. C., Xia, Z., and Lin, W. (2019). A highly efficient blind image quality assessment metric of 3-d synthesized images using outlier detection. *IEEE Trans. Ind. Inform.* 15, 4120–4128. doi: 10.1109/TII.2018.2888861

Jang, C. Y., Kim, S., Cho, K.-R., and Kim, Y. H. (2019). Performance analysis of structural similarity-based backlight dimming algorithm modulated by controlling allowable local distortion of output image. *Displays* 59, 1–8. doi: 10.1016/j.displa.2019.05.001

Ke, R., Li, Z., Tang, J., Pan, Z., and Wang, Y. (2019). Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Trans. Intel. Trans. Syst.* 20, 54–64. doi: 10.1109/TITS.2018.2797697

Li, B., Dai, Y., and He, M. (2018a). Monocular depth estimation with hierarchical fusion of dilated cnns and soft-weighted-sum inference. *Pattern Recognit.* 83, 328–339. doi: 10.1016/j.patcog.2018.05.029

Li, L., Huang, Y., Wu, J., Gu, K., and Fang, Y. (2021a). Predicting the quality of view synthesis with color-depth image fusion. *IEEE Trans. Circ. Syst. Video Technol.* 31, 2509–2521. doi: 10.1109/TCSVT.2020.3024882

Li, L., Zhou, Y., Gu, K., Lin, W., and Wang, S. (2018b). Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness. *IEEE Trans. Multimedia* 20, 914–926. doi: 10.1109/TMM.2017.2760062

Li, L., Zhou, Y., Wu, J., Li, F., and Shi, G. (2021b). Quality index for view synthesis by measuring instance degradation and global appearance. *IEEE Trans. Multimedia* 23:320–332. doi: 10.1109/TMM.2020.2980185

Liao, C.-C., Su, C.-W., and Chen, M.-Y. (2019). Mitigation of image blurring for performance enhancement in transparent displays based on polymer-dispersed liquid crystal. *Displays* 56, 30–37. doi: 10.1016/j.displa.2018.11.001

Ling, S., Li, J., Che, Z., Min, X., Zhai, G., and Le Callet, P. (2021). Quality assessment of free-viewpoint videos by quantifying the elastic changes of multi-scale motion trajectories. *IEEE Trans. Image Process.* 30, 517–531. doi: 10.1109/TIP.2020.3037504

Ling, S., Li, J., Che, Z., Wang, J., Zhou, W., and Le Callet, P. (2020). Re-visiting discriminator for blind free-viewpoint image quality assessment. *IEEE Trans. Multimedia*. doi: 10.1109/TMM.2020.3038305. [Epub ahead of print].

Liu, H., Lei, F., Tong, C., Cui, C., and Wu, L. (2021). Visual smoke detection based on ensemble deep cnns. *Displays* 69:102020. doi: 10.1016/j.displa.2021.102020

Luo, G., Zhu, Y., Weng, Z., and Li, Z. (2020). A disocclusion inpainting framework for depth-based view synthesis. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 1289–1302. doi: 10.1109/TPAMI.2019.2899837

Min, X., Gu, K., Zhai, G., Liu, J., Yang, X., and Chen, C. W. (2018). Blind quality assessment based on pseudo-reference image. *IEEE Trans. Multimedia* 20, 2049–2062. doi: 10.1109/TMM.2017.2788206

Min, X., Zhai, G., Zhou, J., Farias, M. C. Q., and Bovik, A. C. (2020a). Study of subjective and objective quality assessment of audio-visual signals. *IEEE Trans. Image Process.* 29, 6054–6068. doi: 10.1109/TIP.2020.2988148

Min, X., Zhou, J., Zhai, G., Le Callet, P., Yang, X., and Guan, X. (2020b). A metric for light field reconstruction, compression, and display quality evaluation. *IEEE Trans. Image Proc.* 29, 3790–3804. doi: 10.1109/TIP.2020.2966081

Mittal, A., Soundararajan, R., and Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal. Process. Lett.* 20, 209–212. doi: 10.1109/LSP.2012.2227726

Selzer, M. N., Gazcon, N. F., and Larrea, M. L. (2019). Effects of virtual presence and learning outcome using low-end virtual reality systems. *Displays* 59, 9–15. doi: 10.1016/j.displa.2019.04.002

Sugita, N., Sasaki, K., Yoshizawa, M., Ichiji, K., Abe, M., Homma, N., et al. (2019). Effect of viewing a three-dimensional movie with vertical parallax. *Displays* 58, 20–26. doi: 10.1016/j.displa.2018.10.007

Sun, K., Tang, L., Qian, J., Wang, G., and Lou, C. (2021). A deep learning-based pm2.5 concentration estimator. *Displays* 69:102072. doi: 10.1016/j.displa.2021.102072

Tanimoto, M., Tehrani, M. P., Fujii, T., and Yendo, T. (2011). Free-viewpoint tv. *IEEE Signal. Process. Mag.* 28, 67–76. doi: 10.1109/MSP.2010.939077

Tian, S., Zhang, L., Morin, L., and Déforges, O. (2018). Niqsv+: a no-reference synthesized view quality assessment metric. *IEEE Trans. Image Process.* 27, 1652–1664. doi: 10.1109/TIP.2017.2781420

Vu, P. V., and Chandler, D. M. (2012). A fast wavelet-based algorithm for global and local image sharpness estimation. *IEEE Signal. Process. Lett.* 19, 423–426. doi: 10.1109/LSP.2012.2199980

Wang, G., Wang, Z., Gu, K., Jiang, K., and He, Z. (2021). Reference-free dibr-synthesized video quality metric in spatial and temporal domains. *IEEE Trans. Circ. Syst. Video Technol.* doi: 10.1109/TCSVT.2021.3074181

Wang, G., Wang, Z., Gu, K., Li, L., Xia, Z., and Wu, L. (2020). Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain. *IEEE Trans. Image Process.* 29, 1802–1814. doi: 10.1109/TIP.2019.2945675. [Epub ahead of print].

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wang, Z., and Li, Q. (2011). Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* 20, 1185–1198. doi: 10.1109/TIP.2010.2092435

Ye, P., Wu, X., Gao, D., Deng, S., Xu, N., and Chen, J. (2020). Dp3 signal as a neuro-indictor for attentional processing of stereoscopic contents in varied depths within the 'comfort zone'. *Displays* 63:101953. doi: 10.1016/j.displa.2020.101953

Yildirim, C. (2019). Cybersickness during vr gaming undermines game enjoyment: a mediation model. *Displays* 59, 35–43. doi: 10.1016/j.displa.2019.07.002

Yue, G., Hou, C., Gu, K., Zhou, T., and Zhai, G. (2019). Combining local and global measures for dibr-synthesized image quality evaluation. *IEEE Trans. Image Process*. 28, 2075–2088. doi: 10.1109/TIP.2018.2875913

Zhai, G., Wu, X., Yang, X., Lin, W., and Zhang, W. (2012). A psychovisual quality metric in free-energy principle. *IEEE Trans. Image Process*. 21, 41–52. doi: 10.1109/TIP.2011.2161092

Zhang, L., Zhang, L., and Bovik, A. C. (2015). A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process*. 24, 2579–2591. doi: 10.1109/TIP.2015.2426416

Zhang, Z., Xu, C., Yang, J., Tai, Y., and Chen, L. (2018). Deep hierarchical guidance and regularization learning for end-to-end depth estimation. *Pattern Recognit*. 83:430–442. doi: 10.1016/j.patcog.2018.05.016

Zhou, Y., Li, L., Ling, S., and Le Callet, P. (2019). Quality assessment for view synthesis using low-level and mid-level structural representation. *Signal. Process. Image Commun*. 74, 309–321. doi: 10.1016/j.image.2019.03.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Shi, Shao and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.