Computational Visual Media
https://doi.org/10.1007/s41095-019-0131-6 Vol. 5, No. 2, June 2019, 193–208
Research Article

No-reference synthetic image quality assessment with convolutional neural network and local image saliency

Xiaochuan Wang[1], Xiaohui Liang[1] (✉), Bailin Yang[2], and Frederick W. B. Li[3]

Abstract Depth-image-based rendering (DIBR) is widely used in 3DTV, free-viewpoint video, and interactive 3D graphics applications. Typically, synthetic images generated by DIBR-based systems incorporate various distortions, particularly geometric distortions induced by object dis-occlusion. Ensuring the quality of synthetic images is critical to maintaining adequate system service. However, traditional 2D image quality metrics are ineffective for evaluating synthetic images as they are not sensitive to geometric distortion. In this paper, we propose a novel no-reference image quality assessment method for synthetic images based on convolutional neural networks, introducing local image saliency as prediction weights. Due to the lack of existing training data, we construct a new DIBR synthetic image dataset as part of our contribution. Experiments were conducted on both the public benchmark IRCCyN/IVC DIBR image dataset and our own dataset. Results demonstrate that our proposed metric outperforms traditional 2D image quality metrics and state-of-the-art DIBR-related metrics.

Keywords image quality assessment; synthetic image; depth-image-based rendering (DIBR); convolutional neural network; local image saliency

1 Introduction

With the development of mobile devices and wireless network technology, depth-image-based rendering (DIBR) has become a mainstream technology for supporting remote interactive 3D graphics. Example uses include 3DTV [1], free-viewpoint video [2], stereo-view video [3], and 3D interactive graphics systems [4]. In these DIBR-based systems, a virtual view is synthesized based on various known reference views as the input, which comprise texture and depth information. *3D warping* [5] and *hole filling* [1] are typically applied to generate the required virtual views. However, the process of virtual view synthesis is prone to distortions, degrading the visual quality of the synthetic images. Having a proper quality metric for synthetic images is fundamental to ensuring quality of service (QoS) of DIBR-based systems. Specifically, the feedback from synthetic image assessment can be used to govern optimization of reference view compression and transmission.

As illustrated in Fig. 1, geometric distortions, such as holes, cracks, ghost artifacts, and stretching, are the dominant distortions in a DIBR synthetic

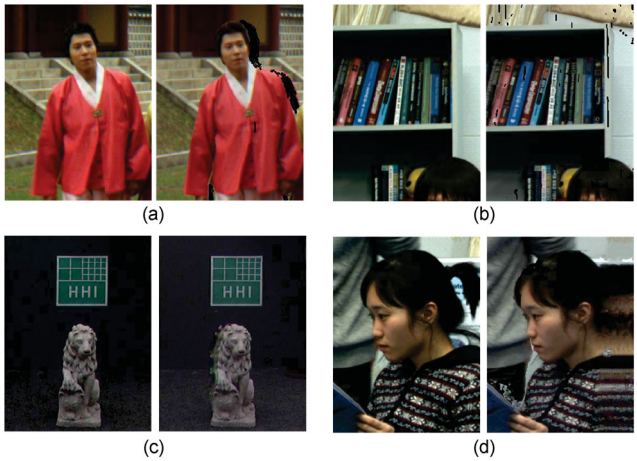

Fig. 1 Geometric distortions in DIBR synthetic images. In each pair, left: undistorted image, right: synthetic image. (a)–(d) exhibit *holes*, *cracks*, *ghost artifacts*, and *stretching*, respectively.

1 State Kay Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China. E-mail: X. Wang, wangxc@buaa.edu.cn; X. Liang, liang_xiaohui@buaa.edu.cn (✉).
2 School of Computer Science & Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China. E-mail: ybl@mail.zjgsu.edu.cn.
3 Department of Computer Science, University of Durham, United Kingdom. E-mail: frederick.li@durham.ac.uk.

Manuscript received: 2018-12-12; accepted: 2019-01-27

image. They mainly result from object dis-occlusion, and rounding errors from 3D warping and hole filling processes. Compared to traditional DCT-based image distortions such as noise, blurring, blocking, and ringing artifacts which are distributed rather uniformly over an image, geometric distortions appear in a non-uniform way and are distributed locally around occlusion regions [6]. Existing 2D image quality assessment (IQA) algorithms focus on structural distortions, and are incapable of properly assessing the visual quality of DIBR synthetic images. So far, only a few works have aimed to evaluate DIBR synthetic images. Most are extensions of existing 2D IQA methods, assuming that DIBR synthetic images follow the same natural scene statistics (NSS) as traditional 2D images [6–9]. Their improvements mainly rely on carefully designed handcrafted features.

In contrast to existing DIBR-related metrics, which heavily rely on handcrafted features, we propose a no-reference (NR) DIBR synthetic image quality assessment method using convolutional neural networks (CNNs) and local image saliency based weighting. Specifically, we exploit the power of CNNs for synthetic image feature extraction, while utilizing the sensitivity of local image saliency to geometric distortions to refine the predicted scores. To overcome the lack of existing training data, we constructed a large DIBR synthetic image dataset with subjective score annotations.

Our main contributions are as follows:

- To our knowledge, we are the first to propose a CNN-based NR-IQA for DIBR synthetic images. In particular, the integration of local image saliency boosts prediction performance.
- We have constructed a new DIBR synthetic image dataset with subjective scores. The capacity and diversity of our proposed dataset is superior to any existing public DIBR image dataset, boosting the training quality and avoiding training bias.
- We have validated the proposed metric on both the public benchmark IRCCyN/IVC DIBR image dataset [10] and our own dataset. Experimental results demonstrate that our method outperforms conventional 2D image metrics and state-of-the-art DIBR-related metrics.

The rest of the paper is organized as follows. Related work is described in Section 2. Section 3 presents our NR-IQA approach, and Section 4 evaluates our proposed algorithm. Application of the proposed metric is demonstrated in Section 5. Finally, Section 6 concludes the paper.

2 Related work

2.1 Image quality assessment

Depending on their need for a priori knowledge of the undistorted image, IQA methods may be broadly categorized as full-reference (FR), reduced reference (RR), and no-reference (NR). In FR-IQA, algorithms typically have full knowledge of the ground truth image, and evaluate image distortion according to pixel error measurements, e.g., SSIM [11]. In contrast, RR-IQA only uses partial information of a reference image for quality evaluation [12]. NR-IQA is the most challenging task, in which algorithms estimate the quality of a distorted image without any information about the ground truth. However, NR-IQA is most suitable for DIBR system usage, since the undistorted image corresponding to a virtual view is typically unavailable. We hence only discuss NR-IQA algorithms in the following.

Most NR-IQA methods are based on NSS priors. Mittal et al. [13] proposed a *Blind/Referenceless Image Spatial Quality Evaluator* (BRISQUE), which extracts point-wise statistics from local normalized luminance signals, measuring image naturalness by the deviations from a natural image model. They also proposed another no-reference metric, *Natural Image Quality Evaluator* (NIQE) [14], without the need for knowing the human subjective score for a distorted image.

Recently, deep learning methods, especially CNNs, have attracted great attention for their powerful image feature extraction capability. Kang et al. [15] firstly introduced CNNs into image quality assessment. In their work, training images are divided into small patches assigned with subjective scores as labels. The small patches are then trained to fit human subjective scores using CNNs. Bosse et al. [16] and Bare et al. [17] improved the prediction performance by weighting the predicted patch scores with image saliency. Bare et al. [17] adopted a more complex network architecture which clusters each minibatch of training patches. In Ref. [18], a pre-trained CNN model is utilized to provide multiple

level features for image quality assessment. GANs are also introduced into NR-IQA [19], where a plausible reference image is generated to assist training. As well as for image quality assessment, deep learning has also been applied in aesthetic evaluation [20]. CNN-based NR-IQA methods have achieved state-of-the-art performance on public 2D image databases, such as LIVE [21], TID2008 [22], and TID2013 [23]. However, no work has been reported for assessing DIBR synthetic images. This is mainly due to the training bias of traditional 2D image datasets, as the features of traditional 2D images and synthetic images are different due to the different natures of their distortions.

2.2 DIBR-related image quality assessment

Previous IQA methods for 2D images are inappropriate for assessing DIBR synthetic images, since the dominant distortions in synthetic images are geometric distortions, as mentioned before. Specifically, holes are mainly induced by object dis-occlusions in a virtual view. Cracks are induced by rounding errors from 3D warping. Ghost artifacts are mainly induced by inaccurate depths, and stretching is due to improper hole filling algorithms. These distortions are quite different from traditional image distortions, such as noise, blurring, blocking, and ringing artifacts induced by DCT-transform based coding and lossy transmission.

Conze et al. [24] aggregated texture, gradient orientation, and contrast information as weighting maps for assessing DIBR synthetic image distortions. Battisti [7] presented an FR synthetic image quality metric. It evaluated a synthetic image by comparing the Kolmogorov–Smirnov distance between the blocks of the synthetic image and the undistorted image. Sandić-Stanković et al. proposed a *Morphological Wavelet Peak Signal-to-Noise Ratio* (MW-PSNR) metric [25] and a *Morphological Pyramids Peak Signal-to-Noise Ratio* (MP-PSNR) metric [26]. Both MW-PSNR and MP-PSNR transform a synthetic image into wavelet domain, and measure the spectral difference between the synthetic image and the undistorted one. Zhou et al. [6] proposed an FR metric for DIBR synthetic images with dis-occluded region discovery. It first detected the dis-occluded regions by comparing the absolute difference between the synthetic image and the undistorted image, and then weighted the predicted quality using the detected dis-occluded regions. Gu et al. [8] proposed an NR method for DIBR synthetic images using local image description. It measured geometric distortions with an auto-regression based NSS model. Tian et al. [9] proposed another NR-IQA method for measuring synthetic image distortions. Four kinds of features, including morphological differences, edges, gradients, and holes ratio, are separately measured and finally aggregated. These DIBR-related metrics achieve significant improvement over conventional IQA metrics, yet heavily rely on handcrafted features.

3 Our approach

We now present the details of our method. As mentioned above, current DIBR-related IQA methods rely heavily on handcraft features, while CNN-based methods suffer from training bias. We hence propose a novel NR-IQA method for synthetic images based on CNNs and local image saliency based weighting. We also address the lack of training data by constructing a new DIBR synthetic image database with sufficient samples.

3.1 Overview

Motivated by previous work, we apply CNNs to train a regression model between predicted image quality scores and human subjective scores. Specifically, the CNN model is assumed to represent the feature subspace of DIBR synthetic images in terms of natural images.

The main bottleneck of CNN-based synthetic image quality prediction is the lack of sufficient training data. Notably, existing CNN-based IQA methods achieve successful results as they are typically trained on very large image databases, e.g., LIVE, CSIQ, TID2008, and TID2013, which contain thousands of images. In contrast, existing public DIBR synthetic image datasets, in particular the IRCCyN/IVC DIBR image dataset, contain only 96 images (including the undistorted images). Our new synthetic image dataset was developed to address the lack of training data.

A CNN model is proposed and trained on our dataset. Particularly, we utilize local image saliency to weight the predicted score, appropriately emphasising the contribution of geometric distortions. The architecture of proposed method is illustrated in Fig. 2. With our trained model, we can predict the

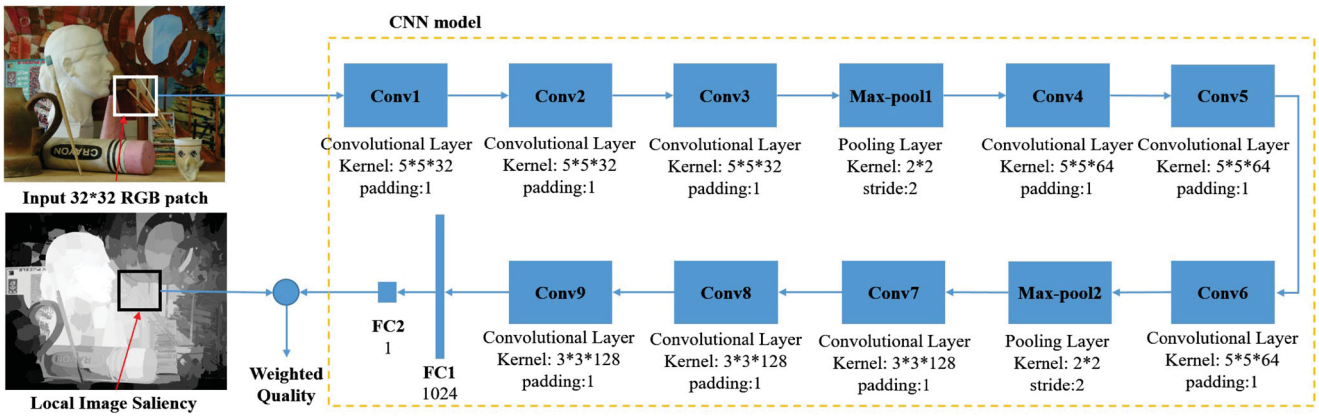

Fig. 2 Architecture of our no-reference synthetic image quality metric. The inputs are small (32×32) patches. The predicted patch scores are weighted by local image saliency.

quality score for test images without knowledge of undistorted versions of them.

3.2 Local image saliency based weighting

Previous work assigns the subjective score of an image to small image patches uniformly [15–17]. It implicitly assumes that the small image patches equally contribute to image quality. In fact, the visual quality of each small image patch is quite different from the whole image quality [27], especially for synthetic images. Suppose a small image patch is exactly covered by a dis-occluded region, and holes dominate an entire patch. As illustrated in Fig. 3, such a patch may be perceived as having better visual quality than that of the whole image. Without knowledge of geometric distortions, a user may simply think that the patch contains a smooth region. Therefore, the strategy of assigning a uniform predicted score to all image patches cannot properly represent the contributions of geometric distortions.

Fig. 3 Visual appearance of image patches containing geometric distortions. Patch A has partial holes, while patch B is dominated by holes. Compared to patch A, patch B is generally perceived as a higher quality image patch, if knowledge of geometric distortions in the whole image is not known.

As performing subjective tests on small image patches is expensive and time-consuming (e.g., a total of 768 subjective tests are required to consider small image patches for each image), a light-weight method of assignment of predicted patch scores is highly desirable. In Ref. [16], the predicted patch score is weighted by image saliency, i.e., salient regions are assigned larger weights. This fits the assumption that observers are generally more sensitive to salient regions, such as the person and chair in Fig. 4(a). The distortions in such salient regions have more influence on the quality of the whole image. However, this only holds for traditional distortions, such as blurring, white noise, and blocking artifacts that are distributed uniformly across the whole image. It is inapplicable to DIBR synthetic images, as geometric distortions in such images are non-uniform and locally-distributed.

Consider Fig. 4. Figure 4(b) shows the saliency map for Fig. 4(a) generated by Ref. [28]. Note that the most salient regions (depicted brighter) are not those regions containing geometric distortions in the synthetic image. For instance, the most salient region in Fig. 4 is the blurred red book, but it is not humanly perceived as distorted. Directly applying image saliency based weighting as proposed in Ref. [27] to the synthetic image thereby overstates the contribution of such regions, while weakening the contribution of local patches containing geometric distortions.

We observe that it makes sense to exploit the difference between the saliency map of a local patch

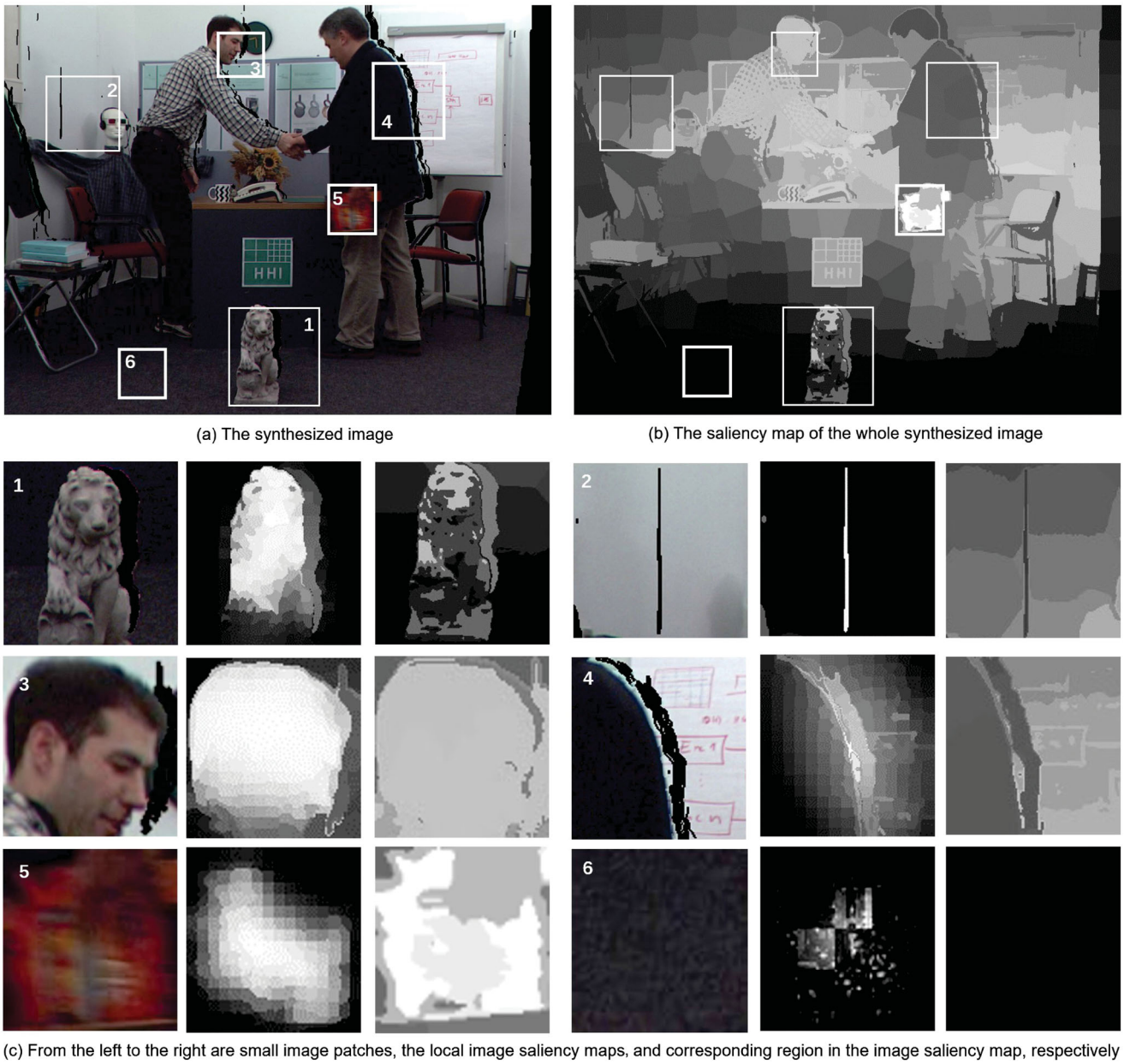

Fig. 4 Saliency maps for a synthetic image and its local patches. (a) Synthetic image. (b) Associated saliency map, with brighter intensity indicating stronger saliency. (c) Six chosen small patches extracted from the synthetic image, the corresponding patch saliency maps using the same saliency model, and the corresponding region extracted from the image saliency map. Note that geometric distortions appear differently in the patch saliency map and the image saliency map.

and its corresponding region of the saliency map for the whole image to help to improve the representation of geometric distortions. As seen in Fig. 4(c), the cracks on the wall are dark (indicating weak saliency) in the whole image but are bright (indicating strong saliency) in the small patch. In reality, human perception is most sensitive to such cracks. We should hence assign a large weight to the corresponding patches. In contrast, the holes appearing at the right side of the lion statue are dark (indicating weak saliency) in both the image saliency map and the patch saliency map. This fits the observation that holes around the lion statue are not perceived to be consistent with the cracks in the white wall. This is partly supported by theories that in the human visual system, texture contrast masking and luminance adaptation conceal distortions to some extent [29]. We can thus give the corresponding patch a small weight. On the other hand, patches containing no geometric distortion share similar appearance of local

patch saliency and corresponding regional saliency in the whole image. For instance, the aforementioned red book with motion blurring appears to be salient in both the patch and the corresponding region of the whole image. However, human perception does not consider motion blurring to be a distortion. In this situation, the contribution of the predicted patch score should be low. The background floor is neither salient at the patch level nor the whole image level, and that should also be considered as unimportant, as shown in Fig. 4(c).

Based on the above observations, we exploit the ratio between the local patch saliency and the corresponding regional saliency in the whole image to represent the contribution of patch scores toward geometric distortions. We define this as *local image saliency*, formulated as follows:

$$c_x = \frac{\sum_{p\in\Omega_x} S(p)}{\sum_{p\in\Omega_x} S'(p)} \tag{1}$$

where Ω_x indicates the region of a small patch. $S(\cdot)$ and $S'(\cdot)$ denote the per-pixel value of patch saliency and the corresponding saliency in the whole image, respectively. The proposed local image saliency is then used to weight the predicted patch scores. For example, a patch with high local image saliency implies that the patch contains clearly visible geometric distortions, and that the predicted score should be increased, and vice versa.

3.3 Network architecture

Our network is mostly inspired by Ref. [15], but is designed to process DIBR synthetic images during preprocessing, and to use local image saliency based weighting.

3.3.1 Preprocessing

Before training, we divide each synthetic image into small patches of size 32×32 pixels. As depicted in Fig. 5, geometric distortions are visible in RGB channels. However, such distortions are concealed after gray-scale transformation and local contrast normalization. Consequently, we abandon gray-scale transform and local contrast normalization, even though they have been widely used in existing CNN-based NR-IQA methods [15, 17]. As a result, important distortion information can be better preserved.

3.3.2 Layers

We use 9 convolutional layers to extract local patch features. Each convolutional layer is followed by a ReLU activation function, which means the local

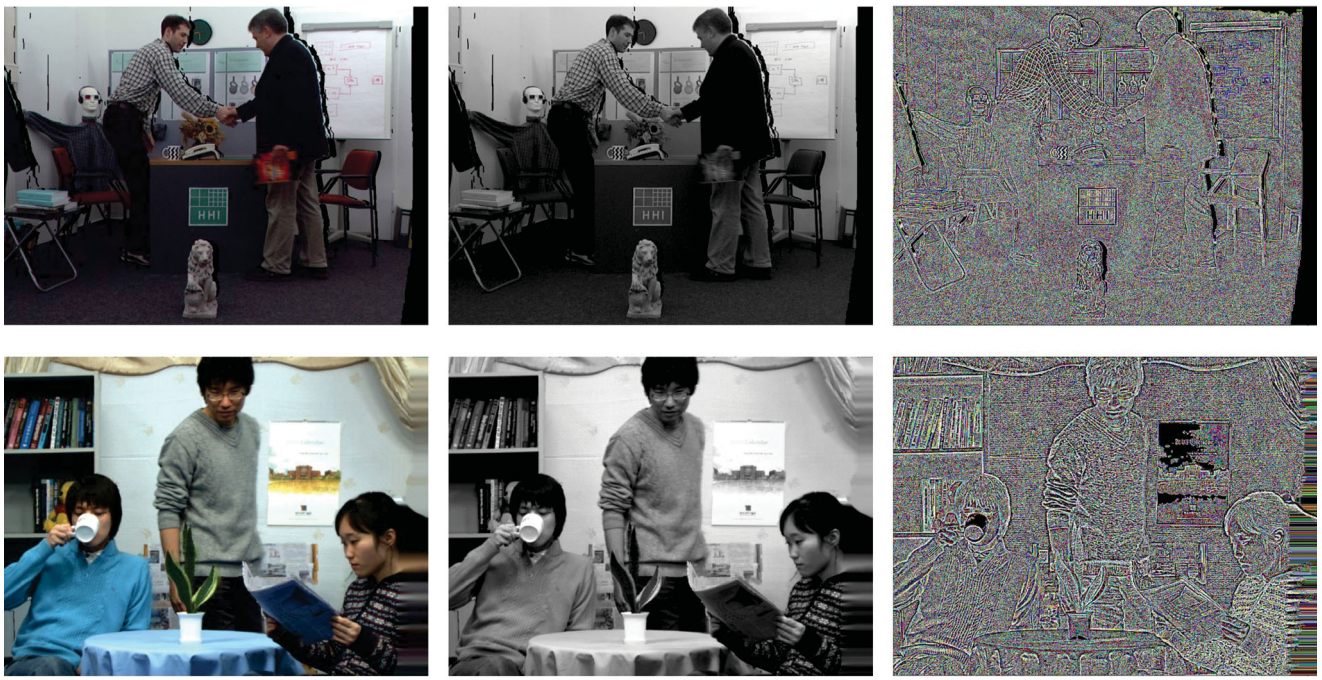

Fig. 5 Visual perception of synthetic images. (a) Two synthetic images. (b) Corresponding gray-scale maps. (c) Visualization of the local normalized maps [15, 17]. Note that holes in regions with high intensity contrast and complex textures are lost after gray-scale transformation and local contrast normalization.

information is extracted into a deeper layer. The convolutional layer can be formulated as

$$C_{j+1} = \max(0, W_j C_j + B_j) \tag{2}$$

where C_j is the feature map of the jth layer, and W_j and B_j are weight and bias respectively. Details of layer configurations as well as kernels are depicted in Fig. 2.

Note that we use a zero-padding strategy, so as to preserve the information at image borders. After three convolutional layers, we apply a max-pooling layer with a 2×2 kernel to enlarge the respective field. We also apply the dropout strategy after the first fully connected layer. The network depth is chosen with the assumption that shallow network architectures capture low-level features while deep network architectures capture semantic features. The effect of network depth is discussed in Section 4.

3.3.3 Optimization

By aggregating the local image saliency based weighting, the loss function is formulated as follows:

$$\min |c_x f(x; \boldsymbol{W}, \boldsymbol{B}) - q^x| \tag{3}$$

where c_x is the local image saliency defined in Eq. (1). x and q^x denote the input small image patch and its assigned subjective quality score, respectively. $f(\cdot)$ outputs the predicted quality score. $\boldsymbol{W}, \boldsymbol{B}$ indicate the trainable weights and biases, respectively. The effectiveness of our proposed local image saliency based weighting is discussed in Section 4. We use the ADAM optimizer to solve this problem.

3.4 Construction of training database

3.4.1 Our DIBR synthetic image database

Until recently, available synthetic image databases with subjective scores were insufficient for training. For instance, the IRCCyN/IVC DIBR image dataset [35] contains only 12 undistorted images and 84 synthetic images. Moreover, these images cover only three scenes: *Book Arrival*, *Newspaper*, and *Lovebird*. All have humans in the center of the scene, which may lead to training bias. The MCL 3D database [36] contains 693 stereoscopic image pairs, which is sufficient for training. However, it lacks subjective scores for each synthetic image. In order to improve training performance, we constructed a new DIBR synthetic image dataset.

A total of 18 reference images were chosen. These reference images ranged from 960×640 to 1920×1080 pixels in size. Twelve reference images were randomly sampled from 3D-HEVC testing video sequences or other typical RGBD databases. Note that the sampled reference images are quite different from those in the IRCCyN/IVC DIBR image dataset. The remaining six reference images were picked from the Middlebury Stereo dataset [34], which only contains indoor objects without people. We specifically chose these reference images to avoid training bias. The reference images are shown in Fig. 6.

Figure 7 shows a scatter plot of spatial information (SI) vs. colorfulness information (CI) for our chosen reference images and IRCCyN/IVC DIBR image dataset, as suggested by Ref. [37]. They show that the SI and CI of our chosen reference images span a larger range than the IRCCyN/IVC DIBR image dataset, indicating that the contents of our dataset are more diverse and more likely to avoid training bias.

For each reference image, we set four camera baselines between the reference view and the virtual

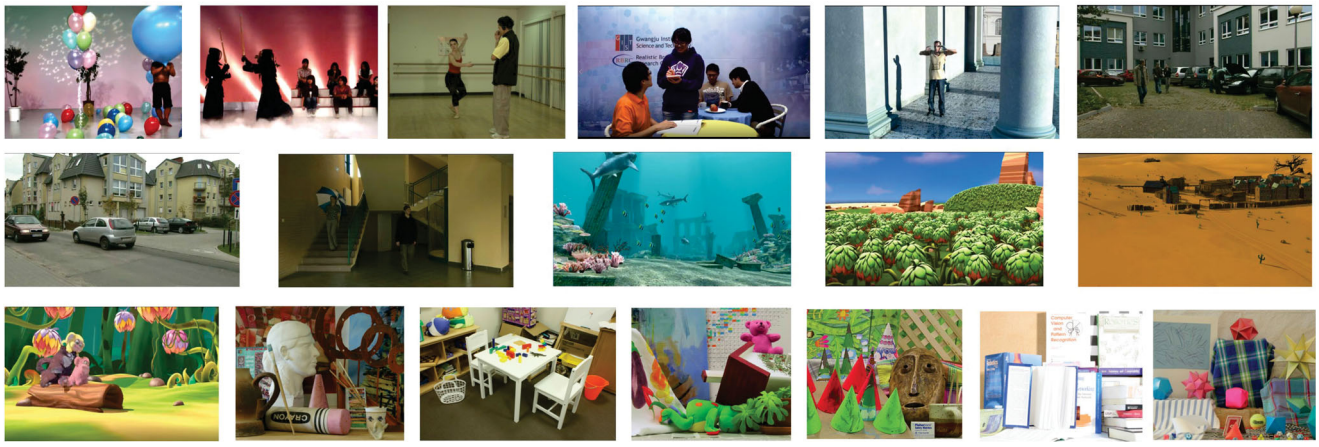

Fig. 6 Reference images from Nayoga Free-viewpoint video dataset [30], Microsoft 3D Video database [31], Poznan Multiview video test sequences [32], Freiburg stereo dataset [33], and Middlebury Stereo dataset [34].

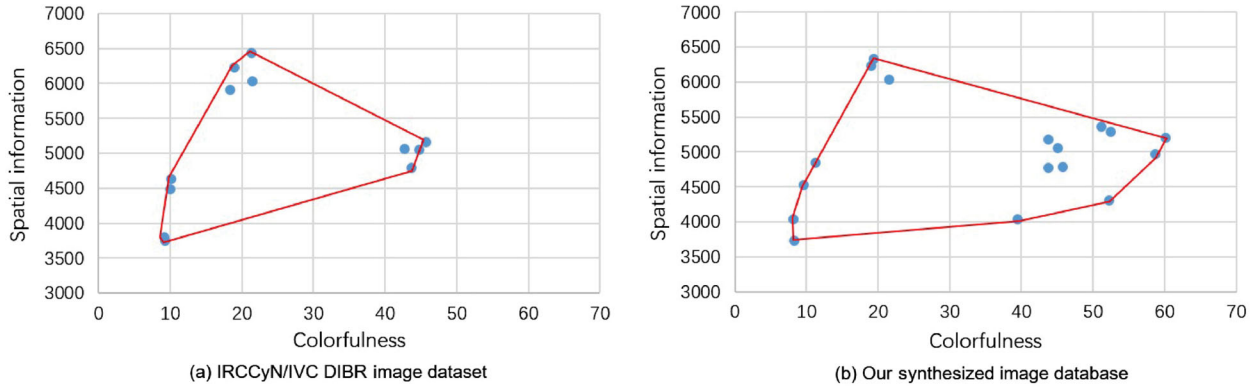

Fig. 7 Spatial information versus colorfulness scatter plots for (a) the IRCCyN/IVC DIBR image dataset and (b) our proposed augmented synthesized image dataset. Red lines indicate the convex hull of the points in each scatter plot, indicating the range of scene diversity.

view. For instance, the camera position of the *Balloons* reference image is denoted by 0, then we select four virtual cameras along the horizon line of the reference camera, while the baselines between the virtual cameras and the reference camera are set to $-2d, -d, +d, +2d$, respectively. d is the preset unit distance. After 3D warping, we conduct 7 hole-filling algorithms on the synthetic images. Finally, we obtain 504 synthetic images. Note that the hole-filling algorithms are the same as those used for the IRCCyN/IVC DIBR image dataset. Details of the hole-filling algorithm are given in Ref. [7]. Compared to the IRCCyN/IVC DIBR image dataset, our new database has over 5 times as many images. Further comparisons are listed in Table 1.

3.4.2 Subjective testing

Since the number of synthetic images was prohibitively large for a double stimulus setup, we instead chose a single stimulus absolute category rating procedure with hidden reference (ACR-HR), as suggested by ITU-T Recommendation P.910 [38]. Each synthetic image was evaluated by 15 observers. Subjective testing was divided into three sub-sessions of 25 min each with a break of five minutes in between to reduce visual fatigue and eye strain. Each testing image was displayed for 15 s, following by a gray image for 5 s. To ensure the robustness of subjective opinion, twelve testing images were randomly displayed repeatedly. The 15 subjects who participated in the test were graduate or undergraduate students with ages ranging from 21 to 31. Two of them had knowledge of IQA, the remainder having no experience of IQA.

Before testing started, the study procedure was explained to each subject. We also obtained verbal confirmation that the subjects had normal or corrected-normal vision. For each sub-session, five images were shown as a warm-up; these had different contents but the same type of distortions as the testing images.

A 24 inch Lenovo X23 LG 0.2 monitor was used as display. It had 16:9 aspect ratio, 0.30 m height, 200 cd·m^{-2} peak luminance, and 1920×1080 display resolution. The testing room was dark with weak ambient lighting. Subjects viewed images from 2.1 m, as suggested in ITU-T Recommendation P.910 [38]. At the end of the image display duration, the test number of the image was displayed on the screen, informing subjects to write down one of the five rankings: *5-Excellent, 4-Good, 3-Fair, 2-Poor, 1-Bad* on their subjective scoring sheets.

3.4.3 Processing of raw subjective scores

The subject rejection procedure outlined in ITU-R BT.500 [39] was used to discard scores from unreliable subjects. The kurtosis of the scores (MOS scores) was firstly used to determine whether the scores assigned by a subject followed a normal distribution. For the normally distributed scores, a subject was rejected whenever more than 5% of the scores assigned by the subject fell outside the range of two deviations from the mean scores; otherwise, the subject was rejected whenever more than 5% of the scores fell

Table 1 Details of our proposed DIBR synthetic image dataset

	IRCCyN/IVC DIBR image dataset	Our image dataset
Scenes	3	18
Reference images	12	18
Content	With people	With & without people
Synthetic images	96	504

outside the range of 4.47 standard deviations from the mean scores. All of the 15 subjects passed the outlier rejection. We further analyzed the scores for the 12 redundant images, finding that most subjects assigned the same scores to these repeated images. This further validated the effectiveness of our subjective testing. Finally, the scores of 15 subjects were averaged.

4 Experimental results

We now provide the details of our experimental settings and give a performance comparison for our proposed DIBR synthetic image quality metric on the benchmark IRCCyN/IVC DIBR image dataset and our own dataset. We also briefly discuss the dependence on proposed strategies, including pre-processing, local image saliency based weighting, and network depth.

4.1 Settings

4.1.1 Training implementation

Two datasets were used in our experiments, including the IRCCyN/IVC DIBR image dataset and our DIBR synthetic image database. We trained the CNN model on our DIBR synthetic image database; the synthetic images were divided into training set, validation set, and testing set according to reference image. The dataset division obeyed the 60%/20%/20% principle. Thus, 10 reference images with their associated distorted images were chosen as training set. The validation set and testing set contained 4 reference images and their distorted images separately. Only the training set and validation set were used during training, while the testing set was kept secret until performance evaluation.

In experiments, we set the ADAM optimizer learning rate $\lambda = 0.0001$, performing stochastic gradient descent (SGD) for 20 epochs in training, and saving the models with the top five Pearson linear correlation coefficient (PLCC) performance on the validation set. For each epoch, the training and validation set were shuffled. We calculated local image saliency weights for the whole image and patches using the saliency model in Ref. [28]. During the testing stage, the predicted scores from the five restored models were averaged.

4.1.2 Evaluation methodology

Three indicators were used to evaluate the performance of our proposed metric, including Pearson linear correlation coefficient (PLCC), root mean square error (RMSE), and Spearman rank order correlation coefficient (SROCC). These indicators measure the consistency, accuracy, and monotonicity between the predicted quality scores and subjective scores. PLCC and SROCC range from 0 to 1, higher values indicating better performance. RMSE ranges from 0 to ∞^+, smaller values indicating better performance.

A total of 13 IQA algorithms were selected for comparison. These methods can be divided into two categories, traditional 2D IQA metrics and DIBR-related IQA metrics. For 2D image quality assessment, we separately choose four FR-IQA methods, including PSNR, SSIM [11], VSNR [40], and FSIM [41], as well as three NR-IQA methods, including BRISQUE [13], NIQE [14], and SSEQ [42]. For DIBR-related methods, four FR-IQA methods, including 3DSwIM [7], MW-PSNR [25], MP-PSNR [26], and SDRD [6], as well as two recently published NR-IQA methods, including APT [8] and NIQSV+ [9], were chosen.

For the sake of fairness of performance comparison, the predicted scores of compared metrics were scaled to the subjective scores, i.e., MOS values via third-order polynomial fitting. The polynomial fitting is conducted as follows, which is suggested by ITU-R BT.500 [39]:

$$MOS_{\mathrm{p}} = as^3 + bs^2 + cs + d \tag{4}$$

where s is the score and a, b, c, d are coefficients of the polynomial fitting function, determined by the predicted results and associated subjective scores. Note that our predicted scores are directly trained to fit the subjective scores, so do not require scaling.

The parameters (if any) in the compared FR-IQA methods were trained on the training dataset, while the predicted scores were fitted using non-linear logistic regression to minimize the errors between the predicted scores and the corresponding subjective scores, as suggested by Ref. [8]. After parameter training, we evaluated each method's performance on the testing dataset. The compared NR-IQA methods were directly evaluated on the testing dataset.

4.2 Performance on the IRCCyN/IVC DIBR image dataset

We now compare the performance of the proposed algorithm on the IRCCyN/IVC DIBR image dataset with state-of-the-art methods. As mentioned before,

we trained the CNN model on the training data of our DIBR image database, where the models with top five PLCC results on the validation dataset were saved. The RMSE, PLCC, and SROCC for our metric using the IRCCyN/IVC DIBR image dataset are listed in Table 2. Our proposed algorithm achieves values of 0.3820, 0.8112, and 0.7520, respectively, which are better than those for competing methods.

From Table 2, we are able to derive two important conclusions.

Firstly, existing IQA algorithms that were designed for traditional 2D images do not perform effectively. The FR-IQA metrics are better than the NR-IQA metrics. FSIM [41] achieves 0.5887, 0.4671, and 0.3286 for RMSE, PLCC, and SROCC, respectively. Note that NR-IQA metrics are not able to predict DIBR synthetic image scores at all well, e.g., NIQE [14] achieves 0.1152 and 0.1181 for PLCC and SROCC, respectively. This is mainly due to dependency on natural image distortion priors. In particular, NIQE predicts image quality by evaluating the effect of distortions in terms of the NSS distribution. As mentioned before, geometric distortions are different from traditional image distortions. The learned model is thus inadequate for assessing DIBR synthetic images.

Secondly, despite the fact that the DIBR-related IQA algorithms perform better than those designed for traditional 2D images, prior methods are still insufficient. The best DIBR-related IQA metric is SDRD [6] that achieves 0.3901, 0.8104, and 0.7610 for RMSE, PLCC, and SROCC, respectively. State-of-the-art NR-IQA metrics, such as APT [8] and NIQSV+ [9] achieve similar performance. Our metric outperforms those two relatively new NR-IQA metrics for DIBR synthetic images, and indeed achieves performance competitive to that of the state-of-the-art FR-IQA metric, SDRD. Note however that SDRD is a full-reference method while ours is independent of reference images.

Table 2 RMSE, PLCC, and SROCC on IRCCyN/IVC DIBR image dataset

	Method	Type	RMSE	PLCC	SROCC
FR	PSNR	2D	0.6018	0.4279	0.4610
	SSIM [11]	2D	0.6185	0.3703	0.3069
	VSNR [40]	2D	0.6614	0.4012	0.4293
	FSIM [41]	2D	0.5887	0.4671	0.3286
	3DSwIM [7]	DIBR	0.4988	0.6623	0.6158
	MW-PSNR [25]	DIBR	0.5351	0.5951	0.6246
	MP-PSNR [26]	DIBR	0.5251	0.6148	0.6274
	SDRD [6]	DIBR	**0.3901**	**0.8104**	**0.7610**
NR	BRISQUE [13]	2D	0.4924	0.3071	0.3201
	NIQE [14]	2D	0.4111	0.1152	0.1181
	SSEQ [42]	2D	0.5258	0.2964	0.2890
	APT [8]	DIBR	0.4546	0.7307	0.7157
	NIQSV+ [9]	DIBR	0.4679	0.7114	0.6668
	Ours	DIBR	**0.3820**	**0.8112**	**0.7520**

4.3 Cross validation

To avoid training bias of our CNN model, we conducted cross validation on our own database. Particularly, we evaluated the RMSE, PLCC, and SROCC of our metric and DIBR-related metrics on the testing set of our database. The results are listed in Table 3.

Our metric achieves the best performance on our DIBR synthetic image database in comparison with other DIBR-related metrics. Note that SDRD [6] is inferior to our method on the new database.

The performance of most existing DIBR-related metrics decreases when tested on our database. This implies that lack of diversity in the IRCCyN/IVC DIBR image dataset has caused training bias. The variation in RMSE on these two databases is shown in Table 4, which shows that RMSE is lower when testing on our database. Note that the RMSE variation of 3DSwIM is the most significant. This is perhaps due to the weighting of face features in 3DSwIM, leading to training bias.

4.4 Ablation study

Several strategies are involved in our method. The most important issues concerning prediction performance are preprocessing, local image saliency based weighting, and network depth. We therefore

Table 3 RMSE, PLCC, and SROCC on testing dataset of our DIBR synthetic image database

	Method	RMSE	PLCC	SROCC
FR	3DSwIM	0.5012	0.6320	0.6117
	MW-PSNR	0.5781	0.5662	0.6028
	MP-PSNR	0.5320	0.6022	0.6113
	SDRD	**0.4071**	**0.7882**	**0.7420**
NR	APT	0.4651	0.7250	0.7081
	NIQSV+	0.4720	0.7106	0.6623
	Ours	**0.3940**	**0.7960**	**0.7461**

Table 4 RMSE on IRCCyN/IVC DIBR image dataset and testing dataset of our DIBR synthetic image database

	3DSwIM	MW-PSNR	MP-PSNR	SDRD	APT	NIQSV+	Ours
IRCCyN/IVC DIBR image dataset	0.6623	0.5951	0.6148	0.8104	0.7307	0.7114	0.8112
Our DIBR synthesized database	0.6320	0.5662	0.6022	0.7882	0.7250	0.7106	0.7960
RMSE performance	−4.80%	−4.50%	−2.00%	−2.70%	−0.70%	−0.10%	−1.80%

conduct an ablation study to demonstrate the effect of these strategies.

4.4.1 Preprocessing

We first evaluated preprocessing. While our preprocessing strategy uses raw images directly, we also implemented gray-scale transformation and local contrast normalization of the training images for comparison; the network architecture remained the same. The RMSE, PLCC, and SROCC values are listed in Table 5.

We can see from Table 5 that our preprocessing strategy achieves better performance on the testing set of our DIBR synthetic image database. It implies that gray-scale transformation and local contrast normalization may discard useful information.

4.4.2 Local image saliency based weighting

To demonstrate the effectiveness of local image saliency based weighting, we separately trained the CNN model with different modalities, i.e., the CNN network without weighting, the same model with image saliency based weighting as deployed in Ref. [17], and our proposed model based on local image saliency weighting. In the first case, the predicted patch scores are averaged to fit the subjective score. In the second case, the predicted patch scores are weighted by image saliency. The utilized image saliency is formulated as follows:

$$c'_x = \frac{\sum_{p\in\Omega_x} S'(p)}{\sum_{p\in I} S'(p)} \tag{5}$$

Note the difference between image saliency based weighting in Eq. (6) and local image saliency based weighting in Eq. (3). Image saliency considers saliency, while local image saliency considers saliency variation between the local region and the whole image. The RMSE, PLCC, and SROCC for the testing dataset of our DIBR synthetic image database are listed in Table 6, which shows that the performance of the unweighted CNN model is greatly improved by using image saliency based weighting, as shown in Ref. [17]. However, our proposed local image saliency based weighting further improves the indicators on the testing dataset. This implies that local image saliency based weighting is better for assessing DIBR synthetic images.

A visualization of local image saliency based weighting is given in Fig. 8. Figure 8(a) represents the saliency map of the entire image, while Fig. 8(b) represents saliency maps of small patches, merged into an entire image-sized map. Figure 8(c) visualizes the actually used local image saliency based weights, as calculated by Eq. (3). Clearly, the weights from the saliency map and local image saliency are quite different. The red box in Figs. 8(a) and 8(c) shows cracks in the wall assigned a low weight by the saliency map but a high weight by our proposed local image saliency: local image saliency based weighting provides a better representation of the contributions of patch scores.

4.4.3 Network depth

A deeper network architecture is suggested [16] to achieve better prediction performance on traditional 2D image databases. We validated this assumption on our augmented DIBR synthetic image dataset. Figure 9 shows how RMSE varies with different network depths, i.e., number of convolutional layers. We observe that RMSE decreases on both the training dataset and validation dataset with increasing network depth, agreeing with the assumption that greater network depth benefits prediction performance. However, the performance gain, significantly decreases when the network depth

Table 5 RMSE, PLCC, and SROCC for the testing set of our DIBR synthetic image database with different preprocessing strategies

	RMSE	PLCC	SROCC
With preprocessing	0.4251	0.7420	0.7122
No preprocessing	**0.3940**	**0.7960**	**0.7461**

Table 6 RMSE, PLCC, and SROCC for the testing dataset of our DIBR synthetic image database with different network modalities

	RMSE	PLCC	SROCC
Without weighting	0.4630	0.6920	0.6761
With image saliency	0.4120	0.7420	0.7228
With local image saliency	**0.3940**	**0.7960**	**0.7461**

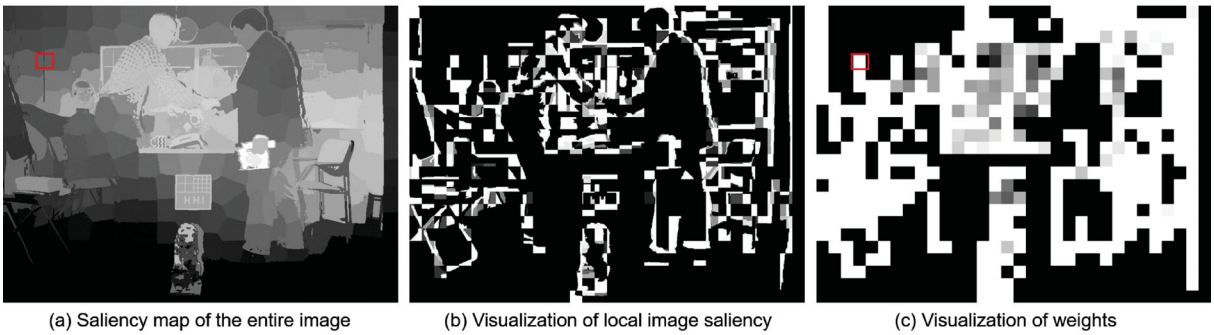

Fig. 8 Visualization of local image saliency based weighting. (a) Saliency map of the entire distorted image. (b) Merged saliency maps of the associated small image patches. All saliency maps were produced by Ref. [28]. (c) Local image saliency based weights, brighter blocks indicating higher weights.

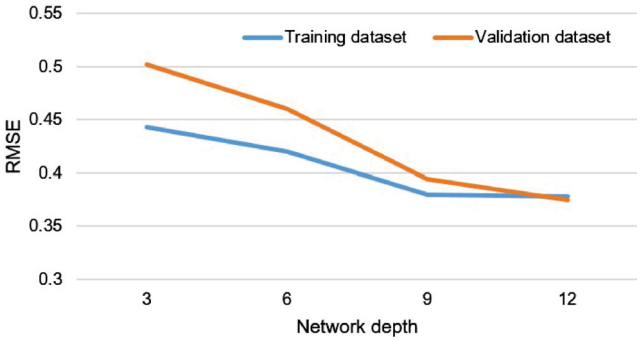

Fig. 9 Performance of CNN models with different network depths (numbers of convolutional layers).

exceeds nine. Also, deeper convolutional layers may lead to overfitting on the validation dataset unless care is taken. In practice, we use a network architecture with nine convolutional layers.

5 Application

The quality of synthetic images is key to the success of DIBR-based systems. For instance, a quality measure can be used to guide the coding of reference texture images and depth map. It can also be used to evaluate hole-filling algorithms. Here we use the proposed synthetic image quality metric to optimize the prediction of reference viewpoints. We first describe the baseline model of reference viewpoint prediction, and then introduce a novel model using our proposed metric.

5.1 Baseline model of reference viewpoint prediction

Suppose a user navigates within a virtual environment. Reference viewpoints are predicted according to user movement, and for each, an associated reference texture image and depth are transmitted to the user-end for virtual view synthesis. Ideally, reference viewpoint prediction is frequent, to reduce errors. However, the bottleneck of reference viewpoint transmission is bandwidth: the reference data which can be transmitted are limited. Previous work [43, 44] adopts a strategy that predicts reference viewpoints with a constant frequency. Shi et al. [45] adopts another mechanism that predicts the reference viewpoint when the MSE between the synthetic image and the undistorted image exceed preset thresholds. We choose these two models as baselines to demonstrate the effectiveness of our proposed metric. Following Ref. [45], we predict reference viewpoints by assessing the quality of the synthetic images. However, our metric requires no reference, and can be directly used to assess the synthetic images without need for the undistorted images.

5.2 Performance

Suppose the user navigates the virtual environment along a horizontal path. The path is equally sampled, and each sample indicates a virtual viewpoint. The positions of these virtual viewpoints can then be denoted as $(\cdots, v_{-1}, v_0, v_1, \cdots)$, where v_0 denotes the initial viewpoint. Figure 10 shows the undistorted image and the synthetic images for v_0. Note that the two synthetic images utilize different reference viewpoint predicted by MSE and our proposed metric.

We can see from Fig. 10 that the two synthetic images can hardly be distinguished. However, the predicted reference viewpoints are v_4 using MSE and v_7 using the proposed metric, respectively. We choose the predicted reference viewpoint as the new initial viewpoint, repeating the reference viewpoint prediction until the virtual viewpoint reaches v_{100}. A

(a) The undistorted image of v_0 (b) The synthesized image with the reference viewpoint v_4 (c) The synthesized image with the reference viewpoint v_7

Fig. 10 Visual quality of synthetic images with different predicted reference viewpoints. (a) Undistorted image of v_0. (b) Synthetic image of v_0 using the reference view of v_4, as suggested by MSE. (c) Synthetic image of v_0 using the reference view of v_7, as suggested by our metric.

total of 25 reference viewpoints are suggested by MSE, while only 17 reference viewpoints are suggested by our proposed metric. By doing so, the transmitted reference data is reduced while the visual quality maintained.

We also simulated virtual environment navigation on a Nexus 5 device. The reference data was transmitted to the client when the quality of the synthetic image fell below a preset threshold. We tested bandwidth required by MSE-based reference viewpoints and ours. See Table 7: our metric saves 29% bandwidth on average in comparison to the metric in Ref. [45].

Table 7 Transmission frequency and average bandwidth cost of different reference viewpoint selection models

Model	Trans. freq.	Avg. bandwidth cost
Bao and Gourlay [43]	5.0 fps	6.90 Mbps
Shi et al. [45]	2.4 fps	3.31 Mbps
Ours	1.7 fps	2.35 Mbps

6 Conclusions

Compared to existing DIBR-related IQA methods, there are some highlights of our work. Firstly, it is the first CNN-based NR-IQA method for DIBR synthetic images, achieving significant performance improvements over state-of-the-art 2D and DIBR-related IQA methods. Our proposal to use local image saliency based weighting further benefits prediction performance. Secondly, we have designed a diverse DIBR synthetic image dataset, which helps to reduce training bias in our CNN model. Although we have achieved competitive performance on DIBR synthetic images, there is still room to improve. For instance, the assignment of patch scores needs further consideration to better fit human perception. In future, we hope to improve the proposed metric by integrating local image saliency in an end-to-end framework.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments. They would also thank Kai Wang and Jialei Li for their assistance in dataset construction and public release. The work was sponsored by the National Key R&D Program of China (No. 2017YFB1002702), and the National Natural Science Foundation of China (Nos. 61572058, 61472363).

References

[1] Fehn, C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In: Proceedings of the SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI, 93–104, 2004.

[2] Smolic, A. 3D video and free viewpoint video: From capture to display. *Pattern Recognition* Vol. 44, No. 9, 1958–1968, 2011.

[3] Smolic, A.; Mueller, K.; Merkle, P.; Kauff, P.; Wiegand, T. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution. In: Proceedings of the Picture Coding Symposium, 1–4, 2009.

[4] Wang, X.; Liang, X.; Yang, B.; Li, F. W. Scalable remote rendering using synthesized image quality assessment. *IEEE Access* Vol. 6, 36595–36610, 2018.

[5] Mark, W. Post-rendering 3D image warping: Visibility, reconstruction, and performance for depth-image warping. Technical Report. Chapel Hill, NC, USA, 1999.

[6] Zhou, Y.; Li, L.; Gu, K.; Fang, Y.; Lin, W. Quality assessment of 3D synthesized images via disoccluded region discovery. In: Proceedings of the IEEE International Conference on Image Processing, 1012–1016, 2016.

[7] Battisti, F.; Bosc, E.; Carli, M.; Le Callet, P.; Perugia, S. Objective image quality assessment of 3D synthesized views. *Signal Processing: Image Communication* Vol. 30, 78–88, 2015.

[8] Gu, K.; Jakhetiya, V.; Qiao, J. F.; Li, X.; Lin, W.; Thalmann, D. Model-based referenceless quality metric of 3D synthesized images using local image description. *IEEE Transactions on Image Processing* Vol. 27, No. 1, 394–405, 2018.

[9] Tian, S.; Zhang, L.; Morin, L.; Déforges, O. NIQSV+: A no-reference synthesized view quality assessment metric. *IEEE Transactions on Image Processing* Vol. 27, No. 4, 1652–1664, 2018.

[10] Bosc, E.; Pepion, R.; Le Callet, P.; Koppel, M.; Ndjiki-Nya, P.; Pressigout, M.; Morin, L. Towards a new quality metric for 3-D synthesized view assessment. *IEEE Journal of Selected Topics in Signal Processing* Vol. 5, No. 7, 1332–1343, 2011.

[11] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* Vol. 13, No. 4, 600–612, 2004.

[12] Sharifi, K.; Leon-Garcia, A. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology* Vol. 5, No. 1, 52–56, 1995.

[13] Mittal, A.; Moorthy, A. K.; Bovik, A. C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* Vol. 21, No. 12, 4695–4708, 2012.

[14] Mittal, A.; Soundararajan, R.; Bovik, A. C. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters* Vol. 20, No. 3, 209–212, 2013.

[15] Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1733–1740, 2014.

[16] Bosse, S.; Maniry, D.; Wiegand, T.; Samek, W. A deep neural network for image quality assessment. In: Proceedings of the IEEE International Conference on Image Processing, 3773–3777, 2016.

[17] Bare, B.; Li, K.; Yan, B. An accurate deep convolutional neural networks model for no-reference image quality assessment. In: Proceedings of the IEEE International Conference on Multimedia and Expo, 1356–1361, 2017.

[18] Kim, J.; Nguyen, A.; Ahn, S.; Luo, C.; Lee, S. Multiple level feature-based universal blind image quality assessment model. In: Proceedings of the 25th IEEE International Conference on Image Processing, 291–295, 2018.

[19] Lin, K.-Y.; Wang, G. Hallucinated-IQA: No-reference image quality assessment via adversarial learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 732–741, 2018.

[20] Zhang, F.-L.; Wu, X.; Li, R.-L.; Wang, J.; Zheng, Z.-H.; Hu, S.-M. Detecting and removing visual distractors for video aesthetic enhancement. I*EEE Transactions on Multimedia* Vol. 20, No. 8, 1987–1999, 2018.

[21] Sheikh, H. R.; Wang, Z.; Cormack, L.; Bovik, A. C. Live image quality assessment database release 2 (2005). 2016. Available at `http://live.ece.utexas.edu/research/quality`.

[22] Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics* Vol. 10, No. 4, 30–45, 2009.

[23] Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; Jay Kuo, C.-C. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* Vol. 30, 57–77, 2015.

[24] Conze, P.-H.; Robert, P.; Morin, L. Objective view synthesis quality assessment. In: Proceedings of the SPIE 8288, Stereoscopic Displays and Applications XXIII, 82881M, 2012.

[25] Sandić Stanković, D.; Kukolj, D.; Le Callet, P. DIBR synthesized image quality assessment based on morphological wavelets. In: Proceedings of the 7th International Workshop on Quality of Multimedia Experience, 1–6, 2015.

[26] Sandić Stanković, D.; Kukolj, D.; Le Callet, P. DIBR-synthesized image quality assessment based on morphological multi-scale approach. *EURASIP Journal on Image and Video Processing* Vol. 2017, 4, 2017.

[27] Heng, W.; Jiang. T. From image quality to patch quality: An image-patch model for no-reference image quality assessment. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1238–1242, 2017.

[28] Zhu, W.; Liang, S.; Wei, Y.; Sun, J. Saliency optimization from robust background detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2814–2821, 2014.

[29] Yang, X.; Ling, W.; Lu, Z.; Ong, E. P.; Yao, S. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication* Vol. 20, No. 7, 662–680, 2005.

[30] Kimata, H.; Kitahara, M.; Kamikura, K.; Yashima, Y. Free-viewpoint video communication using multi-view video coding. *NTT Technical Review* Vol. 2, No. 8, 21–26, 2004.

[31] Zitnick, C. L.; Kang, S. B.; Uyttendaele, M.; Winder, S.; Szeliski, R. High-quality video view interpolation using a layered representation. *ACM Transactions on Graphics* Vol. 23, No. 3, 600–608, 2004.

[32] Domañski, M.; Grajek, T.; Klimaszewski, K.; Kurc, M.; Stankiewicz, O.; Stankowski, J.; Wegner, K. Poznan multiview video test sequences and camera parameters. ISO/IEC JTC1/SC29/WG11 MPEG, M17050, 2009.

[33] Mayer, N.; Ilg, E.; Hausser, P.; Fischer, P.; Cremers, D.; Dosovitskiy, A.; Brox, T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4040–4048, 2016.

[34] Hirschmuller, H.; Scharstein, D. Evaluation of cost functions for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–8, 2007.

[35] Bosc, E.; Pépion, R.; Le Callet, P.; Köppel, M.; Ndjiki-Nya, P.; Morin, L.; Pressigout, M. Perceived quality of DIBR-based synthesized views. In: Proceedings of SPIE 8135, Applications of Digital Image Processing XXXIV, 81350I, 2011.

[36] Song, R.; Ko, H.; Jay Kuo, C.-C. MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source.*Journal of Information Science and Engineering* Vol. 31, 1593–1611, 2015.

[37] Winkler, S. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing* Vol. 6, No. 6, 616–625, 2012.

[38] I. T. Union. ITU-R BT.910. In: Subjective video quality assessment methods for multimedia applications. 1999.

[39] I. T. Union. ITU-R BT.500-12. In: Recommendation: Methodology for the subjective assessment of the quality of television pictures. 1993.

[40] Chandler, D. M.; Hemami, S. S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing* Vol. 16, No. 9, 2284–2298, 2007.

[41] Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* Vol. 20, No. 8, 2378–2386, 2011.

[42] Liu, L.; Liu, B.; Huang, H.; Bovik, A. C. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication* Vol. 29, No. 8, 856–863, 2014.

[43] Bao, P.; Gourlay, D. Low bandwidth remote rendering using 3D image warping. In: Proceedings of the International Conference on Visual Information Engineering. Ideas, Applications, Experience, 61–64, 2003.

[44] Bao, P.; Gourlay, D. A framework for remote rendering of 3-D scenes on limited mobile devices. *IEEE Transactions on Multimedia* Vol. 8, No. 2, 382–389, 2006.

[45] Shi, S.; Nahrstedt, K.; Campbell, R. A real-time remote rendering system for interactive mobile graphics. *ACM Transactions on Multimedia Computing, Communications, and Applications* Vol. 8, No. 3s, Article No. 46, 2012.

Xiaochuan Wang received his M.Sc. degree from Beihang University, Beijing, China, in 2012. He is currently pursuing his Ph.D. degree in the State Key Laboratory of Virtual Reality Technology and System, Beihang University, China. His current research interests include mobile graphics, remote rendering, image quality assessment, and multiview video systems.

Xiaohui Liang received his Ph.D. degree in computer science and engineering from Beihang University in 2002. He is currently a professor in the State Key Laboratory of Virtual Reality Technology and System, Beihang University. His research interests include computer graphics, animation, visualization, and virtual reality.

Bailin Yang received his Ph.D. degree in the Department of Computer Science from Zhejiang University in 2007. He is currently a professor in the Department of Computer and Electronic Engineering of Zhejiang Gongshang University. His research interests are in mobile graphics, real-time rendering, and mobile games.

Frederick W. B. Li received his Ph.D. degree in computer science from City University of Hong Kong in 2001. He is currently an assistant professor at Durham University, UK. Before that, he was an assistant professor in Hong Kong Polytechnic University and project manager of a Hong Kong Government Innovation and Technology Fund (ITF) project. His research

interests include distributed virtual environments, computer graphics, and e-learning systems. Dr. Li has served as a guest editor of special issues of the *International Journal of Distance Education Technologies* and the *Journal of Multimedia*. He has served on conference committees of a number of conferences, including as Program Co-Chair of ICWL 2007-08, 2013, 2015, and IDET 2008-09, and Workshop Co-Chair of ICWL 2009 and U-Media 2009.