IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 7, 2018, accepted December 29, 2018, date of publication January 7, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891070

Quality Assessment of 3D Synthesized Images via Measuring Local Feature Similarity and Global Sharpness

XUEJIN WANG[1], FENG SHAO[1], (Member, IEEE), QIUPING JIANG[1], (Student Member, IEEE), RANDI FU[1], AND YO-SUNG HO[2], (Fellow, IEEE)

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[2]School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea

Corresponding author: Feng Shao (shaofeng@nbu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61622109, in part by the Zhejiang Natural Science Foundation of China under Grant R18F010008, in part by the Natural Science Foundation of Ningbo under Grant 2017A610112, and in part by the by K. C. Wong Magna Fund in Ningbo University.

ABSTRACT Depth-image-based rendering (DIBR) techniques can be used to generate virtual views for free-viewpoint video application. However, the DIBR algorithms will introduce geometric distortions that mainly distribute at the disoccluded regions in the synthesized views. It has been demonstrated that conventional 2-D quality metrics are not suitable for the synthesized views. In this paper, we propose a new quality model for 3-D synthesized images by measuring the block-wise texture similarity and color contrast similarity in critical areas, and the global gradient magnitude deviation. A critical area detection module is first employed using a warping method with morphological operation. Then, the critical areas are partitioned into blocks, which are classified as edge blocks, texture blocks, and smooth blocks by computing discrete cosine transform coefficient values. Block-wise texture similarity and color contrast similarity in the corresponding areas are calculated, which are weighted by the size of critical areas. Furthermore, gradient magnitude deviation is measured to quantify global sharpness. Finally, the two scores are pooled to obtain the overall quality. The experimental results on the IRCCyN/IVC, IETR, and MCL-3-D DIBR image databases indicate that our method achieves higher quality prediction accuracy than the state-of-the-art quality metrics.

INDEX TERMS Quality assessment, depth-image-based rendering, 3D synthesized image, view synthesis.

I. INTRODUCTION

In recent years, 3D video applications, such as 3D Television (3DTV) and Free-viewpoint video (FVV) [1], [2], have received tremendous attention due to the capabilities for providing the viewers with deep depth sensation in viewing the scene from multiple viewpoints. To create dense viewpoints for scene representation, Multiview Video plus Depth (MVD) format [3] is widely used due to its convenience, which can synthesize arbitrary viewpoint via Depth-Image-Based Rendering (DIBR) technique [4]. As explored in [5]–[7], imperfect texture images or depth maps will affect the generation of synthesized views. Thus, the perceptual quality of synthesized views is an important indicator to evaluate the performance of different MVD-based 3D video generators.

Typical DIBR operation consists of two stages, namely warping and rendering. In the warping stage, a texture image is mapped to 3D space based on the corresponding depth information, and followed by an inverse mapping from the 3D space to a new virtual view. Some regions occluded in the original view may become visible in the virtual view, leading to disoccluded regions. In the rendering stage, the disoccluded regions are filled to produce the final synthesized image. However, due to the imperfect rendering, DIBR algorithms may introduce some new types of distortions [8] which are quite different from the traditional distortions, e.g., blur, noise, or compression. Moreover, the DIBR-synthesized artifacts only occur in disoccluded regions. Therefore, objective quality assessment metrics are demanded for predicting the quality of DIBR-synthesized images.

To date, a number of objective quality assessment approaches [9]–[14] have been proposed for DIBR-synthesized images, and several publicly available databases [9], [15], [16] have been released. Although most of these specifically

designed metrics for synthesized images perform better than traditional image quality assessment (IQA) metrics, their performance still has large room for improvement. Towards this end, this paper proposes a new quality assessment metric via measuring local feature similarity and global sharpness. Based on the visual observations that the human perceives blur distortions mostly in texture and edge regions, and is sensitive to color artifacts in homogeneous regions, local texture and color features are measured from different regions to investigate how much these features in the reference image are preserved or changed in the synthesized image. The main contributions of this work are three-folds: 1) We combine local feature similarity and global sharpness for quality assessment of synthesized images to achieve a more consistent evaluation with human perception; 2) We employ discrete cosine transform (DCT) based texture classification method to partition the critical areas into edge block (ED), texture block (TE) and smooth block (SM), and evaluate three types of blocks independently based on their perceptual properties; 3). We measure the texture similarity by comparing the number of edge and texture blocks in the reference and synthesized images weighted by the size of critical areas.

The remainder of this paper is organized as follows. In Section II, we review the related work, detail our method in Section III, and finally present the experimental results in Section IV and the discussion in Section V.

II. RELATED WORK

A. QUALITY ASSESSMENT OF 3D SYNTHESIZED IMAGES

Current quality assessment metrics for synthesized images can be divided into three categories: local quality assessment only for the disoccluded regions, global quality assessment for the synthesized images, and quality assessment combining local and global quality cues.

Only a few local approaches were proposed for 3D synthesized images [9], [17]–[19]. Bosc *et al*. [9] first calculated the distance map between reference and synthesized images, and employed a threshold to extract critical areas. The mean structural similarity on the critical areas was measured as the final quality score. Shao *et al*. [17] proposed a Color and Sharpness of Edge Distortion (CSED) algorithm, in which color distortion was measured by calculating the luminance loss of the disoccluded regions, and the sharpness of edge distortion was measured by calculating a depth weighted proportion of remaining edge to the original edge. Jung *et al*. [18] proposed a Critical Binocular Asymmetry (CBA) metric designed for characterizing the binocular asymmetry property of human eyes. In the method, critical areas were first detected using the synthesized left-, right- view images and the corresponding disparity maps. The average SSIM scores of the critical areas in left- and right- view images were measured as the overall CBA score. Gu *et al*. [19] employed an autoregression prediction model to capture the geometric distortions, and then an optimized procedure was utilized by taking the saliency map into account. However, the performances of the local approaches are limited due to only considering the local distortion areas.

For global approaches, besides traditional 2D quality metrics [20]–[28], some quality metrics were proposed for 3D synthesized images [12], [29]–[31]. Conze *et al*. [29] devised a SSIM-based View Synthesis Quality Assessment (VSQA) metric, which was modulated with three visibility weighting maps, including textural complexity, orientations and image contrast. Battisti *et al*. [12] proposed a 3D Synthesized view Image Quality Metric (3DSwIM) metric relied on statistical features extracted from wavelet subbands of 3D synthesized images. Moreover, based on the assumption that viewers were more sensitive to distortion around human subjects, a skin detector was employed as an approximate visual attention model. Stankovic *et al*. [30] presented an image quality metric, where Morphological bandpass wavelet transform was first employed to decompose the images and multi-scale wavelet mean squared error was obtained. The Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) score was calculated based on multi-scale values. In addition, the Reduced version of MW-PSNR (RMW-PSNR) was presented in [31]. Compared with MW-PSNR, RMW-PSNR only using the wavelet subbands at higher decomposition scales. The Morphological Pyramids Peak Signal-to-Noise Ratio (MP-PSNR) and Reduced version of MP-PSNR (RMP-PSNR) were also presented in [31] to compute the quality score. However, since these methods only analyze global features, the distortions in the disoccluded regions cannot be well characterized.

Recently, some works have been conducted to combine the local and global features to produce better evaluation performance. Li *et al*. [32] proposed a LOcal Geometric distortion in disoccluded regions and global Sharpness (LOGS) index [32], which first measured local geometric distortions in disoccluded regions, and employed a reblurring-based strategy to quantify the global sharpness. The overall quality score was calculated by pooling the scores of both local disoccluded regions and global sharpness. Yue *et al*. [33] designed a similar quality metric, which measured geometric distortions by calculating the similarity between center pixels and its adjacent pixels in disoccluded regions, and the distance between the distorted image and its down-sampled version was measured as global sharpness. Although the two methods have achieved good performance, they still have room for improvement in prediction accuracy.

B. ANALYSIS of DIBR-RELATED DISTORTIONS

The artifacts of a synthesized view mainly locate in disoccluded regions which are not visible in the reference viewpoint, but are visible in the synthesized viewpoint. Different types of artifacts related to the DIBR-synthesized views have been defined [8]. The typical artifacts around disoccluded regions can be listed as follows.

1) *Object shifting*: A region may be slightly translated or resized, e.g., object regions may be resized in the

synthesized image if the depth map is smoothed by Gaussian filtering preprocessing or encoding. This type of artifact can be observed in Fig. 1(b), where the shape of the leaf is larger than that of the reference image. The shape of the vase in the image is also shifted.

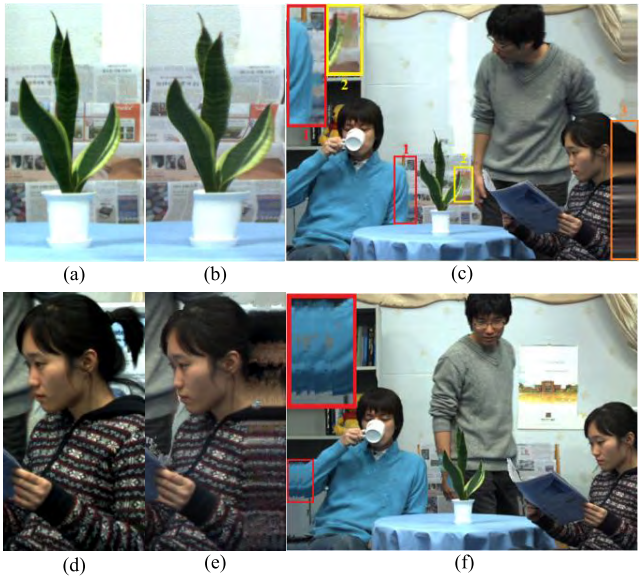


FIGURE 1. Illustrations of the synthesized distortions: (a) Reference image of (b); (b) Shifting/resizing artifacts; (c) Blurry artifacts; (d) Reference image of (e); (e) Incorrect rendering of textured areas; (f) Blocky artifacts.

2) *Blurry regions*: The distortion is induced by the inpainting method to fill the disoccluded regions, which is more noticeable in the boundary of background and foreground. Blurry regions marked by red, yellow and orange rectangles around disoccluded regions can be seen in Fig. 1(c).

3) *Incorrect rendering*: Inpainting methods may fail to reconstruct complex textured areas. A hole filling approach based on patch-wise texture synthesis is employed in [34] to solve this problem, but it may lead to new artifacts in the synthesized patches, which is obviously observed in Fig. 1(e).

4) *Blocking artifacts*: In the process of reconstructing complex textured areas, the use of rectangular patches may also lead to blocky artifacts in disoccluded regions [34], as shown in Fig. 1(f).

5) *Tiny distortions*: Some tiny geometric distortions and illumination errors are perceptually invisible in the synthesized images, but pixel-wise metrics may penalize these distortions.

In addition to the above artifacts caused by imperfect rendering, traditional distortions, such as Gaussian blur, white noise, compression distortion, or transmission error, may be also included in the texture or depth image, affecting the synthesized image, which can be observed in Fig. 2. Therefore, a good objective quality metric for synthesized images should also be robust to these distortions.

FIGURE 2. An example of synthesized distortions: (a) Reference image; (b) Synthesized image.

C. MOTIVATION OF THIS WORK

Based on the above analysis, although local and global quality evaluation issues for 3D synthesized images have been studied in [32] and [33], they still have the following limitations:

1) Due to imperfect DIBR techniques, the DIBR-related distortions are not restricted to disoccluded regions, but also appear near or around these regions (defined as critical areas in our work). Thus, directly evaluating the quality of the disoccluded regions cannot accurately reflect the degree of local distortions [33].

2) For a synthesized image, HVS perceives blur distortions mostly in complex texture and edge regions, while viewers are more intolerable to color artifacts appearing in homogeneous regions. Thus, the distortion in the disoccluded regions cannot be well characterized by only computing the mean value of the difference map [32].

To overcome these limitations, we attempt to enlarge disoccluded regions to obtain the critical areas, and use DCT decomposition to classify the blocks of critical areas into ED, TE and SM blocks. Properties of these blocks are used to evaluate them independently.

III. PROPOSED METHOD

The framework of the proposed method is illustrated in Fig. 3. The local distortions of a synthesized image are evaluated by two components: strength and size of critical areas. Specifically, the strength of critical areas is computed by measuring texture similarity and color contrast that use the size of the critical areas as weight. Meanwhile, gradient magnitude deviation (GMD) between synthesized and reference images is computed to quantity global sharpness. The final quality is a combination of local and global measurements. Some inputs and outputs used in our method are defined in Table 1.

TABLE 1. Definition of each image.

Definition	Symbol
Reference image	S_r
Synthesized image	S_s
Unfilled image	S_u
Warped reference image	S_{wr}
Warped unfilled image	S_{wu}

A. CRITICAL AREA DETECTION

The disoccluded regions in the synthesized image can be extracted from the unfilled image (i.e., with holes).

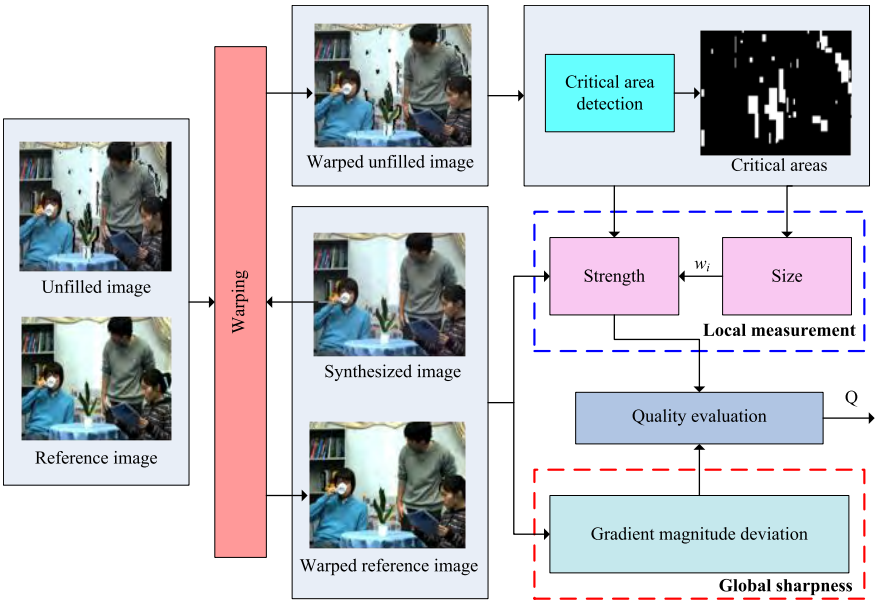


FIGURE 3. Overview of the proposed scheme.

However, using the DIBR algorithm [4], the holes of a synthesized image are removed, and then the image is stretched to its original size. In this case, it is difficult to accurately detect the disoccluded regions from the synthesized image, because geometric displacement between the synthesized and the unfilled images is existed. To solve this issue, we adopt SIFT-flow [35] to establish dense correspondences between two images to overcome the geometric displacement. The SIFT-flow map between S_r and S_s is obtained by minimizing the following objective function:

$$\begin{aligned} E(\mathbf{w}) = & \sum_{\mathbf{p}} \min\left(\left\|S_s(\mathbf{p}) - S_r(\mathbf{p}+\mathbf{w}(\mathbf{p}))\right\|, t\right) \\ & + \sum_{\mathbf{p}} \eta\left(|\mu(\mathbf{p})| + |\nu(\mathbf{p})|\right) \\ & + \sum_{\mathbf{p},\mathbf{q}\in\varepsilon} \{\min\left(\alpha\,|\mu(\mathbf{p}) - \mu(\mathbf{q})|\right) \\ & + \min\left(\alpha\,|\nu(\mathbf{p}) - \nu(\mathbf{q})|, d\right)\} \end{aligned} \quad (1)$$

where $\mathbf{w}(\mathbf{p})$ denotes the SIFT flow vector of pixel $\mathbf{p}$, parameters t and d denote the thresholds to limit the amount of maximum error, and η and α are the weights, $\mu(\mathbf{p})$ and $\nu(\mathbf{p})$ represent the horizontal and vertical components of the flow vector$\mathbf{w}(\mathbf{p})$, and $\mathbf{q}$ denotes the coordinate of the neighbor set of $\mathbf{p}$. Using the estimated SIFT flow vectors, the warped reference image and the warped unfilled image can be obtained, i.e., $S_{wr}(\mathbf{p}) = S_r(\mathbf{p}+\mathbf{w}(\mathbf{p}))$, $S_{wu}(\mathbf{p}) = S_u(\mathbf{p}+\mathbf{w}(\mathbf{p}))$. Fig. 4 (c) illustrates the warping result.

The disoccluded regions can be extracted by detecting the pixels without filling in the warped unfilled image. The obtained regions contain both noticeable distortions and weak distortions, as shown in Fig. 5(a), which are respectively marked by red and yellow rectangles. However, some weak distortions are too faint to draw human's attention, called isolated small "noise", which can be removed if the sizes are less than p pixels. Here, p is empirically set as 20. Those with a small area should be merged with the adjacent large one to share the similar area. Moreover, the distortions in synthesized views are not restricted to the disoccluded regions, but also appear around these regions. Thus, a rectangle-based morphological dilation operation is conducted on the disoccluded regions by applying the minimum rectangles to cover the disoccluded regions to obtain the final critical areas, as shown in Fig. 5(c).

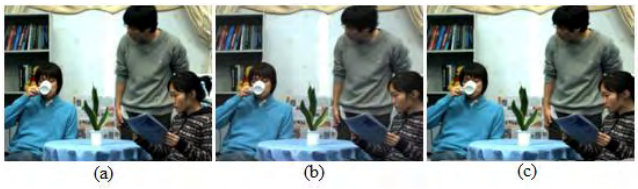


FIGURE 4. Example of warping result: (a) Reference image; (b) Synthesized image; (c) Warped reference image.

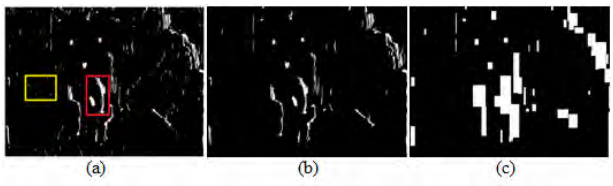


FIGURE 5. Example of critical area detection: (a) Detected regions from synthesized image; (b) Disoccluded regions after removing some isolated small "noise"; (c) Critical areas after morphological operation.

B. DCT-BLOCK PARTITION

DCT coefficients can be classified into direct current (DC) and alternating current (AC), as shown in Fig. 6. The AC energy of DCT coefficients, including low frequency (LF), medium frequency (MF) and high frequency (HF), is an effective measure of the local texture activity [36], [37]. In this paper, we attempt to employ DCT decomposition to classify the blocks in local critical areas.

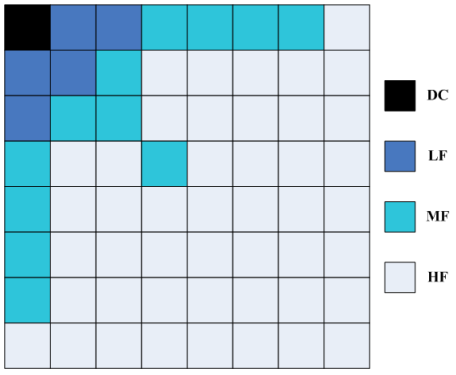


FIGURE 6. DCT block classification.

Specifically, the critical areas of reference and synthesized images are first divided into 8×8-pixel blocks, each of which is transformed into 64 coefficients. Then, each DCT block is assigned a class belonging to ED, TE and SM. Let L, M, and H represent the sums of the absolute DCT coefficient values in the LF, MF and HF bands, respectively, the texture energy of a block is approximated by:

$$E = M + H \tag{2}$$

Since the LF and MF bands reflect edge information, the values of L/M and $(L+M)/H$ can be used to determine if there is an edge [36], [37]. Then, C_1 and C_2 are defined as:

$$C_1 = \left(\bar{L} + \bar{M}\right)/\bar{H}, \quad C_2 = \bar{L}/\bar{M} \tag{3}$$

where $\bar{L}$, $\bar{M}$ and $\bar{H}$ denote the average values of L, M, and H in a block, respectively. The classification rules for blocks are further described as follows.

Case A: Since the texture energy of a smooth block is low, the block is served as a SM if $E \leq \xi_1$.

Compared with the block with low spatial activities, since the thresholds for detecting an edge for a block with high spatial activities needs to be lowered, two sets of ED thresholds are used. Specifically, a block is assigned to ED class if $C_1 \geq v$ or $\max\{C_1, C_2\} \geq \delta$& $\min\{C_1, C_2\} \geq \varepsilon$, where v, δ, and ε are three parameters, $\delta > \varepsilon$, denoted by *Rule*.

Case B: If a block doesn't meet *Rule* and $\xi_1 < E \leq \xi_2$, it is assigned to SM class. Otherwise, it is assigned to ED class.

Case C: If a block doesn't meet *Rule* and $\xi_2 < E \leq \xi_3$, it is assigned to TE class. Otherwise, it is assigned to ED class.

Case D: E > ξ_3. If a block satisfies *Rule* at $\delta = k\delta_1$ and $\varepsilon = k\varepsilon_1$, it is assigned to ED class. Otherwise, it is assigned to TE class.

TABLE 2. Four cases in block classification.

	E	δ	ε	*Rule*	*Class*
Case A	$[0, \xi_1]$	——	——	——	SM
Case B	$(\xi_1, \xi_2]$	δ_1	ε_1	Yes	ED
				No	SM
Case C	$(\xi_2, \xi_3]$	δ_1	ε_1	Yes	ED
				No	TE
Case D	$(\xi_3, \infty]$	$k\delta_1$	$k\varepsilon_1$	Yes	ED
				No	TE

The above four cases are summarized in Table 2. To verify the effectiveness of the classification, we test the classification rule on four different images, including the original image, three synthesized images generated using the A6 [38], A4 [39] and A7 [9] methods respectively. The main distortions of the three synthesized images are incorrect rendering of textured areas, blur and information loss, as shown in Fig. 7. Refer to [36] and [37], the parameters are set as follows: $\xi_1 = 125$, $\xi_2 = 290$, $\xi_3 = 900$, $\delta_1 = 7$, $\varepsilon_1 = 5$, $k = 0.1$, $v = 16$. The number of ED, TE and SM are computed for each image, denoted as N_{ED}, N_{TE} and N_{SM}, respectively.

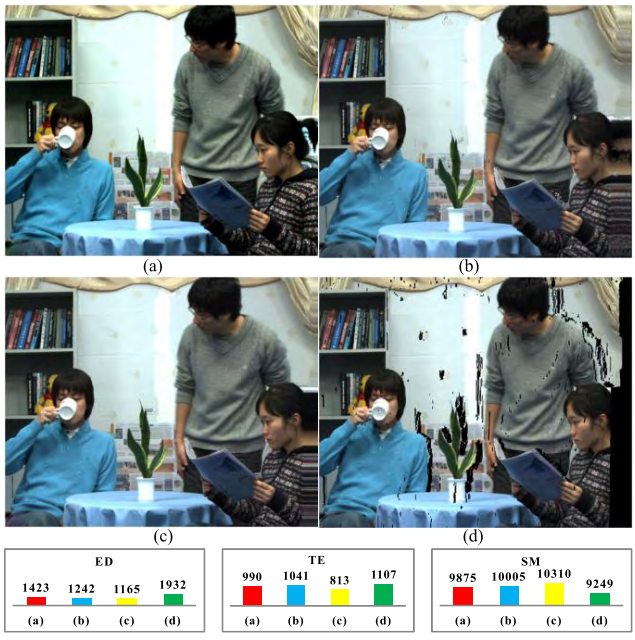


FIGURE 7. Example of block statistics: (a) Original image; (b), (c) and (d) are synthesized images generated using the A6 [38], A4 [39] and A7 [9] method, respectively. For each image, its number of ED, TE and SM are listed.

Compared with the original image, N_{ED} of the synthesized image generated using the A6 method is smaller, while N_{TE} and N_{SM} are bigger. This phenomenon can also be seen from the results of the incorrect texture synthesis, as shown in Fig. 7(b). The increasement of N_{SM} and decreasement of N_{ED} and N_{TE} are in accordance with the case that the main artifact in Fig. 7(c) is blur. Similarly, the unfilled image in Fig. 7(d) result in dramatic increasement of N_{ED} and N_{TE}

and corresponding decreasement of N_{SM}. Since the number of 3 classes varies significantly with different distortions, we may draw the conclusion that the fluctuation of ED and TE can be utilized to measure the degree of distortions.

C. LOCAL FEATURE SIMILARITY MEASUREMENT

As discussed, this paper uses DCT-based texture classification method to extract local features, including block-wise texture similarity and color contrast similarity in critical areas. The estimation procedure is shown in Fig. 8. The images are first converted into the perceptually uniform CIELab color space [40], where L is the luminance channel, a and b are two color channels. DCT decomposition is conducted in the luminance channel. Then, the similarity comparison and color contrast are computed based on the classification.

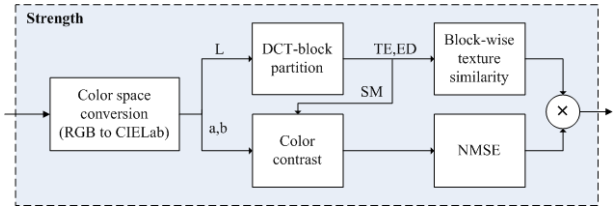


FIGURE 8. Estimation procedure of the Strength block in Fig. 3.

1) BLOCK-WISE TEXTURE SIMILARITY (BS)

BS of complex texture and edge regions reflects the degree of blur distortions in the synthesized image. To measure the variation of ED and TE between synthesized and reference images, the texture similarity of the i-th critical area is calculated as:

$$\widehat{BS}^k(i) = \frac{Num_k^s(i)}{Num_k^r(i)}, \quad k \in \{\text{TE}, \text{ED}\} \tag{4}$$

where $i = 1, \ldots, M$, M is the total number of critical areas in an image, $Num_k^s(i)$ and $Num_k^r(i)$ denote the number of ED or TE in the synthesized and reference image, respectively. As shown in Fig.7, the increasement of N_{ED} or N_{TE} also denotes the degradation of the synthesized image. Thus, if $\widehat{BS}^k(i)$ is larger than 1, the texture similarity can be revised as:

$$BS^k(i) = \begin{cases} \left(\widehat{BS}^k(i)\right)^{-1}, & \text{if} \widehat{BS}^k(i) > 1 \\ \widehat{BS}^k(i), & \text{otherwise} \end{cases} \tag{5}$$

The texture similarity of critical areas is calculated as:

$$BS = \frac{1}{M} \frac{\sum_{i=1}^{M} w_i \left(BS^{\text{TE}}(i) \cdot BS^{\text{ED}}(i)\right)}{\sum_{i=1}^{M} w_i} \tag{6}$$

where $w_i \in (0, 1]$ is the weight of a critical area determining based on the size of the area which can be obtained by calculating the number of pixels in the area. A low BS value indicates low image quality with heavy distortions in critical areas.

2) COLOR CONTRAST SIMILARITY (CC)

Besides the blur distortion, color distortion is another important factor for image synthesis, which may affect human's perception of the scene. Different with the case in complex texture and edge regions, smooth regions are sensitive to color distortions [41]. Based on this assumption, according to the location and size of SM in the luminance channel, the corresponding regions are detected in the color channel, written as:

$$\widehat{\mathbf{I}}_k = \mathbf{I}_k \otimes \mathbf{B}^{\text{SM}}, \quad k \in \{a, b\} \tag{7}$$

where $\mathbf{I}_{\text{k}}$ denotes the color channel, and $\mathbf{B}^{\text{SM}}$ represents the location and size of SM in the luminance channel. The normalized color contrast of the i-th SM is obtained by:

$$C(i) = \sum_{k \in \{\text{a,b}\}} \frac{1}{2c_k^M} \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} \left(\frac{c_k(i,j) - \bar{c}_k}{\bar{c}_k}\right)} \tag{8}$$

where N is the block size. j is the location index of the block. c_k^M denotes the maximum color contrast of a or b for normalization. $\bar{c}_k$ is the means of the block. We adopt normalized mean squared error (NMSE) for the distortion measure instead of a similarity measure such as:

$$\text{NMSE}^{\text{CC}}(C_s, C_r|\theta_1) = \frac{1}{T} \sum_{i=1}^{T} \frac{(C_s(i) - C_r(i))^2}{C_s^2(i) + C_r^2(i) + \theta_1} \tag{9}$$

where C_s and C_r denote the color contrast of the synthesized and reference images, respectively, T is the number of SM and θ_1 is a parameter to avoid unstable results when the denominator is close to zero. Here, $\theta_1 = 0.000001$. The CC in critical areas can be calculated as:

$$CC = \frac{1}{\text{NMSE}^{\text{CC}} + \theta_2} \tag{10}$$

where θ_2 is a model parameter to avoid unstable results when the denominator is close to zero. Here, $\theta_2 = 0.001$. A low CC value indicates low image quality with heavy distortions in critical areas. Then, the quality score of critical areas can be represented as the product of BS and CC, defined as:

$$Q_{local} = BS \cdot CC \tag{11}$$

As an example, we illustrate the local quality score for different synthesized images in Fig. 9, where a higher DMOS value means the lower image quality. The quality scores of the synthesized images generated from a same reference image are in accordance with the subjective measurements. However, for the two synthesized images with the same DMOSs generated from different reference images (e.g., 4.129 vs 4.178), the predicted local quality scores are quite inconsistent. Thus, another global quality measure is needed to compensate the inconsistency.

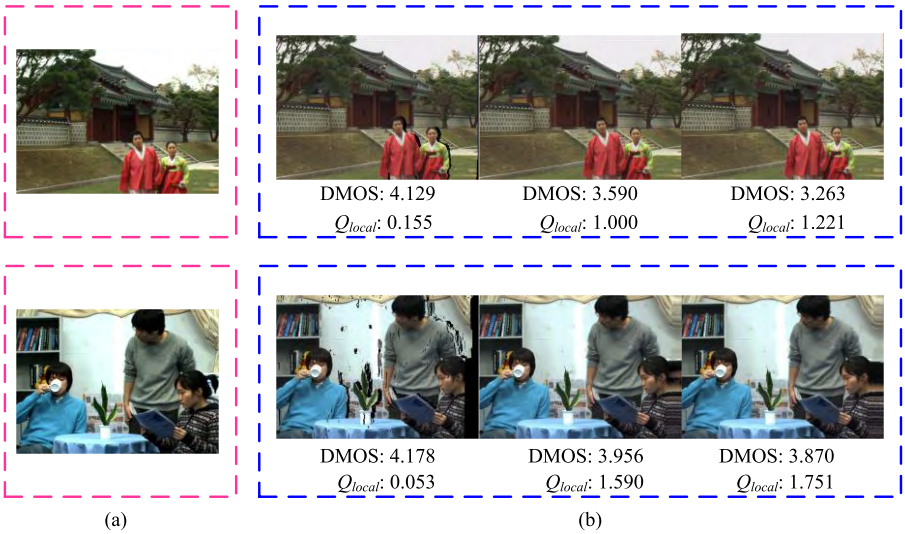


FIGURE 9. Example of local quality scores vs. DMOS for different DIBR algorithms: (a) Reference images; (b) Synthesized images from left to right generated using the A7 [9], A6 [38], and A5 [42] methods, respectively.

D. GLOBAL SHARPNESS ESTIMATION

Although the above texture similarity and color contrast measurement can effectively capture the local synthesized errors, it ignores the global sharpness variation, which can also be perceived by viewers. Thus, we add global sharpness measurement to the proposed quality metric. The image gradient is a popular feature in image quality assessment [22], [43], [44]. In the paper, GMD between synthesized and reference images is employed to quantify the global sharpness. The standard deviation of gradient magnitude is computed as:

$$\delta_t = \frac{1}{L}\sum_{(x,y)\in\Omega}\left[m^t(x,y) - \bar{m}^t\right]^2$$
$$m^t(x,y) = \sqrt{\left(\nabla f_x^t\right)^2 + \left(\nabla f_y^t\right)^2}, \quad t \in (r,s) \qquad (12)$$

where $m^t(x,y)$ denote the gradient magnitude of warped-reference or synthesized image at location (x,y), which can be expressed by the root mean square of horizontal gradient f_x and vertical gradient f_y. The overall sharpness score of a synthesized image is defined as the distance of standard deviation between the synthesized and warped reference images:

$$Q_{global} = \| \delta_s - \delta_r \| \qquad (13)$$

E. QUALITY EVALUATION

Local and global features are complementary as they quantify different aspects of quality degradation. With the quality score Q_{local} of local critical areas and global score Q_{global}, it is desirable to combine them to produce an overall score. The following simple pooling strategy is employed to define the final quality score of a synthesized image:

$$Q = \beta \cdot Q_{local} + \gamma \cdot Q_{global} \qquad (14)$$

where the parameters β and γ are used to balance the relative contributions of local critical areas and global GMD. In the experiment, we empirically set $\beta = 8$, $\gamma = 15$.

IV. PERFORMANCE EVALUATION

A. BENCHMARK DATABASES

In the experiment, three benchmark DIBR databases, including IRCCyN/IVC [9], IETR [16] and MCL-3D [15], are used for performance evaluation. All these databases contain reference views, the corresponding synthesized views obtained by different DIBR algorithms and the associated subjective rating scores. The basic introduction of these datasets is summarized in Table 3.

TABLE 3. Basic information of the used benchmark databases.

Datasets	IRCCyN/IVC	IETR	MCL-3D
Ref. Image. No.	12	10	9
DIBR algorithm No.	7	7	4
Other distortions	NO	NO	YES
Syn. Image No.	84	140	693
Display	2D	2D	Stereo.
Observer No.	43	42	270

The IRCCyN/IVC database is the first released benchmark database for synthesized image quality assessment, which contains 12 reference images and 84 synthesized images generated using seven DIBR approaches. The Mean Opinion Score (MOS) is provided for each synthesized image in the dataset.

IETR database contains 10 reference images and 140 synthesized images generated using seven state-of-the-art DIBR view synthesis algorithms, including VSRS [45], Zhu and Li's [46], Criminisi's [47], Luo's [48], HHF [49], LDI [50] and Ahn and Kim's [51]. Both interview interpolation and single view based synthesis methods are considered

TABLE 4. Performance comparison on the IRCCyN/IVC database. The best results are in bold.

Metric	Type	PLCC	SRCC	KRCC	RMSE
PSNR	2D	0.5098	0.4628	0.3321	0.5728
SSIM	2D	0.5644	0.5245	0.3650	0.5496
FSIM	2D	0.5828	0.4161	0.2738	0.5410
GSM	2D	0.5246	0.4215	0.2750	0.5668
LTG	2D	0.5311	0.4139	0.2900	0.5642
ADD-SSIM	2D	0.6470	0.5611	0.4141	0.5077
Bosc [9]	DIBR	0.5843	0.4905	0.3414	0.5403
VSQA [29]	DIBR	0.5742	0.5233	0.3673	0.5451
MW-PSNR	DIBR	0.5622	0.5757	0.4378	0.5506
RMW-PSNR	DIBR	0.5744	0.6245	0.4960	0.5450
MP-PSNR	DIBR	0.6174	0.6227	0.4833	0.5238
RMP-PSNR	DIBR	0.6772	0.6634	0.5382	0.4899
3DSwIM [12]	DIBR	0.6584	0.6156	0.4496	0.5011
SIQE [10]	DIBR	0.5284	0.4492	0.3269	0.5653
DSQM [11]	DIBR	0.7650	0.7067	0.5382	0.4288
LOGS [32]	DIBR	0.8256	0.7812	0.6071	0.3601
Proposed	DIBR	**0.8512**	**0.8346**	**0.6431**	**0.3146**

in the database. The DMOS is provided for each synthesized image in the dataset.

MCL-3D database [15] contains 693 stereoscopic image pairs from 9 image-plus-depth sources. Six image distortions, including Gaussian blur, additive white noise, down-sampled blur, JPEG and JPEG-2000 (JP2K) compression and transmission error, are added on the texture/depth image before rendering. The distortion caused by four different rendering algorithms was also considered. The MOS is provided for each synthesized image in the dataset.

B. EXPERIMENTAL PROTOCOLS

Considering the non-linearity of subjective scores, we first compute the objective prediction scores, and then use the nonlinear regression to map the scores to subjective ratings based on a five-parameter logistic function:

$$s_p = \mu_1 \left[\frac{1}{2} - \frac{1}{1 + \exp\left(\mu_2 \left(s_o - \mu_3\right)\right)} \right] + \mu_4 s_o + \mu_5 \quad (15)$$

where s_o is the objective score as input, and μ_1, μ_2, μ_3, μ_4 and μ_5 are the parameters to be fitted. To benchmark the performance of image quality assessment metrics, four commonly-used performance indicators are used, namely Pearson Linear Correlation Coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC), Kendall's Rank Correlation Coefficient (KRCC) and RMSE. Specifically, PLCC and RMSE are calculated between the subjective and objective scores to evaluate the prediction accuracy, while SRCC and KRCC are employed to measure prediction monotonicity. For a perfect objective model, we have PLCC $=$ SROCC $=$ KROCC $= 1$ and RMSE $= 0$.

C. OVERALL PERFORMANCE

In this section, we evaluate the proposed method on the IRCCyN/IVC DIBR database and compare it with 16 state-of-the-art approaches, including: 1) the six traditional 2D quality metrics: PSNR, SSIM [20], FSIM [21], GSM [22], LTG [23] and ADD-SSIM [24]; 2) the ten existing DIBR quality metrics: Bosc's [9], VSQA [29], MW-PSNR [30], RMW-PSNR [30], MP-PSNR [31], RMP-PSNR [31], 3DSwIM [12], SIQE [10], DSQM [11] and LOGS [32]. From Table 4, we have derived the following conclusions: 1) traditional 2D image quality metrics are not effective in evaluating the quality of synthesized images. The best performance of 2D metric is obtained by ADD-SSIM, but the PLCC and SRCC are only 0.647 and 0.5611, respectively; 2) most existing DIBR quality metrics perform better than traditional 2D quality metrics, e.g., RMP-PSNR, 3DSwIM, DSQM and LOGS. Among these metrics, LOGS metric produces the best results with PLCC $=$ 0.8256 and SRCC $=$ 0.7812, respectively. In contrast, our method delivers the best performance, significantly better than the LOGS metric.

In Fig.10, we also provide the scatter plots of the predicted quality scores against the DMOS values for some representative DIBR objective metrics, e.g., Bosc *et al.*'s [9], VSQA [29], MW-PSNR [30], MP-PSNR [31], 3DSwIM [12], SIQE [10], DSQM [11] and LOGS [32]) on the IRCCyN/IVC DIBR database. It can be observed that the proposed model can provide relatively compact distribution along the diagonal of the scatter plots.

D. EVALUATION ON IETR AND MCL-3D DATABASES

We also present our results on the other two challenging IETR [16] and MCL-3D [15] databases. Due to the lack of hole images in the IETR database, inspired by the works in [9], critical areas are extracted using a self-adaption threshold. The results in terms of PLCC, SRCC, and RMSE on the two databases are illustrated in Table 5 and Table 6, respectively. From the results, we make the following observations: 1) For the IETR database, the proposed metric outperforms all competing 2D metrics and DIBR metrics except for LOGS metric. The reason may be that local measurements of the proposed metric are highly dependent on the

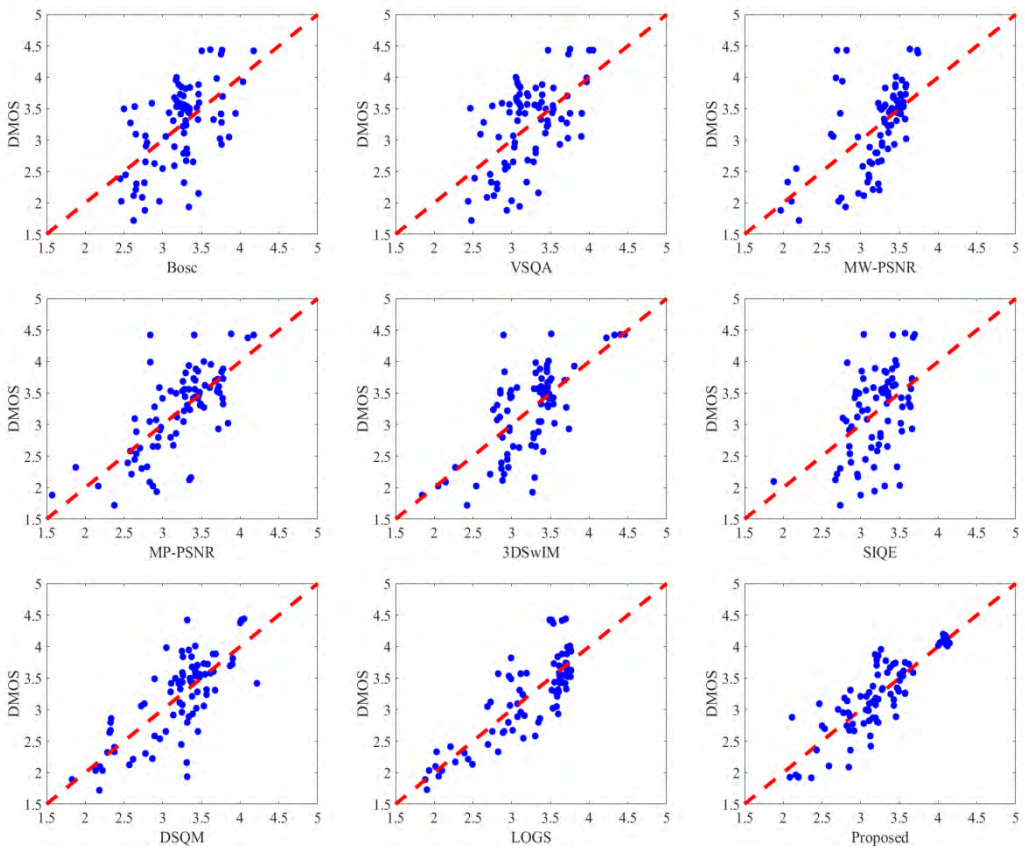


FIGURE 10. Scatter plots of quality indices designed for view synthesis on IRCCyN/IVC database.

TABLE 5. Performance results on IETR database. The best results are in bold.

Metric	Type	PLCC	SRCC	RMSE
PSNR	2D	0.6012	0.5356	0.1985
SSIM [20]	2D	0.4016	0.2395	0.2275
MS-SSIM [25]	2D	0.6162	0.5355	0.1957
IW-PSNR [26]	2D	0.5827	0.4973	0.2019
UQI [27]	2D	0.4346	0.4113	0.2237
PSNR-HVS [28]	2D	0.5982	0.5195	0.1991
VSQA [29]	DIBR	0.5576	0.4719	0.2062
MW-PSNR [30]	DIBR	0.5301	0.4845	0.2106
RMW-PSNR [30]	DIBR	0.5403	0.4946	0.2090
MP-PSNR [31]	DIBR	0.5753	0.5507	0.2032
RMP-PSNR [31]	DIBR	0.6061	0.5873	0.1976
LOGS [32]	DIBR	**0.6687**	**0.6683**	**0.1845**
Proposed	DIBR	**0.6118**	**0.6136**	**0.1961**

accuracy of critical area detection, whereas the database does not have unfilled images; 2) For the MCL-3D database, the performance of the proposed metric is only inferior to PSNR-induced metrics. This is attributable to the fact that PSNR-induced metrics are effective to measure the global image artifacts in the MCL-3D database. Although LOGS and PSNR-induced metrics perform better than the proposed method in one of the three DIBR databases, both of them deliver inferior performances in the other two databases. Therefore, our method has comparatively consistent performances over different databases.

E. IMPACT OF EACH QUALITY COMPONENT

In our model, we used three types of features, i.e., local BS and CC features, and global GMD features, to reveal local and global feature quality cues. In order to investigate the importance of each feature, we separately test the performance using independent feature on the IRCCyN/IVC database. As shown in Table 7, combining three features achieves the best performance. Moreover, the scheme only using BS feature performs better than other scheme only using CC or GMD feature, because most synthesized-related errors are occurred in critical areas. Although the performances of

TABLE 6. Performance comparison on the MCL-3D database. The best results are in bold.

Metric	Type	PLCC	SRCC	RMSE
Bosc	DIBR	0.4453	0.4079	2.3295
VSQA	DIBR	0.5078	0.5120	2.9175
MW-PSNR	DIBR	**0.8012**	**0.8099**	**1.5568**
RMW-PSNR	DIBR	**0.8239**	**0.8308**	**1.4743**
MP-PSNR	DIBR	**0.8169**	**0.8231**	**1.5007**
RMP-PSNR	DIBR	**0.8173**	**0.8250**	**1.4993**
3DSwIM	DIBR	0.6519	0.5683	1.9729
SIQE	DIBR	0.6734	0.6976	1.9233
DSQM	DIBR	0.6995	0.6980	1.8593
LOGS	DIBR	0.7612	0.7577	1.6874
Proposed	DIBR	**0.7910**	**0.7929**	**1.5917**

TABLE 7. Performance of different components.

	PLCC	SRCC	KRCC	RMSE
BS	0.8077	0.7392	0.5578	0.3535
CC	0.2152	0.0098	0.0005	0.5855
GMD	0.4930	0.4257	0.3167	0.5216
ALL	0.8512	0.8346	0.6431	0.3146

CC are very low, if CC component is omitted, the SRCC is reduced to 0.7531, which is far less than the performance of the proposed metric.

F. FURTHER DISCUSSION

Although our method demonstrates higher performance compared with the existing 2D quality metrics and DIBR metrics, the following issues still deserve to be considered in the future work:

1) The preprocessing step with warping methods has positive and negative effects. It can handle the displacement between synthesized and reference images, but the warping method may compensate some geometric distortions in the critical areas. Therefore, suppressing the negative effect of warping methods is expected to further improve the performance of the proposed method.

2) Critical area detection is a significant step which affects the accuracy of local quality evaluation. However, the associated unfilled images are not available in the practical application. Thus, an effective method which only uses synthesized and reference images is demanded for detecting critical areas.

V. CONCLUSION

This paper has presented a quality model for DIBR-synthesized images based on global and local feature analysis. The motivation of this work is to explore some known facts of the human visual system, which first rapidly and unconsciously produces a global perception and then gradually focuses on specific local areas for the perception of image quality, to build a model that is useful for quality evaluation of DIBR-synthesized images. For this purpose, the proposed model is dedicated to characterize the quality of a synthesized image from both local and global perspectives. For local feature analysis, block-wise texture similarity and color contrast in critical areas are calculated to characterize local visual information of a synthesized image. For global feature analysis, the standard deviation of the gradient magnitude map is employed to analyze the global visual information. Finally, local and global features are combined to produce the quality score. Although the proposed method has shown better performance than traditional 2D quality metrics and existing DIBR metrics, it still belongs to a full-reference model requiring a reference image for comparison. Therefore, future work may focus on no-reference metrics to provide a more practical solution for assessing the quality of synthesized images.

REFERENCES

[1] A. Smolic *et al.*, "3D video and free viewpoint video–technologies, applications and MPEG standards," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2006, pp. 2161–2164.

[2] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 67–76, Jan. 2011.

[3] A. Smolic *et al.*, *Multi-View Video Plus Depth (MVD) Format for Advanced 3D Video Systems*, document JTC1/SC29/WG11 and ITU-T SG16 Q 6, San Jose, CA, USA, Apr. 2007.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, May 2004.

[5] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3-D video for view rendering," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 157–167, Feb. 2012.

[6] F. Shao, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Depth map coding for view synthesis based on distortion analyses," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 106–117, Mar. 2014.

[7] F. Shao, Q. Yuan, W. Lin, and G. Jiang, "No-reference view synthesis quality prediction for 3-D videos based on color–depth interactions," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 659–674, Mar. 2018.

[8] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, "Visual quality assessment of synthesized views in the context of 3D-TV," in *3D-TV System With Depth-Image-Based Rendering*. New York, NY, USA: Springer, 2013, pp. 439–473.

[9] E. Bosc *et al.*, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1332–1343, Nov. 2011.

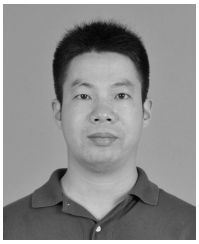
[10] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Objective quality metric for 3D virtual views," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3720–3724.

[11] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Perceptual quality assessment of 3D synthesized images," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 505–510.

[12] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3D synthesized views," *Signal Process., Image Commun.*, vol. 30, pp. 78–88, Jan. 2015.

[13] S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1248–1252.
[14] K. Gu, J.-F. Qiao, P. Le Callet, Z. Xia, and W. Lin, "Using multiscale analysis for blind quality assessment of DIBR-synthesized images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 745–749.
[15] R. Song, H. Ko, and C. C. J. Kuo, "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source," *J. Inf. Sci. Eng.*, vol. 31, no. 5, pp. 1593–1611, 2015.
[16] S. Tian, L. Zhang, L. Morin, and O. Deforges, "A benchmark of DIBR synthesized view quality assessment metrics on a new database for immersive media applications," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2018.2875307.
[17] H. Shao, X. Cao, and G. Er, "Objective quality assessment of depth image based rendering in 3DTV system," in *Proc. IEEE True Vis.-Capture, Transmiss. Display 3D Video*, May 2009, pp. 1–4.
[18] Y. J. Jung, H. G. Kim, and Y. M. Ro, "Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1201–1214, Jul. 2016.
[19] K. Gu, V. Jakhetiya, J.-F. Qiao, X. Li, W. Lin, and D. Thalmann, "Model-based referenceless quality metric of 3D synthesized images using local image description," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 394–405, Jan. 2018.
[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[21] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[22] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
[23] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 506–510.
[24] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.
[25] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.
[26] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[27] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[28] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proc. 3rd Int. Workshop Video Process. Qual. Metrics*, vol. 4, 2007, pp. 1–4.
[29] P.-H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," *Proc. SPIE*, vol. 8288, Feb. 2012, p. 82881M.
[30] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *Proc. IEEE 7th Int. Workshop Qual. Multimedia Exper.*, May 2015, pp. 1–6.
[31] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. IEEE True Vis. Capture, Transmiss. Display 3D Video*, Jul. 2015, pp. 1–4.
[32] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914–926, Apr. 2018.
[33] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for DIBR-synthesized image quality evaluation," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2075–2088, Apr. 2019, doi: 10.1109/TIP.2018.2875913.
[34] P. Ndjiki-Nya *et al.*, "Depth image based rendering with advanced texture synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2010, pp. 424–429.
[35] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.
[36] X. H. Zhang, W. S. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Process.*, vol. 85, no. 4, pp. 795–808, 2005.
[37] H. H. Y. Tong and A. N. Venetsanopoulos, "A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 1998, pp. 428–432.
[38] M. Köppel *et al.*, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 1809–1812.
[39] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3D video systems," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 1–11, Dec. 2009.
[40] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.*, vol. 359, pp. 381–400, Feb. 1985.
[41] S.-H. Bae and M. Kim, "A novel image quality assessment with globally and locally consilient visual quality perception," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2392–2406, May 2016.
[42] P. Ndjiki-Nya *et al.*, "Depth image-based rendering with advanced texture synthesis for 3-D video," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 453–465, Jun. 2011.
[43] D.-O. Kim, H.-S. Han, and R.-H. Park, "Gradient information-based image quality metric," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 930–936, May 2010.
[44] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[45] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, *Reference Softwares for Depth Estimation and View Synthesis*, Standard ISO/IEC JTC1/SC29/WG11 MPEG 20081, 2008, p. M15377.
[46] C. Zhu and S. Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 82–93, Mar. 2016.
[47] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
[48] G. Luo, Y. Zhu, Z. Li, and L. Zhang, "A hole filling approach based on background reconstruction for view synthesis in 3D video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1781–1789.
[49] M. Solh and G. AlRegib, "Hierarchical hole-filling for depth-based view synthesis in FTV and 3D video," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 495–504, Sep. 2012.
[50] V. Jantet, C. Guillemot, and L. Morin, "Object-based layered depth images for improved virtual view synthesis in rate-constrained context," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 125–128.
[51] I. Ahn and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 614–626, Dec. 2013.

XUEJIN WANG received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, China, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree with Ningbo University, Ningbo, China. Her current research interests include image/video processing and quality assessment.



FENG SHAO (M'16) received the B.S. and Ph.D. degrees in electronic science and technology from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively. In 2012, he joined the School of Computer Engineering, Nanyang Technological University, Singapore, as a Visiting Fellow. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. He has published over 100 technical articles in refereed journals and proceedings in the areas of 3-D video coding, 3-D quality assessment, and image perception. He received the Excellent Young Scholar Award from the NSF of China, in 2016.

QIUPING JIANG (S'17) received the Ph.D. degree from Ningbo University, Ningbo, China, in 2018. From 2017 to 2018, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He is currently an Associate Professor with the School of Information Science and Engineering, Ningbo University. His research interests include image processing, visual perception modeling, and computer vision. He was a recipient of the JVCI 2017 Best Paper Award Honorable Mention, as the first author. He is a Reviewer for several prestigious journals and conferences, such as the IEEE T-NNLS, IEEE T-IP, IEEE T-CSVT, IEEE T-MM, IEEE T-SIPN, ICME, and ICIP.

RANDI FU received the M.S. degree from PLA Information Engineering University, China, in 2001. He is currently an Associate Professor with the Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include pattern recognition and remote sensing image processing.

YO-SUNG HO (SM'06–F'16) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, South Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. In 1983, he joined the Electronics and Telecommunications Research Institute (ETRI), Daejon, South Korea. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, NY, USA, where he was involved in the development of the advanced digital high-definition television system. In 1993, he rejoined ETRI as a Technical Staff Member and was involved in the development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, South Korea, where he is currently a Professor with the Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced video coding techniques, digital video and audio broadcasting, 3-D video processing, and content-based signal representation and processing. He is a Fellow of the IEEE.