Depth Image Quality Assessment for View Synthesis Based on Weighted Edge Similarity

Leida Li[1,2], Xi Chen[2,*], Yu Zhou[2], Jinjian Wu[1], Guangming Shi[1]
[1]School of Artificial Intelligence, Xidian University
[2]School of Information and Control Engineering, China University of Mining and University
{lileida,chenxi,zhouy7476}@cumt.edu.cn, jinjian.wu@mail.xidian.edu.cn, gmshi@xidian.edu.cn

Abstract

With the increasing prevalence of multi-view and free-view displays, virtual view synthesis has been extensively researched. In view synthesis, texture and depth images are typically fed into a depth-image-based-rendering (DIBR) algorithm to generate the new viewpoints. In contrast to the enormous amount of research effort on the quality assessment of texture images and rendering process, much less effort has been dedicated to the quality evaluation of depth images. To fill this gap, this paper presents a quality metric of depth images for view synthesis. Depth image represents information relating to the distance of the surfaces of scene objects from a viewpoint, and edge conveys key location information in depth image, which is extremely important in view rendering. Therefore, the proposed metric is developed with emphasis on measuring the edge characteristics of depth images. Firstly, a similarity map is computed between the distorted and reference depth images by combining intensity and gradient information. Then an adaptive weighting map is calculated by integrating depth distance and location characteristics in the depth image. Finally, an edge indication map is computed and utilized to guide the pooling process, producing the overall depth quality score. Extensive experiments and comparisons on the public MCL-3D database demonstrate that the proposed metric outperforms the relevant state-of-the-art quality metrics.

1. Introduction

In recent years, multi-view and free-view videos have become more and more popular. View synthesis is a key technique in these applications. In view synthesis, texture and depth images are utilized to generate the new viewpoints, where depth-image-based-rendering (DIBR) [9] is the most widely used approach. The perceptual quality of synthesized view is mainly determined by three aspects, namely the quality of texture and depth images as well as the DIBR algorithm [2]. Distortions in texture image typically transfer to the synthesized view directly, and the existing quality metrics can handle this properly. Recently, great effort has also been put on the quality evaluation of the rendering operation [6, 20, 1, 21, 14, 33]. In contrast, depth images are used to guide the warping operation in DIBR, and distortions in depth images are quite different from those in texture images and the rendering process. How to accurately evaluate the quality of depth images in view synthesis is still an open problem. In this paper, we try to fill this gap by proposing an objective depth image quality assessment method for the application of view synthesis.

To the authors' best knowledge, only a few metrics have been proposed for predicting the depth image quality in view synthesis. In [8], a novel Blind Depth Quality Metric (BDQM) was proposed to evaluate the compression distortions in depth images. The gradient map of a depth image was first calculated to locate the pixels, which are sensitive to the compression distortion. Then the histogram of compression sensitive pixels and their neighborhood were constructed to evaluate the quality of depth images. In [28], a no-reference depth quality assessment was proposed based on two-step edge misalignment error matching between depth image and texture image. The edge matching was based on the following three criteria, namely spatial similarity, edge orientation similarity and segment length similarity. Then the matching result was utilized to compute the bad point rate (BPR), which was defined as the depth quality score. In [13], a reduced-reference depth quality metric was proposed using a pair of color and depth images. The depth distortions were first calculated based on the edge directions in the neighborhood. Then Gabor filtering was conducted on the texture image to produce a weighting map. Finally, the local depth distortions were pooled into an overall quality score with the guidance of the weighting map.

The aforementioned metrics have achieved notable advances in predicting the quality of depth images. Meantime,

*Corresponding Author: Xi Chen

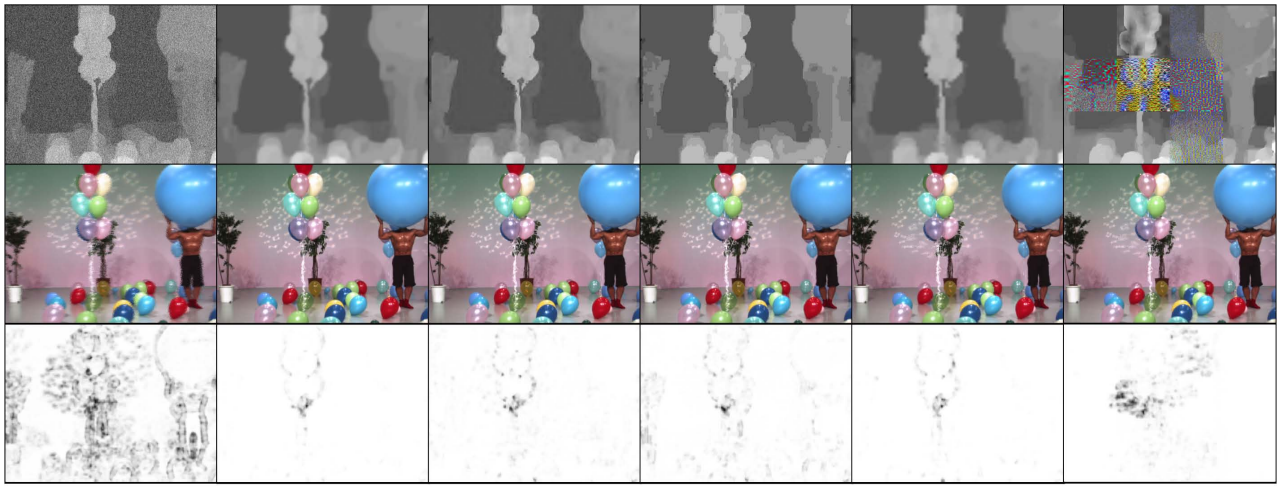

Figure 1. SSIM maps of the synthesized images using distorted depth maps. First row: distorted depth maps; Second row: synthesized images using depth images in the first row and undistorted texture images; Third row: SSIM maps between the synthesized images and reference images. (a) Additive white noise; (b) Gaussian blur; (c) JP2K; (d) JPEG; (e) Down-sampling blur; (f) Transmission error.

it is noted that the method in [8] is specifically designed for compression distortion. In [28, 13], texture image is further needed for predicting the depth image quality. In practice, depth images may be subject to different kinds of distortions, which is similar to natural scene images. Therefore, it is highly desirable to develop a depth image quality metric that is capable of evaluating diversified distortions. Ideally, the evaluation can be done using the depth image directly. With these considerations, this paper presents a depth image quality metric for view synthesis, which can be used to evaluate the general distortions in depth images without referencing to the corresponding texture images. The design philosophy of the proposed method is to measure the edge characteristics in the depth images, which convey rich information in the rendering process and thus directly influence the quality of the synthesized view. To be specific, a similarity map is first generated between the distorted and reference depth images, which takes into account both intensity and gradient characteristics. Then an adaptive weighting map is computed by combining the depth distance and location prior. Finally, an edge indication map is calculated and utilized to guide the pooling, producing the overall depth quality score. The performance of the proposed metric is evaluated on the MCL-3D view synthesis quality database. Extensive results demonstrate that the proposed method outperforms the state-of-the-arts.

2. Distortion Analysis of Depth Images

In view synthesis, original views are warped to the 3D space followed by an inverse warping to the target view. In this process, depth image is used to guide both the forward and backward warping. Since depth image represents the distance of objects in a scene from a viewpoint, it typically consists of many homogeneous regions. Edges in an depth image convey key information relating to the varying distances of objects, which have great influence on the quality of view synthesis. These characteristics are quite different from those of natural scene images. To have an intuitive understanding of the characteristics of depth distortions, in Figure 1 we show some distorted depth maps in MCL-3D database [23]. Since depth image is not directly used for human viewing, here we further show the corresponding synthesized images and the SSIM [24] maps, where darker region indicates more severe distortion. In this example, undistorted texture images are adopted in the rendering process, so the degradations in the synthesized images are purely introduced by the distorted depth maps.

From the SSIM maps in Figure 1, it is easily observed that the distortions in the synthesized images are mainly concentrated in the edge regions. This is most pronounced in Figure 1(a), where the additive white noise is present. For example, obvious noise exists inside the balloon and human body in the depth map, however in the synthesized image the observed distortions are mainly distributed around their boundaries. This further confirms the fact that the distortions in view synthesis mainly come from the edge regions in the depth maps. Following this observation, the proposed depth quality metric is designed with emphasis on measuring the distortions in edge regions.

3. Proposed Depth Quality Metric

The flowchart of the proposed depth quality metric is shown in Figure 2. It consists of three main stages, namely similarity map generation, weighting map generation and

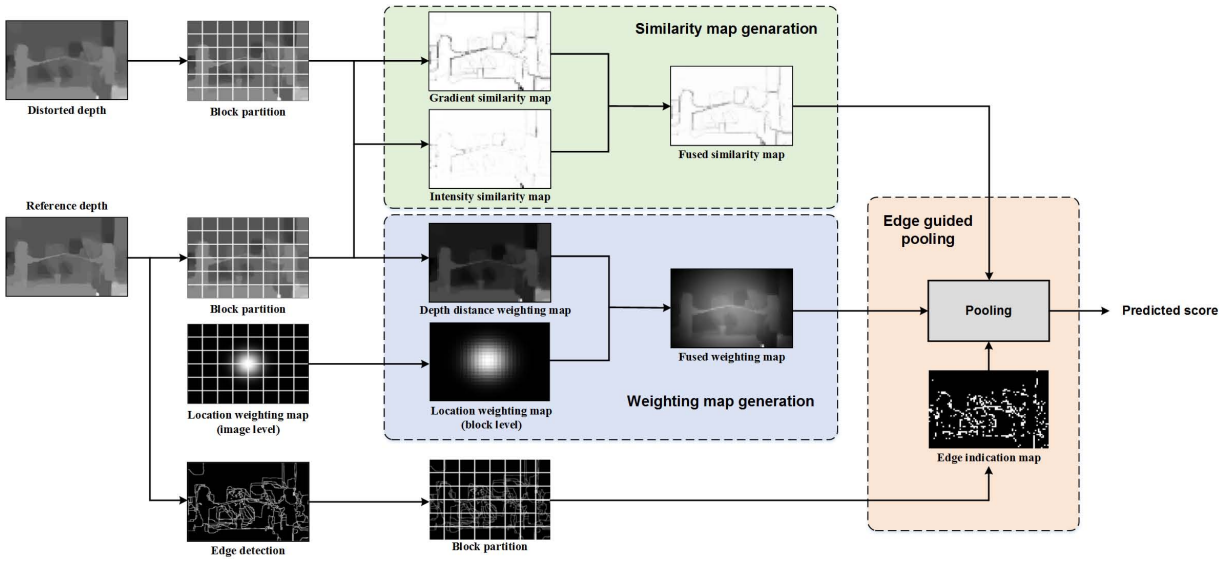

Figure 2. Flowchart of the proposed depth image quality metric.

edge guided pooling. All of them operate based on block partition. In the subsequent subsections, we shall detail on each of them respectively.

3.1. Similarity Map

In the proposed metric, the similarity between a distorted depth image and the corresponding reference image is calculated in both intensity and gradient domains after block partition. In this paper, the reference and distorted depth images are denoted by $\boldsymbol{I}^r$ and $\boldsymbol{I}^d$, and both of them have size $W \times H$. In implementation, we divide the image into blocks of size $M \times M$. The reference and distorted intensity block sets are denoted by $\{\boldsymbol{I}^r_{ij}\}$ and $\{\boldsymbol{I}^d_{ij}\}$, where $i = 1, 2, \cdots, \lfloor \frac{W}{M} \rfloor, j = 1, 2, \cdots, \lfloor \frac{H}{M} \rfloor$, and $\lfloor \cdot \rfloor$ denotes the floor operation. Accordingly, the gradient block sets are denoted by $\{\boldsymbol{G}^r_{ij}\}$ and $\{\boldsymbol{G}^d_{ij}\}$, respectively.

In order to compute the intensity similarity, the average depth intensity of a block is first calculated as

$$v^k_{ij} = \frac{1}{M^2} \sum_{x=1}^{M} \sum_{y=1}^{M} \boldsymbol{I}^k_{ij}(x, y) \tag{1}$$

where $k \in \{r, d\}$ denotes the reference or distorted depth image. Then the intensity similarity can be calculated as

$$S^I_{ij} = \frac{2 \cdot v^r_{ij} \cdot v^d_{ij} + c_1}{(v^r_{ij})^2 + (v^d_{ij})^2 + c_1} \tag{2}$$

where c_1 is a small constant to avoid numerical instability.

Image gradient is typically calculated by convolving with a linear filter. Taken into account computation complexity and anti-noise capability, the Prewitt filter [7] is adopted. Given a depth image block $\boldsymbol{I}^k_{ij}, k \in \{r, d\}$, the gradient map is calculated as

$$\boldsymbol{G}^k_{ij} = \sqrt{(\boldsymbol{G}^k_{ijx})^2 + (\boldsymbol{G}^k_{ijy})^2} \tag{3}$$

$$\boldsymbol{G}^k_{ijx} = \boldsymbol{I}^k_{ij} * \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix} \tag{4}$$

$$\boldsymbol{G}^k_{ijy} = \boldsymbol{I}^k_{ij} * \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix} \tag{5}$$

where $\boldsymbol{G}^k_{ijx}$, $\boldsymbol{G}^k_{ijy}$ denote the gradients in horizontal and vertical directions respectively, and $*$ denotes the convolution. The gradient similarity is then calculated as

$$S^G_{ij} = \frac{1}{M^2} \sum_{x=1}^{M} \sum_{y=1}^{M} \frac{2 \cdot \boldsymbol{G}^r_{ij}(x, y) \cdot \boldsymbol{G}^d_{ij}(x, y) + c_2}{(\boldsymbol{G}^r_{ij}(x, y))^2 + (\boldsymbol{G}^d_{ij}(x, y))^2 + c_2} \tag{6}$$

where c_2 is another small constant, which function similarly with c_1.

With the intensity and gradient similarities, we conduct a simple fusion of them to obtain the overall similarity map, which is defined as

$$S_{ij} = (S^G_{ij})^{\lambda} \cdot (S^I_{ij})^{(1-\lambda)} \tag{7}$$

where $\lambda \in [0, 1]$ is a weight to balance the relative importance between gradient and intensity.

An characteristic of the Human Visual System (HVS) is that human eyes are only sensitive to distortions above a visibility threshold. Considering this, a threshold is employed to process the initial similarity map as follows

$$S_{ij} = \begin{cases} T_S, & if \ \ S_{ij} \geq T_S \\ S_{ij}, & otherwise \end{cases} \tag{8}$$

Figure 3. An example of generated weighting map. (a) Original depth image; (b) Weighting map generated using our method.

where T_S denotes the visibility threshold. Note that S_{ij} is a similarity measure, so lower value indicates bigger difference.

3.2. Weighting Map

In digital images, regions have different contributions to the perceived quality, which is mainly due to the masking effect. In the quality assessment community, visual saliency has been the most popular technique to adapt to the characteristics of the HVS [31]. However, different from natural images, depth images are not for human consumption. So the existing saliency models for natural scene images are not applicable to depth images. In this part, we propose a new weighting strategy for depth images by combining a location prior and a depth distance measure.

It has been demonstrated that human eyes tend to pay more attention to the objects near image center [11], i.e. location prior. Therefore, distortions near the image center cause more damage to the visual quality. Location prior has been adopted in the existing visual saliency modelling [30]. Following this principle, the block-wise location weight can be computed as

$$W_{ij}^{L} = exp\left(-\frac{\|\boldsymbol{L}_{ij}^{r} - \boldsymbol{C}\|_2^2}{\sigma_L^2}\right) \tag{9}$$

where $\boldsymbol{L}_{ij}^{r}$ denotes the coordinate of the block center in relative to the whole image center $\boldsymbol{C}$, and σ_L is a constant to control the weight decaying from inside to outside.

Another characteristic in depth perception is that people are more easily attracted by objects closer to their eyes. Inspired by this, a block depth distance weight is further defined as

$$W_{ij}^{D} = exp\left(\frac{(v_{ij}^{r})^2}{\sigma_D^2}\right) \tag{10}$$

where $\{v_{ij}^{r}\}$ is the average of a depth block $\boldsymbol{I}_{ij}^{r}$, σ_D is another constant to control the decaying rate.

With the location weighting map and depth distance weighting map, we further fuse them to get the final weighting map

$$W_{ij} = W_{ij}^{L} \cdot W_{ij}^{D}. \tag{11}$$

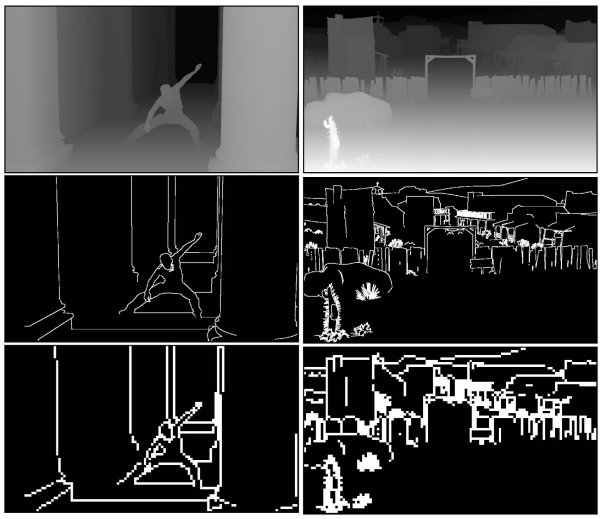

Figure 4. Top: Depth images; Middle: Binary maps detected by Canny operator; Bottom: Edge indication maps.

Figure 3 shows an example of weighting map generated using the proposed method, where brighter regions represent bigger weights and thus are more important in depth quality perception. It is evident that the salient regions detected are consistent with human perception.

3.3. Edge Indication Map

As stated before, the proposed metric is designed with emphasis on measuring the edge characteristics in depth maps. Since our method operates block-wisely, we propose to generate an edge indication map to facilitate the subsequent pooling. The Canny edge detector [4] is first applied on $\boldsymbol{I}^{r}$ to produce a binary edge map $\boldsymbol{C}^{r}$.Then the edge map is also divided into non-overlapping blocks with size $M \times M$, which is denoted by $\{\boldsymbol{C}_{ij}^{r}\}$. Then the number of edge pixels in a block is used to determine whether it is an edge block or not. If a block is a edge block, it is assigned a label '1', otherwise label '0'. Then we get the edge indication map as follows

$$E_{ij} = \begin{cases} 1, & sum(\boldsymbol{C}_{ij}^{r}) \geq \alpha \times M^2 \\ 0, & sum(\boldsymbol{C}_{ij}^{r}) < \alpha \times M^2 \end{cases} \tag{12}$$

where $sum(\cdot)$ counts the number of edge pixels in a block, and $\alpha \in [0, 1]$ controls the classification threshold.

Figure 4 shows two examples of the edge indication maps, from which we know that the edge blocks can be accurately selected.

3.4. Edge Guided Pooling

The overall depth quality score is generated by pooling the similarity map and weighting map with the edge indication map as a guidance. Specifically, with S_{ij}, W_{ij} and E_{ij},

edge guided pooling is performed as

$$S = \frac{\sum_{i=1}^{\lfloor \frac{W}{M} \rfloor} \sum_{j=1}^{\lfloor \frac{H}{M} \rfloor} E_{ij} \cdot S_{ij} \cdot W_{ij}}{\sum_{i=1}^{\lfloor \frac{W}{M} \rfloor} \sum_{j=1}^{\lfloor \frac{H}{M} \rfloor} E_{ij} \cdot W_{ij}}. \quad (13)$$

The final quality score is converted to the range (0, 1] using a logarithmic function [15],

$$Q = log_{(1-T_S)}(1 - S) \quad (14)$$

where T_S is the visibility threshold used in equation (8). A higher score indicates better quality.

4. Experimental Results and Discussions

4.1. Experimental Settings

The performance of the proposed metric is evaluated on depth images from MCL-3D database [23], which was built for view synthesis quality assessment. The database contains nine 2D-image-plus-depth scenes, each with three viewpoints. There are six different types of distortions in the database, including additive white noise (AWN), gaussian blur (GB), down-sampling blur (DB), JPEG compression(JPEG), JPEG2000 compression(JP2K) and transmission error (TE). For each distortion type, four distortion levels are included. There are three configurations in the database, namely view synthesis using (1) undistorted texture and distorted depth, (2) distorted texture and undistorted depth, (3) distorted texture and distorted depth. Since we focus on depth quality, we use images in scenario (1). In this case, undistorted texture and distorted depth images are used in DIBR, so the distortions in the synthesized images are mainly caused by the degraded depth. In total, 648 depth images are used in our experiment. Since there is no annotation for the quality of depth images, we use the Mean Opinion Score (MOS) of the corresponding synthesize image as ground truth.

Four criteria are used for performance evaluation, including Pearson Linear Correlation Coefficient (PLCC), Root Mean Square Error (RMSE), Spearman Rank order Correlation Coefficient (SRCC) and Kendalls Rank Correlation Coefficient (KRCC). PLCC and RMSE measure prediction accuracy, while SRCC and KRCC measure prediction monotonicity. Higher PLCC, SRCC, KRCC values and lower RMSE value represent better performance. Before computing PLCC and RMSE, a five-parameter logistic function is employed to map the predict scores to the range of the subjective scores [3]:

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

where β_1, β_2, β_3, β_4, β_5 are the fitting parameters. Other parameters in the proposed metric are set as follows, $M = 16$, $\alpha = 0.1$, $c_1 = 0.001$, $c_2 = 0.009$, $\lambda = 0.85$, $T_S = 0.998$, $\sigma_L = 114$, $\sigma_D = 122$.

Metric	PLCC	SRCC	KRCC	RMSE
PSNR	0.7947	0.6972	0.5219	0.9726
SSIM [24]	0.8069	0.6418	0.4723	0.9464
GMSD [29]	0.7560	0.7386	0.5510	1.0489
MSSSIM [26]	0.8022	0.6283	0.4635	0.9565
IWSSIM [25]	0.8223	0.6819	0.5112	0.9117
FSIM [32]	0.8290	0.6498	0.4860	0.8961
GSM [16]	0.8151	0.6571	0.4880	0.9281
IGM [27]	0.7991	0.6759	0.4997	0.9633
VIF [22]	0.5268	0.5294	0.3918	1.5893
MAD [12]	0.7786	0.7226	0.5423	1.0053
VSI [31]	0.6796	0.5944	0.4241	1.1754
PSNRHVSM [18]	0.7421	0.7025	0.5240	1.0740
VSNR [5]	0.6072	0.5662	0.4100	1.2730
BDQM [8]	0.3591	0.3623	0.2464	1.7031
BPR [28]	0.5938	0.5539	0.4024	1.2637
Proposed	**0.8816**	**0.8436**	**0.6588**	**0.7562**

Table 1. Performances comparison on the MCL-3D database.

4.2. Performance Evaluation

Comparison with State-of-the-arts. The performance of the proposed metric is compared with two representative depth quality metrics, i.e., BDQM [8] and BPR [28]. We also tested 13 state-of-the-art natural scene image quality metrics, with the aim to have an intuitive understanding on how they perform on depth images. These metrics are PSNR, SSIM [24], GMSD [29],MSSSIM [26], IWSSIM [25], FSIM [32], GSM [16], IGM [27], VIF [22], MAD [12], VSI [31], PSNRHVSM [18] and VSNR [5]. The experimental results are summarized in Table 1. It is evident that the proposed method achieves the best performance among all the tested metrics. The BDQM metric is specifically designed for compression distortion, so it does not perform very well on the whole database, which contains six types of distortions.

To intuitively show the superiority of the proposed metric, Figure 5 shows the scatter plots of the predicted scores versus the subjective scores by different quality metrics. As can be seen from Figure 5, the proposed metric produces the best fitting result and the scatter points gather most closely around the fitted curve. These results indicate that the proposed metric can produce quality scores that correlate best with the subjective ratings.

Comparison of Edge Detectors. In our implementation, the Canny edge detector is adopted to generate the edge map. For comparison, we also included the results when other commonly used edge detectors are utilized, including Roberts [19], Sobel [10], Prewitt [7] and Log [17]. Table 2 summarizes the experimental results. It is easily observed from the table that the Canny edge detection method achieves the best performance. This may attribute to the advantage of Canny operator that it incorporates a mechanism to prevent one edge from multiple responses.

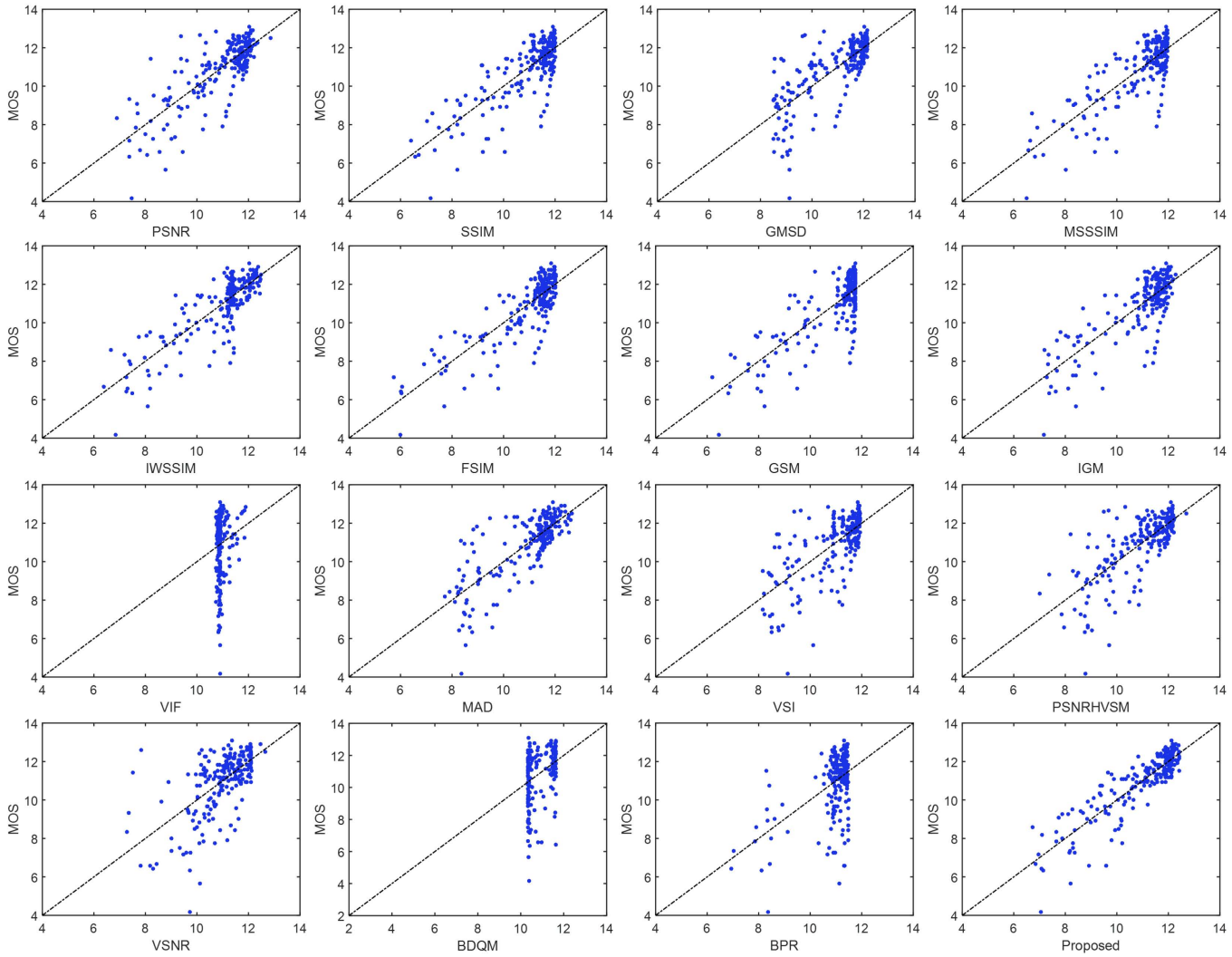

Figure 5. Scatter plots of the subjective scores in the MCL-3D database versus the objective scores predicted by different IQA metrics.

Detector	PLCC	SRCC	KRCC	RMSE
Roberts [19]	0.7820	0.7066	0.5295	0.9986
Sobel [10]	0.7643	0.6974	0.5199	1.0333
Prewitt [7]	0.7664	0.6991	0.5202	1.0292
Log [17]	0.8594	0.8094	0.6242	0.8191
Canny [4]	**0.8816**	**0.8436**	**0.6588**	**0.7562**

Table 2. Performances of the proposed metric when different edge detectors are adopted.

Impact of Block Size. To evaluate the influence of block size on the performance of the proposed method, we tested the performances when different block sizes are adopted, ranging from 4×4 to 64×64. Table 3 summarizes the experimental results of different block sizes, where the best result has been marked in boldface. It can be observed that when block size is 16×16, the performance of the proposed metric is the best. Therefore, block size 16×16 is adopted in this work.

Size	PLCC	SRCC	KRCC	RMSE
4×4	0.7545	0.7852	0.5974	1.0515
8×8	0.8421	0.8152	0.6225	0.8248
16×16	**0.8816**	**0.8436**	**0.6588**	**0.7562**
32×32	0.8624	0.8158	0.6259	0.8126
64×64	0.6213	0.6066	0.4577	1.1838

Table 3. Performances of the proposed metric adopting different block sizes.

Impact of Edge Indication Map. In the proposed metric, edge indication map is employed to reject the non-edge blocks during the pooling. In order to demonstrate the superiority of the edge indication map, we conduct a comparative study. Specifically, we test the performance of the proposed metric by removing the edge indication map while keeping other elements unchanged. Figure 6 shows

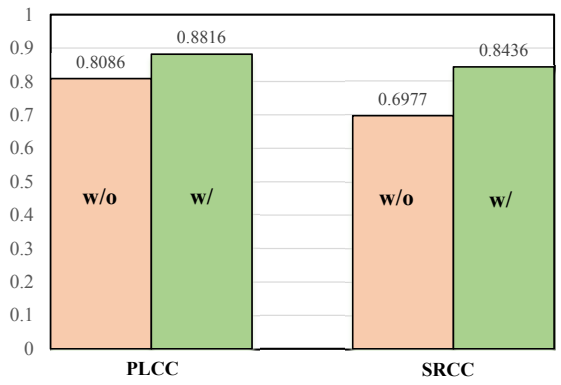

Figure 6. Impact of the edge indication map on the performance of the proposed method.

the performance of the proposed metric with/without using the edge indication map during the pooling stage. It is obvious that by incorporating the edge guided pooling, the performance of the proposed metric improves by a large margin. To be specific, after incorporating the edge indication map, the PLCC and SRCC values are improved by 9.03% and 20.91%, respectively. As a result, the proposed edge indication map is very effective in locating the visually important regions, which facilitate the overall evaluation.

Metric	Without map		With map		Performance gain	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
SSIM [24]	0.8069	0.6418	0.8349	0.8045	3.47%	25.35%
GMSD [29]	0.7560	0.7386	0.8031	0.7687	6.23%	4.08%
MSSSIM [26]	0.8022	0.6283	0.8302	0.7581	3.49%	20.66%
FSIM [32]	0.8290	0.6498	0.8321	0.8036	0.37%	23.67%
GSM [16]	0.8151	0.6571	0.8473	0.8175	3.95%	24.41%

Table 4. Performances of representative IQA metrics with/without using the edge indication map.

In order to further demonstrate the universality of the proposed edge indication map, we replace the similarity map generation part of the proposed method using five popular quality metrics (SSIM [24], GMSD [29], MSSSIM [26], FSIM [32], GSM [16]) and test their performances before and after using the edge indication map. Table 4 summarizes the experimental results as well as a statistics of the performance gains. It is known from the table that the performances of all tested quality metrics improve after incorporating the edge guided pooling. particularly for the monotonicity criterion SRCC, four of them delivered more than 20% performance gains, which is significant. As discussed earlier, distortions in the non-edge regions have little damage to the quality of the synthesized views. The proposed edge indication map is designed by considering the inherent characteristics of depth images, so it is very helpful in determining the visually important edges.

Evaluation of Weighting Map. In the proposed metric, location prior and depth distance of the blocks are combined to generate the weighting map, which is inspired by the HVS. To demonstrate the effectiveness of the weighting map, five representative IQA metrics, including SSIM [24], GMSD [29], MSSSIM [26], FSIM [32], VSI [31], are tested by incorporating the proposed weighting map. In SSIM and MSSSIM, average pooling is used. In GMSD, standard deviation pooling strategy is adopted. In FSIM, phase congruency pooling strategy is employed. In VSI, SDSP saliency detection map [30] is used as pooling map. The proposed method adopts average pooling strategy for performance comparison. To save space, only PLCC and SRCC values are reported here. Table 5 summarizes the performances of the six metrics.

Metric	Original pooling		Propsoed pooling		Performance gain	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
SSIM[24]	0.8069	0.6418	0.7755	0.6047	-3.90%	-5.79%
GMSD[29]	0.7560	0.7386	0.8077	0.6600	6.84%	-10.64%
MSSSIM[26]	0.8022	0.6283	0.7832	0.5932	-2.37%	-5.59%
FSIM[32]	0.8290	0.6498	0.7957	0.6097	-4.02%	-6.17%
VSI[31]	0.6796	0.5944	0.6934	0.5801	2.03%	-2.40%
Proposed	0.8639	0.8353	0.8816	0.8436	2.05%	0.99%

Table 5. Performances of the representative IQA metrics with different pooling strategies.

It can be seen from the table that the performances of most traditional natural scene quality metrics drop after adopting the proposed pooling strategy, while the proposed method achieves a better performance. This indicates that the proposed pooling strategy is designed for depth images, and it is not readily applicable to natural scene images. Although the performance of the proposed metric with average pooling is worse than the final pooling, it still outperforms all the other state-of-the-art metrics. These results not only demonstrate the necessity of the proposed weighting map, but also further confirm the validity of the similarity calculation of the edge blocks.

Performance on Individual Distortion Types. In this part, we evaluate the performances of all the aforementioned metrics on different distortion types in the MCL-3D database. The experimental results are listed in Table 6. Due to space limit, we only report the results on PLCC and SRCC, and in experiment we find that the results on KRCC and RMSE lead to similar conclusion. It can be observed from the table that FSIM achieves the best performance on noise distortion, and MAD delivers the best performance on transmission distortion. For the other four types of distortions, the proposed metric performs the best. Furthermore, all metrics, except VIF and BDQM, perform better on AWN distortion than other types of distortions. The reason may be explained as follows. As has been observed in section 2, AWN distortion in the depth images cause the most serious damage to the quality of synthesized views, so the AWN

Metric	AWN		GB		JP2K		JPEG		DB		TE	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
PSNR	0.8428	0.8347	0.6249	0.5287	0.6293	0.3198	0.6514	0.6429	0.6164	0.4411	0.6063	0.5845
SSIM [24]	0.8606	0.8623	0.4698	0.4261	0.5470	0.3791	0.6818	0.6815	0.6020	0.5455	0.7629	0.7901
GMSD [29]	0.8150	0.7906	0.8008	0.4883	0.6916	0.3836	0.6306	0.5418	0.7887	0.4703	0.6741	0.6771
MSSSIM [26]	0.9046	0.8943	0.4478	0.4443	0.6443	0.3574	0.6850	0.6778	0.4704	0.4842	0.7539	0.7676
IWSSIM [25]	0.8701	0.8892	0.5504	0.5310	0.4821	0.4220	0.6976	0.6581	0.4939	0.4560	0.7736	0.7532
FSIM [32]	**0.9256**	**0.9441**	0.4717	0.3782	0.6602	0.3131	0.6597	0.6199	0.5735	0.5222	0.7750	0.7651
GSM [16]	0.8809	0.8913	0.3827	0.2004	0.6889	0.3929	0.6524	0.6506	0.3977	0.1883	0.7150	0.7391
IGM [27]	0.8273	0.8320	0.4703	0.4059	0.6789	0.3518	0.5846	0.5027	0.5846	0.5652	0.7670	0.7811
VIF [22]	0.5991	0.5837	0.4463	0.4096	0.4967	0.4333	0.5807	0.5499	0.3828	0.2340	0.7536	0.7776
MAD [12]	0.8110	0.7777	0.7749	0.3598	0.5975	0.4278	0.6006	0.5554	0.7713	0.5036	**0.8327**	**0.8486**
VSI [31]	0.7948	0.8502	0.3317	0.2258	0.3039	0.2778	0.5106	0.3305	0.3285	0.3292	0.7115	0.7459
PSNRHVSM [18]	0.8199	0.8275	0.6077	0.4985	0.6136	0.2039	0.5856	0.5129	0.6178	0.4281	0.6762	0.6680
VSNR [5]	0.8052	0.8064	0.4479	0.4319	0.2535	0.1872	0.3716	0.3365	0.2585	0.3803	0.4054	0.5111
BQDM [8]	0.4504	0.4184	0.3046	0.2607	0.5284	0.3538	0.6626	0.6621	0.2524	0.2809	0.5680	0.4813
BPR [28]	0.6263	0.5602	0.4942	0.5384	0.6193	0.5307	0.6072	0.4962	0.4551	0.5490	0.4519	0.5271
Proposed	0.8931	0.8644	**0.8817**	**0.8267**	**0.8141**	**0.8149**	**0.8351**	**0.8538**	**0.8053**	**0.8294**	0.7712	0.7775

Table 6. Performances of different quality metrics on different distortion types in MCL-3D Database.

distortion is easily captured and evaluated.

The proposed metric performs well on AWN, GB, DB, JPEG, JP2K, but performs slightly worse on TE distortion. Since the TE distortion usually occurs in random positions, so it may not be present in the edge blocks. This may be the reason why the proposed metric is not that effective in evaluating the transmission error distortion. Based on these results, we can draw the conclusion that the proposed metric achieves the best overall performance.

5. Conclusion

In this paper, we have proposed a novel depth image quality metric for view synthesis. The proposed method is based on the observation that distortions in depth images mainly affect the edge regions, which in turn degrade the quality of the synthesized images. The proposed metric is a three-step approach, which comprises similarity map generation, weighting map generation and edge guided pooling. All the three stages are designed by considering the characteristics of depth images. So the proposed metric is more effective in evaluating the depth distortions. This has been verified by extensive experiments based on a view synthesis quality database. Comparisons with the state-of-the-art quality metrics also confirm the superiority of the proposed metric.

An inspiration of the proposed metric is that in view synthesis, the quality of a synthesized view can be measured based on the quality of depth and texture images before performing the actual rendering process, which is usually computationally expensive. By this means, a DIBR algorithm can automatically reject 'bad' inputs (regardless of texture or depth images), so that an expected quality of synthesized images can be guaranteed. The current work only focuses on the evaluation of depth images. As future work, it would be interesting to further incorporate texture images to design a pre-rendering depth-texture quality metric for measuring the quality of synthesized views.

6. Acknowledgements

This work is supported by National Natural Science Foundation of China (61771473, 61379143), Natural Science Foundation of Jiangsu Province (BK20181354), Six Talent Peaks High-level Talents in Jiangsu Province (XYDXX-063), and Qing Lan Project of Jiangsu Province.

References

[1] Federica Battisti, Emilie Bosc, Marco Carli, Patrick Le Callet, and Simone Perugia. Objective image quality assessment of 3D synthesized views. *Signal Processing: Image Communication*, 30(1):78–88, 2015.

[2] Emilie Bosc, Romuald Pepion, Patrick Le Callet, Martin Koppel, Patrick Ndjiki-Nya, Muriel Pressigout, and Luce Morin. Towards a new quality metric for 3-D synthesized view assessment. *IEEE Journal of Selected Topics in Signal Processing*, 5(7):1332–1343, 2011.

[3] Kjell Brunnstrom, David Hands, Filippo Speranza, and Arthur Webster. VQEG validation and ITU standardization of objective perceptual video quality metrics [standards in a nutshell]. *IEEE Signal Processing Magazine*, 26(3):96–101, 2009.

[4] John Canny. A computational approach to edge detection. In *Readings in Computer Vision*, pages 184–203. Elsevier, 1987.

[5] Damon M Chandler and Sheila S Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 16(9):2284–2298, 2007.

[6] Pierre-Henri Conze, Philippe Robert, and Luce Morin. Objective view synthesis quality assessment. In *Stereoscopic Displays and Applications XXIII*, volume 8288, page 82881M. International Society for Optics and Photonics, 2012.

[7] R. O. Duda and P. E Hart. Pattern classification and scene analysis. *IEEE Transactions on Automatic Control*, 19(4):462–463, 2003.

[8] Muhammad Shahid Farid, Maurizio Lucenteforte, and Marco Grangetto. Blind depth quality assessment using histogram shape analysis. In *2015 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–5. IEEE, 2015.

[9] Christoph Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291, pages 93–105. International Society for Optics and Photonics, 2004.

[10] Ramesh Jain, Rangachar Kasturi, and Brian G Schunck. *Machine vision*. McGraw-Hill New York, 1995.

[11] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2106–2113. IEEE, 2009.

[12] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010.

[13] Thanh-Ha Le, Seung-Won Jung, and Chee Sun Won. A new depth image quality metric using a pair of color and depth images. *Multimedia Tools and Applications*, 76(9):11285–11303, 2017.

[14] Leida Li, Yu Zhou, Ke Gu, Weisi Lin, and Shiqi Wang. Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness. *IEEE Transactions on Multimedia*, 20(4):914–926, 2018.

[15] Leida Li, Hancheng Zhu, Gaobo Yang, and Jiansheng Qian. Referenceless measure of blocking artifacts by Tchebichef kernel analysis. *IEEE Signal Processing Letters*, 21(1):122–125, 2014.

[16] Anmin Liu, Weisi Lin, and Manish Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, 2012.

[17] R Muthukrishnan and Miyilsamy Radha. Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology*, 3(6):259, 2011.

[18] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, and Vladimir Lukin. On between-coefficient contrast masking of DCT basis functions. In *Proceedings of the Third International Workshop on Video Processing and Quality Metrics*, volume 4, 2007.

[19] Lawrence G Roberts. *Machine perception of three-dimensional solids*. PhD thesis, Massachusetts Institute of Technology, 1963.

[20] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet. DIBR synthesized image quality assessment based on morphological wavelets. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2015.

[21] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet. Multi–scale synthesized view assessment based on morphological pyramids. *Journal of Electrical Engineering*, 67(1):3–11, 2016.

[22] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–709. IEEE, 2004.

[23] Rui Song, Hyunsuk Ko, and C. C. Jay Kuo. MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source. *Journal of Information Science & Engineering*, 31(5):1593–1611, 2015.

[24] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[25] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2011.

[26] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1398–1402. IEEE, 2003.

[27] Jinjian Wu, Weisi Lin, Guangming Shi, and Anmin Liu. Perceptual quality metric with internal generative mechanism. *IEEE Transactions on Image Processing*, 22(1):43–54, 2013.

[28] Sen Xiang, Li Yu, and Chang Wen Chen. No-reference depth assessment based on edge misalignment errors for T+D images. *IEEE Transactions on Image Processing*, 25(3):1479–1494, 2016.

[29] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014.

[30] Lin Zhang, Zhongyi Gu, and Hongyu Li. SDSP: A novel saliency detection method by combining simple priors. In *2013 IEEE International Conference on Image Processing*, pages 171–175. IEEE, 2013.

[31] Lin Zhang, Ying Shen, and Hongyu Li. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014.

[32] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.

[33] Yu Zhou, Liu Yang, Leida Li, Ke Gu, and Lijuan Tang. Reduced-reference quality assessment of DIBR-synthesized images based on multi-scale edge intensity similarity. *Multimedia Tools and Applications*, 77(16):21033–21052, 2018.