# Combining Local and Global Measures for DIBR-Synthesized Image Quality Evaluation

Guanghui Yue, Chunping Hou, Ke Gu, Tianwei Zhou, and Guangtao Zhai, *Member, IEEE*

***Abstract*— Depth-image-based-rendering (DIBR) techniques are significant for 3D video applications, e.g., 3D television and free viewpoint video (FVV). Unfortunately, the DIBR-synthesized image suffers from various distortions, which induce an annoying viewing experience for the entire FVV. Proposing a quality evaluator for DIBR-synthesized images is fundamental for the design of perceptual friendly FVV systems. Since the associated reference image is usually not accessible, full-reference (FR) methods cannot be directly applied for quality evaluation of the synthesized image. In addition, most traditional no-reference (NR) methods fail to effectively measure the specifically DIBR-related distortions. In this paper, we propose a novel NR quality evaluation method accounting for two categories of DIBR-related distortions, i.e., geometric distortions and sharpness. First, the disoccluded regions, as one of the most obvious geometric distortions, are captured by analyzing local similarity. Then, another typical geometric distortion (i.e., stretching) is detected and measured by calculating the similarity between it and its equal-size adjacent region. Second, considering the property of scale invariance, the global sharpness is measured as the distance between the distorted image and its downsampled version. Finally, the perceptual quality is estimated by linearly pooling the scores of two geometric distortions and sharpness together. Experimental results verify the superiority of the proposed method over the prevailing FR and NR metrics. More specifically, it is superior to all competing methods except APT in terms of effectiveness, but greatly outmatches APT in terms of implementation time.**

***Index Terms*— Perceptual quality, image quality assessment (IQA), view synthesis, DIBR, geometric distortion, sharpness.**

## I. Introduction

THREE-DIMENSIONAL (3D) videos have received considerable attention and gained rapidly in recent decades due to their strong immersive perception [1]–[4]. As one important application of 3D videos, free viewpoint video (FVV) tends to technically mature and bids fair to the developmental direction of the next-generation video technologies [5]. Generally, the content in FVV is recorded by multiple cameras at different viewpoints and displayed by special technology. Viewers can view the 3D scene from multiple viewpoints when perceiving FVVs. With the increasing number of views, it inevitably aggravates the task of storage and transmission [6]. To solve this issue, Multi-view-Video-plus-Depth (MVD) technique has been designed for 3D representation and can be used as the data format of FVV [7]. Using MVD, limited textural views and their associated depth maps are required to be coded and transmitted. The remaining virtual views can be synthesized from decoded textural views and depth maps via the Depth-Image-Based-Rendering (DIBR) algorithm on the receiver side [8], [9]. Unfortunately, DIBR is not a flawless technology and produces distortions on the synthesized views. A low-quality view may induce an annoying perceptual experience in the entire FVV, but existing solutions are powerless in accurately estimating those distortions [10]. As a preliminary study of evaluating quality of FVVs, effectively and efficiently estimate the perceptual quality of DIBR-synthesized views has attracted increasing attention from scholars [10]–[15].

Over the past decades, extensive discussions have been made on the perceptual image quality assessment (IQA) [16]–[21]. Subjective evaluation, as the most reliable measurement, directly reflects intuitive subjective feelings. However, it fails in large-scale practical applications since it is time-consuming and expensive. Actually, subjective evaluation is usually treated as the ground truth of objective evaluation, which can be generally divided into three categories. To be specific, full-reference (FR) and reduced-reference (RR) methods respectively require entire and partial reference information to estimate the quality of the distorted image. While, no-reference (NR) methods do not require any reference information.

To date, dozens of objective methods have been designed for evaluating 2D images [19], [21]–[26]. However, as previously revealed, they are incapable of solving the quality assessment problem of DIBR-synthesized images [10], [12], [15]. Hence, specific methods for DIBR-synthesized images have been developed as the demand for the development of FVVs has increased. For example, Bosc *et al.* [12] conducted the pioneer work that investigated the reason why traditional IQA methods failed to evaluate DIBR-synthesized images and appealed for the attention to designing new IQA solutions.

Manuscript received January 10, 2018; revised July 6, 2018 and August 22, 2018; accepted September 30, 2018. Date of publication October 15, 2018; date of current version December 19, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61520106002, Grant 61731003, and Grant 61471262, and in part by the China Scholarship Council. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Khan M. Iftekharuddin. *(Corresponding author: Ke Gu).*

G. Yue, C. Hou, and T. Zhou are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: yueguanghui@tju.edu.cn; hcp@tju.edu.cn; tianwei@tju.edu.cn).

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

G. Zhai is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@gmail.com).

Digital Object Identifier 10.1109/TIP.2018.2875913

Conze *et al.* [13] focused on DIBR-induced artifact detection and provided an FR metric by extending an existing 2D IQA metric. Sandić-Stanković *et al.* [27] first decomposed an image into morphological wavelet subbands at multiple scales. Then, the mean squared error (MSE) value was calculated in each subband between the synthesized image and reference image. Finally, the quality of the synthesized image was obtained by pooling the MSE values across scales. Likewise, they also improved the performance by replacing the morphological wavelets with morphological pyramids [28]. Battisti *et al*. [14] designed an FR metric via comparison of statistical features of Haar wavelet subbands between the synthesized image and the reference image. To ensure shifting-resilience, a registration step was required before the feature comparison. Besides, considering that the observer was more sensitive to human-related regions, a skin detection step was conducted to weight the final quality score. Zhou *et al*. [29] proposed a novel FR metric by detecting disoccluded regions and the overall quality was estimated via a weighted pooling function. The performance was further improved by combining global sharpness [15]. Assuming that edge information could well reflect distortions, Zhou *et al*. [30] quantified the edge intensity and orientation between the synthesized image and the reference image individually. The image quality was estimated by pooling the two components together.

To promote the development of this research area, we propose a novel NR quality assessment algorithm for DIBR-synthesized images in this paper. Specifically, we mainly consider geometric distortions (i.e., the disoccluded region and stretching) and sharpness. More specifically, according to local characteristics of natural images, disoccluded regions are first detected by comparing the relationships between the local pixel and its surrounding pixels. Then, the stretching region is captured via threshold detection. To quantify the impacts of stretching on the final performance, we compute the similarity between the stretching region and its equal-size adjacent region. Next, the global sharpness is estimated by measuring the distance between the synthesized image and its downsampled version. Finally, the image quality is estimated via a linear pooling procedure of geometric distortions and sharpness. Compared to those reference-based methods, the proposed method does not require any reference information and may have potentials with broader range of applications. Experimental results on IRCCyN/IVC DIBR image database (developed by the *Institut de Recherche en Communications et Cyberntique de Nantes*) [12] show that the proposed method is consistent well with subjective feelings and simultaneously balances the effectiveness and implementation time. More concretely, it is superior to existing mainstream IQA methods and is slightly inferior to APT in terms of effectiveness. However, it costs less implementation time and only requires less than hundredth of the time required on APT.

The rest of this paper is organized as follows. Section II introduces the related works and motivation of this paper. In Section III, we describe the proposed method in detail. Section IV presents the experimental protocol, results on IRCCyN/IVC DIBR image database. Finally, the conclusion is drawn in Section V. For the reader's convenience, we present the used abbreviations in the Appendix.

## II. Related Works and Motivation

### A. NR Image Quality Assessment Methods

In the literature, a great variety of mature NR IQA methods have been designed for 2D images [17]. These methods are capable of evaluating the commonly occurring distortions, like blurriness/sharpness, blockiness and contrast change. Broadly speaking, they can be classified into two categories. One is the specific-purpose method, which is designed for evaluating the specific distortion type mentioned above. For example, Liu and Heynderickx [31] proposed an effective method for measuring blockiness by considering the structural information and visual masking effect. Wang *et al*. [32] extracted gradient statistical properties and formed a blind blurriness evaluator by utilizing an extreme learning machine. In [33], an NR blurriness assessment method was introduced via analysis of local texture. Gu *et al*. [34] considered multiple properties (e.g., naturalness, sharpness, brightness and colorfulness) and built a regression module to evaluate image contrast. Another category is the general-purpose method, which is designed to simultaneously evaluate multiple distortion types. Currently, most general-purpose methods are learning-based. To be specific, they first extracted quality-sensitive features (e.g., gradient, luminance, texture, entropy and regularity change of natural scene statistics). Then, the quality predictor was built by bridging the relationship between these features and the quality scores via machine learning tools [35]–[38]. Although these methods have obtained reliable performance in the quality assessment of 2D images, they fail or are unsuitable for DIBR-synthesized images. Two reasons explain this phenomenon. 1) The synthesized view suffers different distortions compared to these common distortions (more details are given in Section II-B), therefore, the aforementioned methods show their inability. 2) Most of the aforementioned methods rely on a sufficient dataset for training the prediction model. However, due to the limited amount of data, it is unsuitable for training a reliable model for the quality assessment of DIBR-synthesized images.

Currently, the quality assessment of DIBR-synthesized images is still in its infancy, and only a few attempts have been made to blindly estimate their quality. Gu *et al*. [10] conducted a pioneering work to blindly estimate the quality of DIBR-synthesized images. In that approach, an autoregression prediction model (named APT) was built to detect the geometric distortion, and then an optimized procedure was utilized by taking the saliency map into account. Assuming that a high-quality image and the synthesized image possessed different responses on morphological operations, Tian *et al*. [39] proposed a morphological based approach (named NIQSV) to detect the geometric distortion and consequently measure the image quality. Later, they extended NIQSV (to NIQSV+) to further improve the performance by integrating scores of NIQSV, stretching and black hole together [40]. These attempts are far from completely solving the quality assessment problem of DIBR-synthesized images. On the one

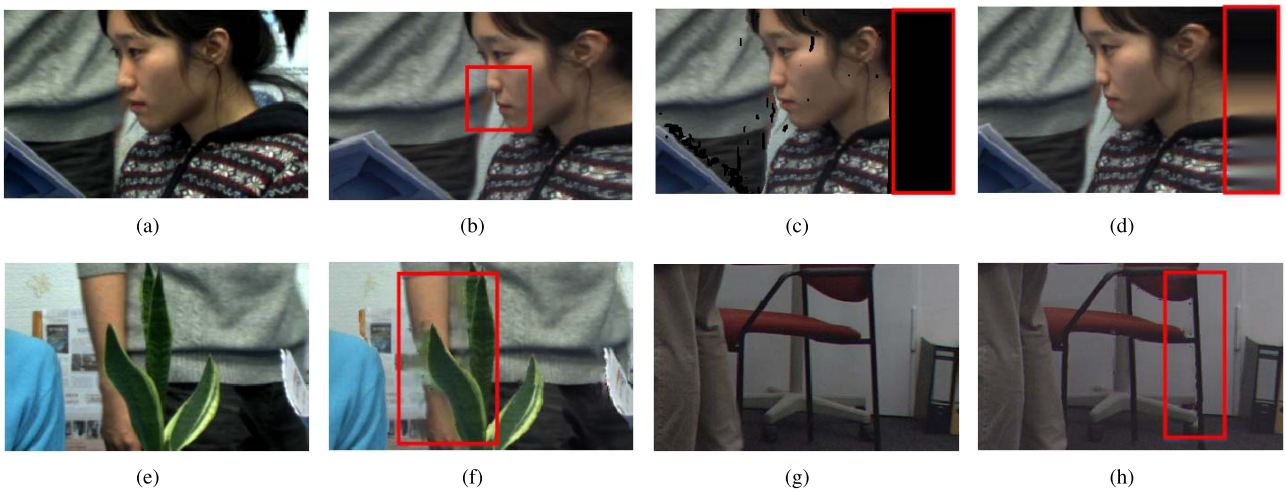



Fig. 1. Illustrations of the synthesized distortions: (a) reference view of (b)−(d), (b) object shifting, (c) disoccluded region, (d) stretching, (e) reference view of (f), (f) blurry region, (g) reference view of (h), and (h) crumbling. In each figure, there may be multiple distortions, we highlight only the representative distortion with a red rectangle.

hand, APT is computationally expensive although it has high performance. On the other hand, both NIQSV and NIQSV+ leave a lag on effectiveness behind their high efficiency. Therefore, large room remains open for researchers on either balancing the effectiveness and efficiency or improving each part individually/simultaneously.

### *B. Motivation*

Different from common ones, DIBR-related distortions are mainly generated during DIBR synthesis and exhibit diverse appearances. Specifically, warping, as the first stage of DIBR synthesis, aims to map the reference view to a 3D Euclidean space with the supervision of depth information. Then, the target view can be generated from the 3D Euclidean space via an inverse mapping operation [8]. Unfortunately, such an operation inevitably induces geometric displacements in the form of the disoccluded regions. In general, a rendering operation is subsequently required to restore these geometric displacements. However, a blurry artifact may be produced on the synthesized image when adopting imperfect rendering algorithms. Meanwhile, stretching occurs on the left or right side of the image since rendering algorithms may fail to work on this region. Moreover, the quality of the depth information also plays an important role in determining the final appearance of the synthesized image. For instance, a simple preprocessing operation (e.g., compression and filter) will obviously affect the synthesized contents. For convenience, we give an intuitive illustration in Fig. 1.

Fig. 1(a) is the local portion of one reference image in the IRCCyN/IVC DIBR image database, while Fig. 1(b) - 1(d) are its distorted versions. Specifically, Fig. 1(b) suffers from object shifting distortion (as marked by the red rectangle, the woman's nose is squeezed), which is mainly induced by error depth information. Fig. 1(c) shows a distorted image that contains disoccluded regions. Generally, the disoccluded region is one of the most obvious distortion types in the synthesized image and should be focused on during quality assessment model construction. Stretching, as shown in Fig. 1(d), is another typical distortion. Such distortion often occurs on the left or right side of the image, where rendering algorithms are powerless. Fig. 1(e) presents another reference image and its synthesized version suffered from blurry region distortion is shown in Fig. 1(f). Theoretically, a blurry region is generated during in-painting the disoccluded region using imperfect rendering algorithms. Fig. 1(g), containing the crumbing distortion, is the synthesized view of Fig. 1(h). Generally, crumbing may occur along the object edge when the depth map is under wavelet-based compression. Apart from these distortions, other artifacts, such as flickering and slight geometric distortion, may also be produced. It is clear that all these distortion types are quite different from common distortions, such as blockiness, noise, contrast change, etc. This is why traditional methods designed for 2D images are not proper for evaluating these synthesized distortions.

To effectively evaluate the quality of the synthesized image, one must focus on detecting the aforementioned distortions and find a suitable solution to quantify each distortion. Moreover, since the reference image of the DIBR-synthesized image is not actually accessible, a reference-free method is more favored and welcomed. In this paper, we propose an effective and efficient NR quality assessment algorithm for the synthesized image. The proposed method mainly considers the typical distortions, i.e., disoccluded region, stretching and global sharpness. Extensive experiments exhibit its effectiveness and efficiency.

## III. Proposed Quality Assessment Scheme

In this section, we introduce the proposed method in detail. More specifically, the proposed method treats the measurement of the geometric distortion and global sharpness as the key problem to be solved for evaluating the quality of the synthesized image. The geometric distortion is measured by two ways, i.e., disoccluded region evaluation and stretching strength evaluation. Whereas, the global sharpness is evaluated

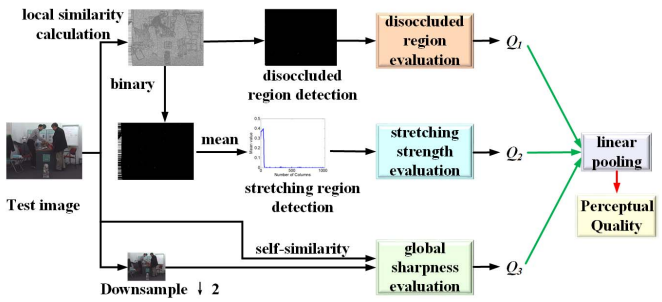



Fig. 2. Framework of the proposed scheme.

via the analysis of image self-similarity. Fig. 2 systematically depicts the framework of the proposed method.

### A. Disoccluded Region Evaluation

Digital images are full of complex structures and textures with strong correlations among pixels, meaning that one pixel is quite akin to its adjacent pixels in a local region. Benefiting from this, human beings can successfully recognize and comprehend image contents. During DIBR synthesis, the warping operation is utilized and induces undesirable geometric distortion in the form of disoccluded regions. Fig. 3(a) gives an intuitive comparison between disoccluded region and natural image patches. For convenience, we respectively mark the typical disoccluded region, smooth region, texture region and edge region by red, green, blue and orange rectangles. Meanwhile, these rectangles are enlarged for better visualization. It is clear that, as for natural patches (e.g., the green, blue and orange regions), pixels in the local region contain intrinsic regularity and form special patterns, like edge, texture, etc. By contrast, pixels in the disoccluded region (i.e., the red region) are the same and the associated region shows flat appearance.

Based on these observations, we attempt to detect disoccluded regions via the analysis of local image similarity. Such similarity is estimated by analyzing the relationship between one pixel and its adjacent pixels. Fig. 4 provides a brief illustration about the analysis procedure. In the figure, the light blue circle denotes the center pixel, while the remaining circles represent its adjacent pixels. Here, we take eight adjacent pixels as an example. For an adjacent pixel, the relationship (e.g., the gray-level difference) between it and the center pixel is calculated. This operation starts from the top circle and is conducted in the clockwise direction. In each case, we highlight the considered adjacent circle with a pink surface. Overall, we can totally obtain eight relationships between the center pixels and its adjacent pixels. The local similarity can be reflected by integrating these relationships together. In previous studies of texture classification and face recognition, LBP was widely verified and used because of its effective ability to excavate local correlation [41]–[43]. We therefore utilize LBP to measure the local correlation in this paper. It is worth emphasizing that, although some works have employed LBP in solving the IQA problem [35], [44]–[46], they have intrinsic differences compared to this work. Actually, those works directly took LBP as quality-sensitive features, which were trained by a machine learning tool to generate the quality evaluator. In contrast, the proposed method uses only the LBP to detect the disoccluded region.

Revisiting Fig. 4, the relationship between the center pixel $n_c$ and its $i$-th adjacent neighbor $n_i$ can be calculated by:

$$s(I(n_i), I(n_c)) = \begin{cases} 1, & if\ I(n_i) \geq I(n_c) \\ 0, & if\ I(n_i) < I(n_c) \end{cases} \tag{1}$$

where $I(n_c)$ denotes the gray value of center pixel, while $I(n_i)$, $i = \{0, 1, 2, 3, \cdots, P\text{-}1\}$ stands for the gray value of circularly symmetric neighbor pixel. $P$ is the number of surrounding neighbors. Apparently, for one pixel, a total of $P$ comparison results are obtained by using Eq. (1). Then, these comparison results can be encoded via binomial factor $2^i$ according to their locations:

$$\Lambda_P = \sum_{i=0}^{P-1} s(I(n_i), I(n_c)) \cdot 2^i. \tag{2}$$

By definition [43], LBP only possesses the gray invariance, yet it is lack of good attributes for object transformations, i.e., the rotation invariance. To expiate this, a rotation invariant uniform LBP map ($\Lambda_P^{riu2}$) was designed and formally defined as:

$$\Lambda_P^{riu2} = \begin{cases} \sum_{i=0}^{P-1} s(I(n_i), I(n_c)), & if\ \mu(\Lambda_P) \leq 2 \\ P+1, & otherwise \end{cases} \tag{3}$$

where $\mu$ is calculated as the number of bitwise transitions. It can be expressed as:

$$\begin{aligned} \mu(\Lambda_P) = &\|s(I(n_0), I(n_c)) - s(I(n_{P-1}), I(n_c))\| \\ &+ \sum_{i=1}^{P-1} \|s(I(n_i), I(n_c)) - s(I(n_{i-1}), I(n_c))\|. \end{aligned} \tag{4}$$

According to Eqs. (1)−(4), we can totally obtain $P+2$ different types of patterns from one uniform LBP map. Obviously, the parameter $P$ affects the characteristics of LBP descriptor. In this work, we set $P = 8$ according to discussion in Section IV-C. Besides, the distance between the centered pixel and its surrounding pixels may also have a direct influence on the detection results. With increasing the distance, LBP will no longer reflect local similarity anymore. Therefore, we set the distance as 1 since it, to the greatest extent, reflects the local information.

Broadly speaking, LBP is able to clarify different natural images or patches by analyzing textural information [41]. In reality, natural images or patches consist of various contents and their histograms of LBP patterns are consequently complex. In contrast, the disoccluded region has flat surface and lacks regularity. Hence, LBP can be treated as an effective tool for distinguishing natural image patches and the disoccluded patches. To illustrate this, an intuitive example is given in Fig. 3(b), where four sub-figures are involved. From the top left to bottom right, these sub-figures are LBP histograms of picture regions marked by the red, green, blue and orange rectangles in Fig. 3(a), respectively. For better visualization,

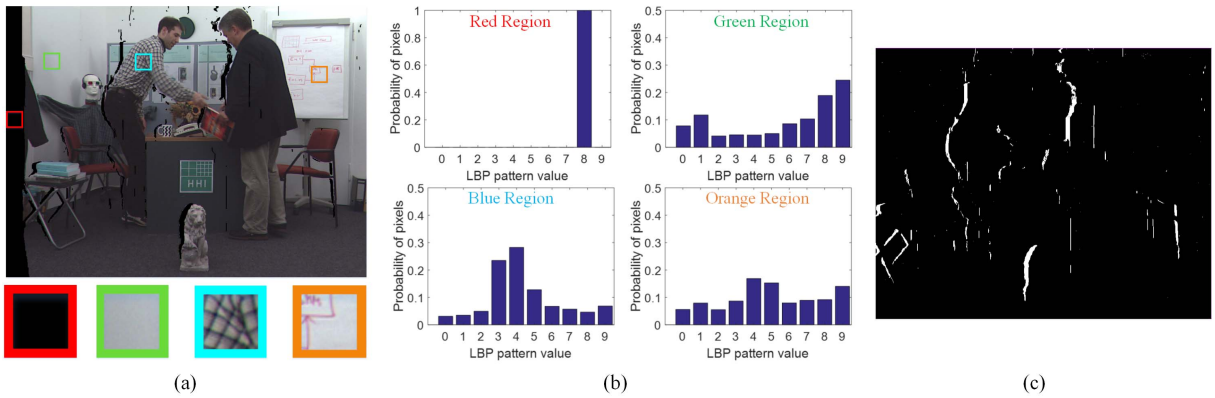



Fig. 3. Disoccluded region detection. (a) Comparison of disoccluded region and natural image patches. Typical patches are marked (and accordingly enlarged) by colorful rectangles (i.e., red, green, blue and orange), which contain disoccluded region, smooth, texture and edge, respectively. (b) LBP histograms of picture regions marked by the red, green, blue and orange rectangles. (c) Detection results. For better visualization, we inverse $D_r$ values.

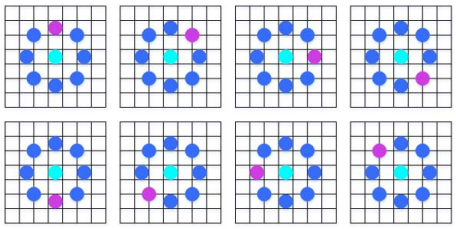


Fig. 4. Procedure of measuring relationship between one pixel and its adjacent pixels.

the texts are also highlighted by the associated color according to regions they represent. Considering Fig. 3(a) and Fig. 3(b) together, we can derive some meaningful conclusions: 1) LBP histograms of natural image patches are sufficiently complex and vary based on image content, i.e., smooth, texture and edge; whereas, 2) that of the disoccluded region is simple with only one bin equaling to one. Such regularity is not difficult to explain: all pixels in the disoccluded region have the same gray value, and thus, for a pixel, its gray value always equals to its neighbors. That is, all pixels in the disoccluded region maintain the same LBP encoding format. According to the above analysis, we have confidence that LBP can detect the disoccluded region effectively.

After obtaining the LBP maps, the disoccluded region map $D_r$ can be obtained via a binary operation:

$$D_r = \begin{cases} 0, & if\ \Lambda_P^{riu2} = 8 \\ 1, & otherwise \end{cases}, \tag{5}$$

where value 0 indicates the disocculded region. To eliminate some isolated small "noise", we subsequently clean the LBP map with the Gaussian filter. Moreover, given that some non-disoccluded regions, such as sky and white wall, may also possess bald appearance, we remove these regions via a threshold. Specifically, one connected region in LBP map is labeled as non-disoccluded region if its area is larger than 10% of the whole image. Fig. 3(c) shows the detection result of Fig. 3(a) using the local similarity analysis mentioned above. In Fig. 3(c) (as well as Fig. 5(b)), the white region has value 1, while the black region has value 0. For better visualization, we inverse $D_r$ values obtained by Eq. (5) in Fig. 3(b). As can be seen, the disoccluded region is well detected in the form of white region.

After obtaining the disoccluded region map $D_r$, our next concern is how to utilize it to estimate the image quality. In the literature, a vast majority of quality assessment methods were designed based on the natural scene statistics (NSS) model, which quantifies the distortion by analyzing the corruption degree of NSS regularity. Unfortunately, as discussed in reference [10], NSS model was incapable of reliably evaluating the quality of DIBR-synthesized images. Apart from NSS model, another well-proved thought is to measure the similarity between the distorted image and its reference image [19], [22], [47]. In this paper, we imitate the benchmark work (SSIM) and define a quality evaluator for the disoccluded region. Suppose that we have a reference image and its disoccluded region map is $D_R$, then, the quality ($Q_1$) of the synthesized image in terms of measuring the disoccluded region can be calculated as follows:

$$Q_1 = \frac{1}{K}\sum_{k=1}^{K}\frac{2D_r(k)\cdot D_R(k)+\varepsilon}{D_r(k)^2+D_R(k)^2+\varepsilon} \tag{6}$$

where $k$ indicates the pixel index, $K$ is the total number of pixels in an image, and $\varepsilon$ is a small positive constant to ensure the stability. Since the reference image (in an ideal situation) is free of disoccluded regions, all values in $D_R$ are the unit (with value of 1). Then, Eq. (6) can be rewritten as follows:

$$Q_1 = \frac{1}{K}\sum_{k=1}^{K}\frac{2D_r(k)\cdot \mathbf{1}+\varepsilon}{D_r(k)^2+\mathbf{1}+\varepsilon} \tag{7}$$

It is clear that the term $D_r(k)^2+\mathbf{1}$ in the denominator is not less than $\mathbf{1}$, indicating that the denominator cannot be zero. Next, we remove $\varepsilon$ and rewrite Eq. (7) as:

$$Q_1 = \overbrace{\frac{1}{K}\sum_{k\in K_0}\frac{2D_r(k)}{D_r(k)^2+\mathbf{1}}}^{Non-disoccluded\ region} + \overbrace{\frac{1}{K}\sum_{k\in K_1}\frac{2D_r(k)}{D_r(k)^2+\mathbf{1}}}^{Disoccluded\ region} \tag{8}$$


(a) (b) (c)

Fig. 5. Illustration of the stretching region. (a) One typical image contains a stretching region. For convenience, two sub-regions are highlighted by colorful rectangles. (b) Coarse stretching region map. (c) Average element values of each column in (b).

where $K_0$ and $K_1$ ($K = K_0 + K_1$) are pixel sets for the non-disoccluded and disoccluded regions, respectively. According to the previous definition (e.g., Eq. (5)), all pixel values in the non-disoccluded (disoccluded) region are 1 (0). Hence, Eq. (8) can be further simplified as:

$$\begin{aligned} Q_1 &= \frac{1}{K}\sum_{k\in K_0}\frac{2\cdot\mathbf{1}}{\mathbf{1}^2+\mathbf{1}} + \frac{1}{K}\sum_{k\in K_1}\frac{2\cdot\mathbf{0}}{\mathbf{0}^2+\mathbf{1}} \\ &= \frac{K_0}{K} \\ &= 1-\frac{K_1}{K} \end{aligned} \tag{9}$$

From Eq. (9), one can observe that the estimation of $Q_1$ does not rely on any reference information even though we suppose a "reference image" aforehand. In this sense, the proposed method is reference-free. Besides, Eq. (9) reveals another important information; that is, the area $K_1$ of the disoccluded region determines the synthesized quality. A larger $K_1$ will cause a smaller $Q_1$. To ensure that $K_1$ and $Q_1$ are positively correlated, we further process $Q_1$ as $Q_1 = 1 - Q_1$.

### *B. Stretching Strength Evaluation*

Stretching, induced by a failed in-painting operation, mainly occurs on the left or right side of an image. One typical example is given in Fig. 5(a). On the left side of Fig. 5(a), the edges of wall, coat and chair stretch to the left. It is obvious that the stretching region has the simple texture or structure with intrinsic differences as compared to natural images. In this section, we propose a simple but effective measurement to evaluate the stretching strength.

First, given an image, we calculate its LBP map using Eqs. (1)−(4). As expected, the stretching region possesses simple patterns while natural regions exhibit complex patterns. More specifically, most LBP patterns in the stretching region have the same value (i.e., 8). Then, we binarize the LBP map and obtain the coarse stretching region map:

$$D_s = \begin{cases} 1, & if\ \Lambda_P^{riu2} = 8 \\ 0, & otherwise. \end{cases} \tag{10}$$

As can be seen in Fig. 5(b), the outline of stretching region is primarily drawn by white regions. Whereas, almost all elements in natural region are zero. With this observation, the average element value in each column is computed to detect fine stretching region. As shown in Fig. 5(c), compared to natural region, the average element values are very high in stretching region. Moreover, an abrupt decline occurs on the common boundary between the stretching region and natural region. With this observation, we finally detect the fine stretching region with a threshold. More specifically, one column is involved in stretching region only if its average element value is larger than $T_1$. It is worth emphasizing that, as the stretching region only occurs in the left or right side of an image with a connected area, those qualified columns which do not connect the left or right side are ignored. In this study, we empirically set $T_1 = 0.2$.

After capturing the stretching region, our next concern is how to evaluate its strength, which indicates its impact on perceptual quality. In Fig. 5(a), two boxes are utilized to label different stretching sub-regions. Intuitively, the region in the red box seems to produce less annoyance than that in the green box although they are in the same stretching width. This may be attributed to the fact that, compared to the region in the green box, the region in red box is more similar to its adjacent natural region, and looks more natural. Therefore, we estimate the stretching strength by calculating the similarity between the stretching region and its equal-size adjacent natural region. The more similar these two regions are, the less the stretching will be perceptible. Since the human visual system is more sensitive to structures, we calculate the similarity in gradient domain. Given a stretching region $S_I$, its gradient magnitude can be computed as:

$$G_s = \sqrt{(S_I \otimes p_x)^2 + (S_I \otimes p_y)^2} \tag{11}$$

where $\otimes$ is the convolution operation; $p_y$ and $p_x$ are filter kernels in the vertical and horizontal directions, respectively. In this study, we employ Prewitt filters, which are expressed below:

$$p_y = p_x^T, \quad p_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix} \tag{12}$$

where symbol $^T$ denotes a transpose operation. Using Eq. (11), we can also obtain the gradient map $G_n$ of the adjacent equal-size natural region of the stretching region. Then,

the similarity between them can be estimated as follows:

$$S_g = \frac{2G_s \cdot G_n + T_2}{G_s^2 + G_n^2 + T_2} \tag{13}$$

where $T_2 = 0.01$ and is used to avoid the problem of division-by-zero. Finally, the stretching strength is evaluated as the standard deviation ($Q_2$) of $S_g$.

$$Q_2 = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(S_j - \overline{S_g})^2} \tag{14}$$

where $J$ is the element number in $S_g$; $S_j$ is the $j$-th element in $S_g$; $\overline{S_g}$ is the mean value of $S_g$. Obviously, a larger $Q_2$ corresponds to a lower perceptual quality.

### C. Global Sharpness Evaluation

Apart from the disoccluded region and stretching region, sharpness is another factor that should not be neglected in quality assessment of DIBR-synthesized images. In synthesized images, the blurring effect, caused during warping blur depth map, mainly exhibits around the transitions of background and foreground. In the literature, many sharpness measurements have been proposed via analysis of edge and gradient [48]–[51]. Here, we provide a simple but effective method to measure the global sharpness of the synthesized image. The key strategy of the proposed method is the estimation of inter-scale self-similarity.

It is well known that natural images are scale-invariant, which indicates that an image shows a similar pixel distribution and exhibits similar property as its translated, zoom in and zoom out versions [52]. Suppose that distortions may distort such characteristics, a sharpness evaluator is designed in this subsection. Given an image $I_0$, we first obtain its down-scale version $I_1$ by a downsampling operation with factor 2. Then, the global sharpness ($Q_3$) of $I_0$ is evaluated via estimating its self-similarity, that is, measuring the distance of standard deviations between $I_0$ and $I_1$:

$$Q_3 = \frac{1}{N}\sum_{n=1}^{N}\sqrt{|\delta_{0,n}^2 - \delta_{1,n}^2|} \tag{15}$$

where $N$ is the total number of non-overlapping image block. $\delta_{0,n}^2$ and $\delta_{1,n}^2$ are the standard deviations of the $i$-th image block in $I_0$ and $I_1$, respectively. It should be noted that, during standard deviation calculation, the block size of $I_1$ is also processed by a downsample operation with factor 2 as that of $I_0$. It is obvious that, a small value of $Q_3$ denotes a small difference between $I_0$ and $I_1$, thereby indicating a high perceptual quality with good sharpness.

### D. Perceptual Quality Estimation

With these quality scores of the disoccluded region ($Q_1$), stretching strength ($Q_2$) and global sharpness ($Q_3$), our next task is how to effectively pool them together to infer the overall quality. According to the analysis in Sections III-A−III-C, all these scores keep monotonous with perceptual quality. Therefore, we directly integrate them linearly:

$$Q = \alpha_1 \cdot Q_1 + \alpha_2 \cdot Q_2 + \alpha_3 \cdot Q_3 \tag{16}$$

**Algorithm 1** Proposed Quality Assessment Scheme

**Input:** Color image $I_c$;
**Output:** Quality score $Q$

1: Convert color image $I_c$ to gray-scale image $I$;
2: Compute the disoccluded region using Eqs. (1)−(5);
3: Evaluate the disoccluded region $Q_1$ using Eqs. (6)−(9);
4: Compute the stretching region using Eq. (10);
5: Evaluate stretching strength $Q_2$ using Eqs. (11)−(14);
6: Evaluate global sharpness $Q_3$ using Eq. (15);
7: Compute the image quality score $Q$ by linearly combining $Q_1$, $Q_2$ and $Q_3$ using Eq. (16).

where the parameters $\alpha_1$, $\alpha_2$ and $\alpha_3$ ($\alpha_3 = 1 - \alpha_1 - \alpha_2$) are employed to balance the relative contribution of each part. During implementation, we set them as 0.9787, 0.0143 and 0.0070 based on the discussion in Section IV-C. To clarify the proposed method, the main steps are summarized as follows (in Algorithm 1).

## IV. Experimental Results and Analysis

In this section, we first introduce the experimental protocol. Then, the effectiveness of the proposed method is validated through extensive comparisons. Additionally, the analysis about the achieved performance is conducted.

### A. Experimental Protocol

*1) Test Platform:* In this study, IRCCyN/IVC synthesized image database is chosen as the test platform [12]. A total of 84 synthesized images (with size $1024 \times 768$) are created from 12 reference images selected from three MVD sequences in this database. For each reference image, it is processed by 7 different DIBR algorithms (A1−A7). A brief introduction about A1−A7 is given below.

- A1: the depth map is filtered to eliminate discontinuities [8]. Borders are cropped and the image is interpolated to its original size.
- A2: the depth map is also preprocessed as that in A1 [8]. Borders are inpainted as described in [53].
- A3: it is proposed by Mori *et al.* [54] and is adopted by the 3D video group as a reference software. It uses the inpainting method to fill the disoccluded area [53].
- A4: the disoccluded area is filled with the help of depth information [55]. The depth values surrounding the disoccluded area are examined row-wise to search the background color samples for filling the disoccluded area.
- A5: the missing part in the virtual view is filled by the known contents [56].
- A6: based on A5, it also considers the temporal information to improve the synthesis in the disoccluded area [57].
- A7: it corresponds to the originally synthesized view without any postprocessing.

The subjective scores are given in the form of mean opinion score (MOS) from 21 observers through subjective evaluations at viewing distance of 6 times the screen height using a "Pair Comparison" protocol method. Since during subjective evaluation, the reference images are hidden, in order to use the

subjective ratings for FR IQA methods, the MOS should be transformed into difference mean opinion score (DMOS) [14]:

$$DMOS = 5 + M_{syn} - M_{ref} \tag{17}$$

where $M_{syn}$ and $M_{ref}$ denote the MOS values of the synthesized image and its reference image, respectively.

*2) Competing Methods:* To investigate the effectiveness and superiority of the proposed method, we compare the proposed method with numerous recently developed IQA methods. According to the application, these methods can fall into two categories. Both categories contain FR metrics and NR methods. Methods in the first category were designed for 2D images, including the following:

- SSIM [22], which estimates image quality by combining the deviations between the distorted image and reference image in terms of contrast, luminance and structural similarities.
- IW-SSIM [23], which is deviated from SSIM by taking the information content weight into consideration.
- FSIM [47], which adopts the similar idea of SSIM and evaluates quality by computing the similarity between the distorted image and reference image in terms of gradient magnitude and phase congruency.
- MAD [58], which utilizes two strategies to separately estimate perceptual distortions in high-quality and low-quality images.
- ADD-SSIM [19], which changes the pooling measure of SSIM and designs a pooling model via analysis of distortion distribution.
- VSNR [59], which considers visual masking, near- and supra-threshold properties, and visual summation effects in the wavelet domain to estimate image quality.
- NIQE [60], which builds a model by analyzing a corpus of high-quality images and evaluates the quality of the distorted image by computing the distance between its features and the model.
- IL-NIQE [61], which deviates from NIQE by considering more features during building the model.
- QAC [62], which estimates image quality using a set of the learnt quality-aware centroids.

In the second category, six methods that specifically proposed for DIBR-synthesized images are involved, including:

- 3D-SWIM [14], which measures the histogram distance of Haar wavelet between the reference block and its matched block in the synthesized view to evaluate the image quality.
- MW-PSNR [27], which decomposes the reference and synthesized images using the morphological wavelet transform. The quality score is obtained by pooling the multi-scale wavelet peak to noise ratio across scales.
- MP-PSNR [28], which is similar to MW-PSNR, but replaces the morphological wavelet by morphological pyramids.
- NIQSV [39], which utilizes morphological operators to estimate the distortions.
- NIQSV+ [40], which is the extended version of NIQSV by further considering the stretching distortion.
- APT [10], which takes the free energy model (plus a threshold) to detect the geometric distortion for quality assessment.

*3) Evaluation Criteria:* Three commonly used evaluation criteria, including the Spearman rank correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC) and Root mean-squared error (RMSE), are employed to evaluate and compare the proposed method with existing IQA methods in this study. Among these criteria, PLCC reflects the prediction accuracy, SRCC is computed to represent the prediction monotonicity, and RMSE evaluates the prediction consistency. With experience, a smaller value of RMSE, unlike those of PLCC and SRCC, indicates superior performance of the tested method. As suggested by the video quality expert group [63], the estimated score of an objective IQA method should be fed into a nonlinear regression function before computing PLCC and RMSE. In this study, we choose the same five-parameter logistic regression function as the competing methods for comparison fairness. By definition, it can be expressed as:

$$f(x) = \tau_1 \cdot \left[\frac{1}{2} - \frac{1}{1 + e^{\tau_2 \cdot (x - \tau_3)}}\right] + \tau_4 \cdot x + \tau_5 \tag{18}$$

where $x$ is the estimated score set of test images by an objective IQA method. Let $x$ as the input and the DMOS values as the output of Eq. (18), parameters $\{\tau_1, \tau_2, \cdots, \tau_5\}$ can be fitted by optimizing the difference between DMOS values and the fitted scores $f(x)$. With the help of Eq. (18), $x$ is mapped to $f(x)$, which has the similar magnitude range as that of DMOS. Therefore, it is fair to make comparisons on prediction accuracy among different objective methods after processing their estimated scores by Eq. (18).

*4) Implementation Detail:* As introduced in Section III, two parameters, including parameter $P$ used in the LBP calculation, and local block size $B$ used in the sharpness evaluation, may have directly affect the overall performance. In this study, we respectively set them as 8 and $32 \times 32$ based on experiments in Section IV-C.

### B. Performance Comparison

To ensure the comparison fairness, all these methods are implemented on the same test environment (Matlab R2016b software) by running the released or provided codes by the authors. Comparison results are given in Table I. For reader's convenience, we highlight the top two methods with boldface in each type. From the table, some interesting and meaningful conclusions can be clearly derived. First, the methods designed for 2D images are incompetent for evaluating quality of the synthesized image. Across the considered nine metrics, MAD obtains the best performance with PLCC $= 0.608$, SRCC $= 0.599$ and RMSE $= 0.529$. This directly verifies their incapacity on evaluating the synthesized image and reflects the necessity of designing specific methods for the synthesized image. Besides, scrupulous readers may also observe that, among these methods, FR methods generally exhibit higher performance than NR methods. This is not difficult to understand as FR methods possess reference information, and thus, they have more advantages over NR methods. Second, methods (i.e., designed for the synthesized image) in the second

TABLE I

PERFORMANCE COMPARISONS. THE TOP TWO RESULTS IN EACH TYPE ARE HIGHLIGHTED WITH BOLDFACE

| Metrics | Type | Category | PLCC | SRCC | RMSE |
|---|---|---|---|---|---|
| SSIM [22] | FR | 2D image | 0.485 | 0.437 | 0.582 |
| IW-SSIM [23] | FR | 2D image | 0.583 | 0.405 | 0.541 |
| FSIM [47] | FR | 2D image | 0.583 | 0.415 | 0.541 |
| MAD [58] | FR | 2D image | 0.608 | 0.599 | 0.529 |
| ADD-SSIM [19] | FR | 2D image | 0.551 | 0.467 | 0.556 |
| VSNR [59] | FR | 2D image | 0.437 | 0.385 | 0.599 |
| MW-PSNR [27] | FR | synthesized image | 0.562 | 0.576 | 0.551 |
| 3D-SWIM [14] | FR | synthesized image | **0.658** | **0.616** | **0.501** |
| MP-PSNR [28] | FR | synthesized image | **0.617** | **0.623** | **0.524** |
| NIQE [60] | NR | 2D image | 0.437 | 0.374 | 0.599 |
| IL-NIQE [61] | NR | 2D image | 0.494 | 0.526 | 0.579 |
| QAC [62] | NR | 2D image | 0.352 | 0.311 | 0.623 |
| NIQSV [39] | NR | synthesized image | 0.643 | 0.483 | 0.510 |
| NIQSV+ [40] | NR | synthesized image | 0.654 | 0.490 | 0.504 |
| APT [10] | NR | synthesized image | **0.731** | **0.716** | **0.455** |
| Proposed | NR | synthesized image | **0.675** | **0.652** | **0.462** |

category generally outperform those in the first category. This is because the synthesized image has different distortion types and existing algorithms in the second category successfully catch and measure such distortions to some extent. By contrast, methods for 2D images fail to do this but are capable of measuring common distortions, like blurriness, blockiness, contrast change and noise. Third, the proposed method is superior to all competing FR methods without the help of reference information. More concretely, our method obtains approximately 3.6 % dominance of SRCC than the best FR methods considered (i.e., 3D-SWIM). Last but not the least, compared to these NR methods, the proposed method still occupies a dominant position. To be specific, it outperforms all competing NR methods but leaves a slight gap to the best metric APT. Despite this, as discussed in the next subsection, the proposed method has an obvious advantage in implementation time over APT. Overall, the proposed method exhibits powerful superiority than all competing methods considered.

For an intuitive comparison, we further give the scatter plots of the estimated objective scores (these scores are actually post-processed by Eq. (18)) versus DMOS values, as shown in Fig. 6. As can be seen, the points of the proposed method and APT look more monotonic and linear compared with other competing metrics, and maintain more consistency in line with subjective feelings. For convenience, the PLCC and SRCC values of each IQA method are given in the brackets on the associated caption respectively.

As introduced in Section III, three quality-related distortion components are comprehensively considered and subsequently integrated toward the generation of the proposed algorithm. In this paragraph, another experiment is conducted to investigate the contribution of each component (i.e., disoccluded region evaluation, stretching strength and sharpness evaluation) on the final performance. Table II presents the results of PLCC, SRCC and RMSE. In the table, "$Q_1$", "$Q_2$" and "$Q_3$" respectively correspond to the situation that only uses the score of disoccluded region evaluation, stretching strength and sharpness evaluation. As expected, all components have

TABLE II

PERFORMANCE OF EACH COMPONENT CONSIDERED

| Metrics | PLCC | SRCC | RMSE |
|---|---|---|---|
| $Q_1$ | 0.598 | 0.486 | 0.534 |
| $Q_2$ | 0.483 | 0.450 | 0.583 |
| $Q_3$ | 0.532 | 0.302 | 0.539 |
| Overall | 0.675 | 0.652 | 0.462 |

TABLE III

SRCC VALUES OF DIFFERENT DIBR ALGORITHMS (A1−A7)

| Criteria | A1 | A2 | A3 | A4 | A5 | A6 | A7 |
|---|---|---|---|---|---|---|---|
| SRCC | 0.667 | 0.859 | 0.857 | 0.286 | 0.603 | 0.821 | 0.964 |

positive effects on the final performance. Taking Tables I−II together, it is clear that the performance of each component stays at a medium level compared with competing NR methods. However, the performance, denoted as "Overall", is greatly increased when combining these components linearly using Eq. (16). In summary, the proposed method obtains considerable results and has potentials to effectively solve the IQA problem for synthesized images.

In addition, we also examine the effectiveness of the proposed method on each DIBR algorithm (i.e., A1−A7). For this purpose, the dataset is first divided into seven parts according to the processing algorithms of each image. Then, the perceptual quality scores of all images in each part are estimated using the proposed method. Finally, the experimental results are assessed via the SRCC value for each part, as shown in Table III. It is apparent that the proposed method gets passable performance on most algorithms, such as A1−A3, and A6−A7. But simultaneously, it fails to evaluate algorithms A4 and A5. This may be attributed to A4 and A5 filling the disoccluded regions with depth or texture information, which unexpectedly produces blurry around these disoccluded regions. Since our method places more effort on detecting and evaluating the disoccluded region, stretching region and sharpness, it unfortunately exposes its insufficient ability to evaluate quality of images processed by algorithms A4 and A5. Besides, as shown in Fig. 1, multiple distortion types may be involved in one image. However, most existing metrics (e.g., NIQSV, APT as well as the proposed method) merely consider some of distortion types, they accordingly achieve limited overall performance and retain bias on different DIBR algorithms. In spite of this, our method selectively chooses some more significant distortion types and obtains an encouraging performance with enviable superiority. This motivates us to insistently move forward towards completely solving this challenge.

It is clear that APT and proposed method lead to the NR methods. In this paragraph, we further investigate whether these two leading methods can be combined to generate a more effective NR method. Specifically, the combined method is defined as follows:

$$Q_c = \beta_1 \cdot Q + \beta_2 \cdot APT \tag{19}$$

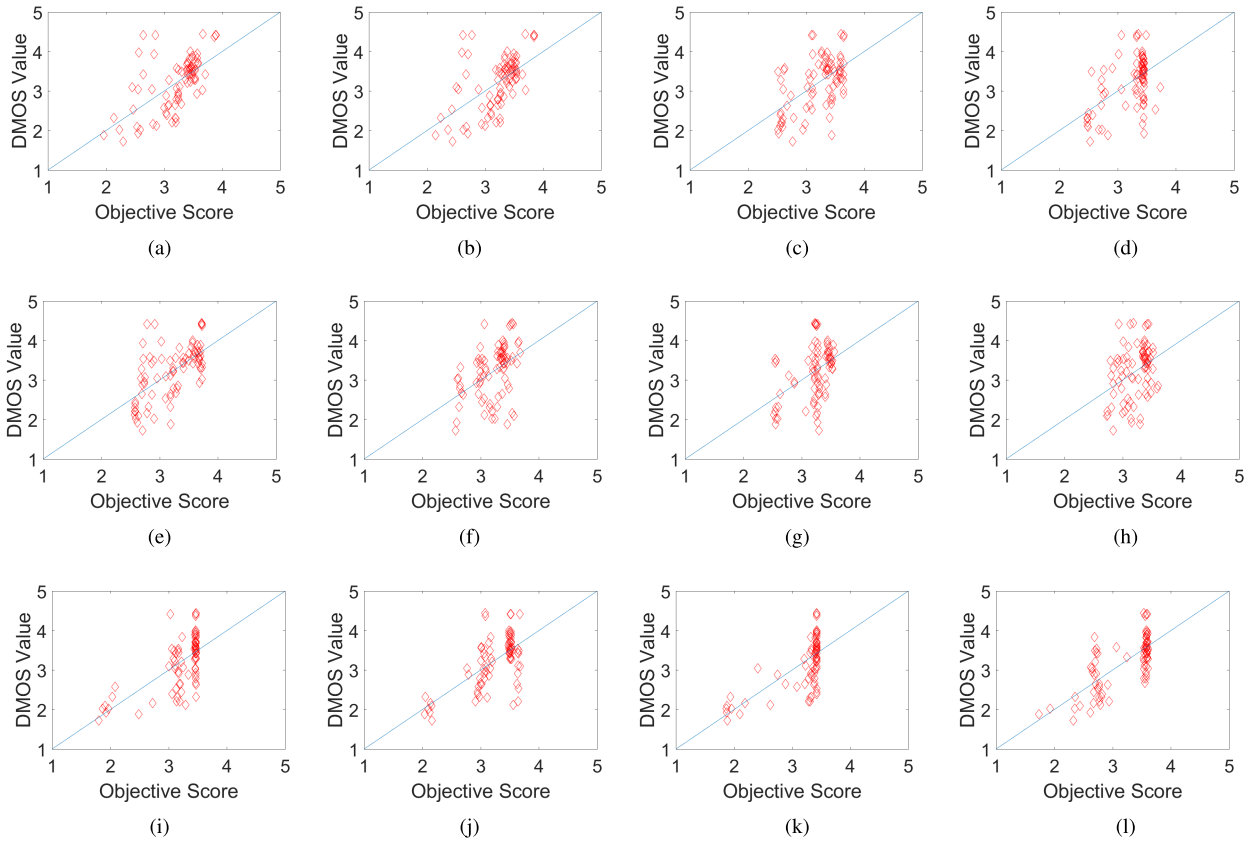



Fig. 6. Scatter plots of the estimated objective scores vs. DMOS values. Since we pay more attention on comparisons with NR methods, only partial FR methods are involved here. For convenience, the name of each objective method is given in the caption of the associated subfigure. Meanwhile, the PLCC and SRCC values are given in the brackets respectively. (a) MP-PSNR [28], (0.617,0.623). (b) MW-PSNR [27], (0.562,0.576). (c) ADD-SSIM [19], (0.551,0.467). (d) SSIM [22], (0.485,0.437). (e) MAD [58], (0.608,0.599). (f) IL-NIQE [61], (0.494,0.526). (g) NIQE [60], (0.437,0.374). (h) QAC [62], (0.352,0.311). (i) NIQSV [39], (0.643,0.483). (j) NIQSV+ [40], (0.654,0.490). (k) APT [10], (0.731,0.716). (l) Proposed, (0.675,0.652).

where $\beta_1$ and $\beta_2$ are weighting parameters to linearly combine the proposed method and APT. Fig. 7 illustrates the SRCC values of the combined method with different $\beta_1$ and $\beta_2$ settings. It is observed that the combined method obtains the best result (SRCC $= 0.726$) when $\beta_1$ and $\beta_2$ are set as $0.6 \times 10^{-3}$ and 323.5, respectively. Meanwhile, we can obtain PLCC $= 0.764$ with this parameter setting. By comparing the results in Table I, one can find that the combined method is superior to all NR methods. More concretely, it brings approximately 3% and 1% performance gains of PLCC and SRCC, respectively, compared to the most effective method APT. With this observation, we can selectively choose different methods according to different practical NR applications. If an application focuses more on effectiveness, one can choose the combined method. Whereas, the proposed method should be selected when the efficiency is necessary.

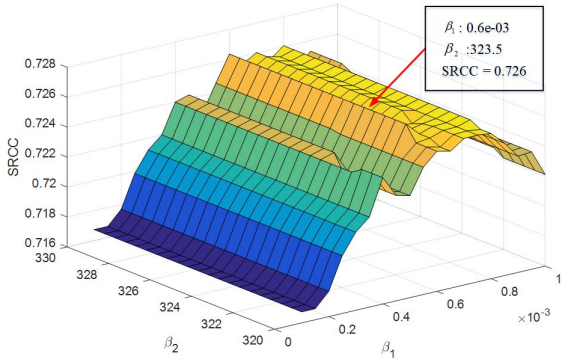



Fig. 7. The impact of parameters ($\beta_1$ and $\beta_2$ used in Eq. (19)) on the performance of the combined method.

### C. Parameters Sensitivity

In this subsection, we investigate the impacts of parameters on experimental results. Totally, two parameters are considered, i.e., parameter $P$ used in LBP calculation and local block size $B$ used in sharpness evaluation. Here, we explore their impacts separately. One parameter should be fixed when another parameter's impact is explored. First, we set $B$ as $32 \times 32$ and choose a $P$ value from $\{4, 8, 16\}$ in a sequential order. Fig. 8(a) depicts the performance under different arrangements. Clearly, parameter $P$ has slight influence on the overall performance. To be specific, when $P = 8$ and $P = 16$, the associated results are similar and superior to the case of $P = 4$. This phenomenon is possibly attributed to that only considering the surrounding 4 pixels is not enough for reflecting local similarity. As the considered surrounding pixels increase, the local similarity seems to be well reflected

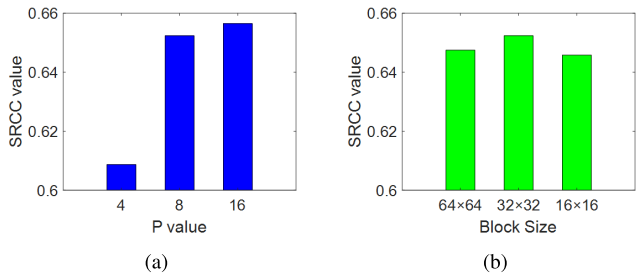



Fig. 8. The impact of parameters on the overall performance. (a) SRCC vs. $P$, (b) SRCC vs. $B$.

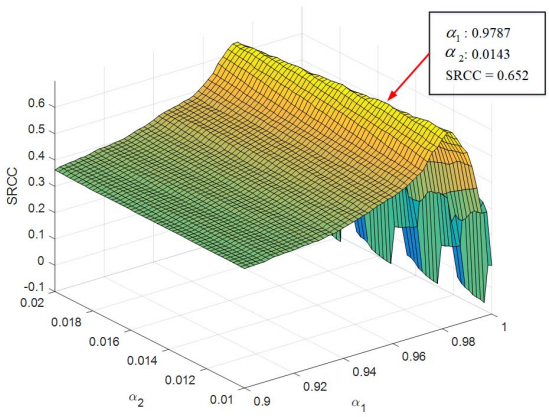



Fig. 9. Impact of parameters ($\alpha_1$ and $\alpha_2$ used in Eq. (16)) on the performance of the proposed method.

and the performance tends to be stable when $P$ hovers between 8 and 16. Since the increase in surrounding pixels burdens the computational cost, parameter $P$ is arranged as 8 in this study. Next, we fix $P = 8$ and change $B$ from $64 \times 64$, $32 \times 32$ to $16 \times 16$. As illustrated in Fig. 8(b), the block size $B$ does not affect the result too much. Therefore, we set it as $32 \times 32$ due to its slight advantage on performance.

It is worth noting that $\alpha_1$, $\alpha_2$ and $\alpha_3$ in Eq. (19) also have direct influence on the overall performance. To avoid their influence, they are chosen during the value selection of $P$ and $B$ for achieving the best performance. For example, let $P = 8$ and $B = 32 \times 32$. $\alpha_1$ ranges from 0.01 to 0.02 in steps of 0.0003 and $\alpha_2$ ranges from 0.9 to 1 in steps of 0.0003. Once they are chosen, $\alpha_3$ is determined by $\alpha_3 = 1 - \alpha_1 - \alpha_2$. Fig. 9 shows the SRCC values under different $\alpha_1$ and $\alpha_2$ settings. By observation, we find that the proposed method obtains the best result (SRCC $= 0.652$) when $\alpha_1 = 0.9787$ and $\alpha_2 = 0.0143$. Actually, the above parameter selection procedure is conducted for every $P$ and $B$ combination. Finally, we set $\alpha_1 = 0.9787$, $\alpha_2 = 0.0143$, $P = 8$ and $B = 32 \times 32$ for achieving the best performance of the proposed method.

### D. Implementation Time

Apart from effectiveness, implementation time is another concerned factor during designing IQA methods. In fact, a time-saving method is more favored and welcomed. Here, we also conduct an experiment to investigate the implementation time of the proposed method and compare it with competing NR methods. To avoid inadvertently mistakes, all methods are implemented on the same test platform (MATLAB R2016b software executed on a 3.2 GHz processor with 16 GB RAM, Windows 7 Pro 64-bit desktop). For each method, the total time consumed is recorded by functions (*tic* and *toc*) on the entire IRCCyN/IVC image database. Table IV summaries the processing time per image (in seconds) of seven NR methods. It can be intuitively observed that most methods except APT are time-saving and fast to implement (with less than 0.5s for processing an image). More especially, the proposed method only requires about 0.2s. In contrast, APT requiring autoregression procedure is time-consuming and costs more than hundreds of times than these time-saving methods. Taking the results in Table I into account together, it seems that there is a tradeoff between the effectiveness and implementation time. To be specific, these competing methods perform well on either effectiveness or implementation time, but none of them considers these two elements simultaneously. By contrast, the proposed method obtains considering performance (is slightly inferior to APT) and requires shorter implementation time (is slightly inferior to NIQSV and NIQSV+). In other words, the proposed method well balances the effectiveness and implementation time, therefore, it may have more potentials than competing methods.

TABLE IV
RUN TIME COMPARISON AMONG DIFFERENT NR METHODS (IN SECONDS)

| Metric | Time (s) | Metric | Time (s) |
|---|---|---|---|
| NIQE [60] | 0.489 | IL-NIQE [61] | 5.907 |
| NIQSV [39] | 0.097 | NIQSV+ [40] | 0.113 |
| APT [10] | 90.454 | QAC [62] | 0.212 |
| Proposed | 0.196 | | |

### E. Discussion

The MVD format together with DIBR algorithms provide us a new solution for reducing data amount of FVV during transmission and storage. Proposing quality assessment method for DIBR-synthesized images is fundamental and critical for proposing a perceptual friendly FVV system. In this study, we have designed a novel quality assessment method for DIBR-synthesized images. Through several experiments, it is shown that the proposed method exhibits considerable performance in terms of effectiveness and implementation time. To be specific, it is better than all competing FR methods and is slightly inferior to the state-of-the-art NR method APT in terms of effectiveness. Meanwhile, the proposed method is time-saving and only leaves a small gap (about 0.1s) to the fastest method NIQSV. More importantly, it requires obviously shorter implementation time than the most effective method APT. Specifically, only 0.196s is used by the proposed method while more than 90s is used by APT for processing an image. Overall, as revealed in Table I and Table IV, further works still require for performance improvement. For example, more efforts can be paid on effectiveness improvement by considering and quantifying more DIBR-related distortions, such

as object shifting, blurry and crumbling. Meanwhile, other measures (such as variation, entropy and moment statistics of pixels in local region) can also be utilized to assist the proposed solution in analyzing the local similarity for estimating the disocculded region, stretching, etc. It is worth noting that, time-saving features should be preferentially considered to supplement currently used features for improving effectiveness or replace currently used features for reducing implementation time. Apart from improving performance, another issue should be addressed is how to extend the proposed method for evaluating the quality of entire DIBR-synthesized FVVs in the future work. For this purpose, the visual angle, binocular perception model and the temporal distortion (e.g., flickering) should be taken into account.

## V. Conclusion

In this paper, we make an attempt to estimate the quality of DIBR-synthesized images without requiring reference information. The proposed method mainly considers three DIBR-related distortions, i.e., disoccluded region, stretching and global sharpness. Specifically, a local similarity analysis is first used to detect disoccluded regions. Then, the stretching region is determined via the analysis of local similarity map followed by a threshold. Next, the global sharpness is evaluated by measuring inter-scale self-similarity. At the end, these three components are measured separately and combined linearly to infer the overall perceptual quality. Extensive experiments on the IRCCyN/IVC database prove the superiority of the proposed method. More concretely, it obtains considerable performance (PLCC $=$ 0.675, SRCC $=$ 0.652 and RMSE $=$ 0.462) and outmatches the mainstream FR IQA methods. Besides, compared to those NR methods, it is merely inferior to APT and occupies the second place. However, our method is more time-saving (only requiring 0.196s to process an image) and therefore possesses more application potentials than APT. Moreover, by combining the proposed method and APT together, we can produce a more effective method. This gives us broader choices for coping with different practical NR applications.

## Appendix

For reader's convenience, we list the abbreviations used in this paper as below.

| Abbreviations | Full Name |
|---|---|
| 3D | Three-dimensional |
| FVV | Free Viewpoint Video |
| MVD | Multi-view-Video-plus-Depth |
| DIBR | Depth-Image-Based-Rendering |
| FR | full-reference |
| RR | reduced-reference |
| NR | no-reference |
| IQA | image quality assessment |
| MSE | mean squared error |
| IRCCyN/IVC | the name of the image database |

## References


[1] B. Javidi and F. Okano, *Three-Dimensional Television, Video, and Display Technologies*. New York, NY, USA: Springer-Verlag, 2002.

[2] G. Yue, C. Hou, J. Lei, Y. Fang, and W. Lin, "Optimal region selection for stereoscopic video subtitle insertion," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2739756.

[3] J. Lei *et al.*, "Shape-preserving object depth control for stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2749146.

[4] W. Yan, C. Hou, J. Lei, Y. Fang, Z. Gu, and N. Ling, "Stereoscopic image stitching based on a hybrid warping model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 1934–1946, Sep. 2017.

[5] 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.

[6] A. Vetro, A. M. Tourapis, K. Müller, and T. Chen, "3D-TV content storage and transmission," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 384–394, Jun. 2011.

[7] Y. Liu *et al.*, "A novel rate control technique for multiview video plus depth based 3D video coding," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 562–571, Jun. 2011.

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

[9] L. Wang, C. Hou, S. Qi, and L. Jiang, "Analysis of maximum tolerant depth distortion in view synthesis," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 7909–7927, 2018.

[10] 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. 2017.

[11] F. Battisti and P. Le Callet, "Quality assessment in the context of FTV: Challenges first answers and open issues," *IEEE COMSOC MMTC Commun.-Frontiers*, vol. 11, no. 2, pp. 22–27, 2016.

[12] 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.

[13] P.-H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," *Proc. SPIE*, vol. 8288, p. 82881M, Feb. 2012.

[14] 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.

[15] 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.

[16] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[17] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[18] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2018.2851984.

[19] 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.

[20] G. Yue, C. Hou, T. Zhou, and X. Zhang, "Effective and efficient blind quality evaluator for contrast distorted images," *IEEE Trans. Instrum. Meas.*, to be published, doi: 10.1109/TIM.2018.2868555.

[21] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.

[22] 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.

[23] 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.

[24] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

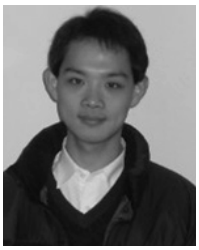
[25] G. Yue, C. Hou, K. Gu, S. Mao, and W. Zhang, "Biologically inspired blind quality assessment of tone-mapped images," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2525–2536, Mar. 2018.

[26] X. Zhang, W. Lin, S. Wang, J. Liu, S. Ma, and W. Gao, "Fine-grained quality assessment for compressed images," *IEEE Trans. Image Process.*, to be published, doi: 10.1109/TIP.2018.2874283.

[27] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *Proc. IEEE 7th Int. Workshop Quality Multimedia Exper. (QoMEX)*, May 2015, pp. 1–6.
[28] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. IEEE 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2015, pp. 1–4.
[29] Y. Zhou, L. Li, K. Gu, Y. Fang, and W. Lin, "Quality assessment of 3D synthesized images via disoccluded region discovery," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Sep. 2016, pp. 1012–1016.
[30] Y. Zhou, L. Li, K. Gu, Z. Lu, B. Chen, and L. Tang, "DIBR-synthesized image quality assessment via statistics of edge intensity and orientation," *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 8, pp. 1929–1933, 2017.
[31] H. Liu and I. Heynderickx, "A no-reference perceptual blockiness metric," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 865–868.
[32] S. Wang, C. Deng, B. Zhao, G.-B. Huang, and B. Wang, "Gradient-based no-reference image blur assessment using extreme learning machine," *Neurocomputing*, vol. 174, pp. 310–321, Jan. 2016.
[33] G. Yue, C. Hou, K. Gu, and N. Ling, "No reference image blurriness assessment with local binary patterns," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 382–391, Nov. 2017.
[34] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.
[35] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.
[36] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[37] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[38] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[39] 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.
[40] S. Tian, L. Zhang, L. Morin, and O. Déforges, "NIQSV+: A no-reference synthesized view quality assessment metric," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1652–1664, Apr. 2018.
[41] Z. Guo, X. Wang, J. Zhou, and J. You, "Robust texture image representation by scale selective local binary patterns," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 687–699, Feb. 2016.
[42] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.
[43] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[44] Q. Li, W. Lin, and Y. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93–103, May 2016.
[45] G. Yue, C. Hou, K. Gu, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2722–2732, Oct. 2018.
[46] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, Apr. 2018.
[47] 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.
[48] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.
[49] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
[50] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.
[51] Y. Zhan and R. Zhang, "No-reference image sharpness assessment based on maximum gradient and variability of gradients," *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1796–1808, Jul. 2018, doi: 10.1109/TMM.2017.2780770.
[52] D. Kersten, "Predictability and redundancy of natural images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2395–2400, 1987.
[53] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph. Tools*, vol. 9, no. 1, pp. 23–34, 2004.
[54] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Process., Image Commun.*, vol. 24, nos. 1–2, pp. 65–72, 2009.
[55] 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, p. 438148, 2009.
[56] 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.
[57] M. Köppel *et al.*, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 1809–1812.
[58] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, 2010.
[59] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[60] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[61] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[62] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 995–1002.
[63] Video Quality Experts Group (VQEG). (2003). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.vqeg.org/

**Guanghui Yue** received the B.S. degree in communication engineering from Tianjin University, Tianjin, China, in 2014, where he is currently pursuing the Ph.D. degree in information and communication engineering with the School of Electrical and Information Engineering. He has been a Joint Ph.D. Student with the School of Computer Science and Engineering, Nanyang Technological University, since 2017. His research interests include bioelectrical signal processing, multimedia quality assessment, 3-D image visual discomfort prediction, and pattern recognition.

**Chunping Hou** received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively. Since 1986, she has been with the School of Electronic and Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3-D Imaging Institute. Her current research interests include 3-D image processing, 3-D display, wireless communication, and the design and applications of communication systems.



**Ke Gu** received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include image analysis, environmental perception, quality assessment, and machine learning. He was the Leading Special Session Organizer in VCIP 2016 and ICIP 2017. He received the Best Paper Award of the IEEE TRANSACTIONS ON MULTIMEDIA, the Best Student Paper Award from the IEEE International Conference on Multimedia and Expo in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He serves as a Guest Editor for the *Digital Signal Processing* journal. He is currently an Associate Editor of the IEEE ACCESS and *IET Image Processing*. He is a reviewer of 20 top SCI journals.

**Tianwei Zhou** received the B.S. degree in automation from Tianjin University, Tianjin, China, in 2014, where she is currently pursuing the Ph.D. degree in control science and engineering with the School of Electrical and Information Engineering. She has been a Joint Ph.D. Student with the Department of Electrical and Computer Engineering, National University of Singapore, since 2017. Her current research interests include synchronization, networked control systems, event-triggered control, and image processing.

**Guangtao Zhai** (M'10) received the B.E. and M.E. degrees from Shandong University, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen–Nuremberg, Germany. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing. He received the National Excellent Ph.D. Thesis Award from the Ministry of Education of China in 2012.