Quality Assessment of DIBR-Synthesized Images by Measuring Local Geometric Distortions and Global Sharpness

Leida Li, *Member, IEEE*, Yu Zhou, Ke Gu, Weisi Lin, *Fellow, IEEE*, and Shiqi Wang, *Member, IEEE*

***Abstract*—Depth-image-based rendering (DIBR) is a fundamental technique in free viewpoint video, which is widely adopted to synthesize virtual viewpoints. The warping and rendering operations in DIBR generally introduce geometric distortions and sharpness change. The state-of-the-art quality indices are limited in dealing with such images since they are sensitive to geometric changes. In this paper, a new quality model for DIBR-synthesized view images is presented by measuring LOcal Geometric distortions in disoccluded regions and global Sharpness (LOGS). A disoccluded region detection method is first proposed using SIFT-flow-based warping. Then, the sizes and distortion strength of local disoccluded regions are combined to generate a score. Furthermore, a reblurring-based strategy is proposed to quantify the global sharpness. Finally, the overall quality score is calculated by pooling the scores of local disoccluded regions and global sharpness. Experiments on four public DIBR-synthesized image/video databases show the superiority of the proposed metric over the state-of-the-art quality models. The proposed method is further adopted for boosting the performances of existing quality metrics and benchmarking DIBR algorithms, both achieving very promising results.**

***Index Terms*—DIBR, disoccluded region, quality evaluation, ranking-based weighting, sharpness, view synthesis.**

I. Introduction

THE increasing prevalence of 3D television and Free Viewpoint Television has drawn extensive interest in virtual viewpoint synthesis [1], [2]. Depth-Image-Based Rendering (DIBR) is a fundamental issue in view synthesis. Objective quality models for view synthesis are highly desired, which can be used for benchmarking DIBR methods and guiding algorithm design. While a lot of efforts have been made on DIBR algorithms, there is a lag behind on the corresponding quality evaluation of DIBR-synthesized views, which may restrict future development of DIBR approaches.

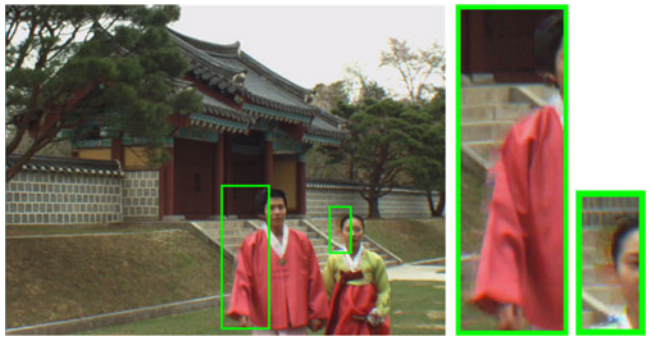

Fig. 1. Distortions in disoccluded regions of a DIBR-synthesized image.

Typically, DIBR consists of two stages, namely warping and rendering. In the warping stage, reference views are mapped to 3D Euclidean space with the guidance of depth information, followed by an inverse mapping from the 3D space to the target view. The warping operation introduces geometric displacements between the target and reference views, resulting in disoccluded regions. Rendering is then employed to restore the disoccluded regions, producing the final synthesized image. Due to the imperfect rendering, distortions are often present in the disoccluded regions, as shown in Fig. 1. Therefore, in the quality assessment of DIBR-synthesized images, local geometric distortions in disoccluded regions should be measured. In the literature, many quality metrics for 2D images have been reported [16], [39], [40], [42]. However, these metrics are all based on an assumption that the distorted and reference images are exactly aligned, i.e., no geometric displacements exist. Therefore, they are limited in handling DIBR-synthesized images.

In the literature, a number of works have been carried out towards the quality assessment of DIBR-based view synthesis [2]–[15], which are briefly reviewed here.

1) Bosc's method [2]. The difference between the reference and synthesized views was first computed. Then a threshold was employed to extract the critical areas. The quality score of the synthesized image was defined as

Manuscript received June 2, 2016; revised March 24, 2017, May 23, 2017, and September 3, 2017; accepted September 3, 2017. Date of publication October 6, 2017; date of current version March 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grants 61771473 and 61379143, in part by the Fundamental Research Funds for the Central Universities under Grant 2017XKQY084, and in part by the Qing Lan Project. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abdulmotaleb El Saddik. *(Corresponding author: Yu Zhou.)*

L. Li and Y. Zhou are with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: lileida@cumt.edu.cn; zhouy7476@cumt.edu.cn).

K. Gu is with 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).

W. Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2760062

the mean Structural SIMilarity (SSIM) [16] score in the critical areas.

2) Color and Sharpness of Edge Distortion (CSED) [3] measure. The luminance component in SSIM [16] was used to measure the color distortion in disoccluded regions. Then the sharpness of edge distortion was measured. These two aspects were combined to generate the final quality score.
3) 3D DIBR-based Video Quality Measure (3VQM) [4]. A distortion-free depth estimation method was first proposed. Then the temporal outliers/inconsistencies and spatial outliers were derived and integrated to measure the quality of DIBR-based videos.
4) View Synthesis Quality Assessment (VSQA) metric [5]. In VSQA, a 2D image quality metric was first employed to generate the similarity/distortion signal between the reference and synthesized images. Then three weighting maps were computed, including textural complexity, gradient orientations and image contrast. The final score was calculated by pooling the similarity/distortion map using the weighting maps.
5) Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) [6]. The reference and synthesized views were first decomposed using the morphological wavelet transform at several scales. Then the mean squared error values were computed for the detail subbands. Multiscale Wavelet Mean Squared Error (MW-MSE) was obtained by pooling the MSE values across scales. Finally, the MW-PSNR score was calculated based on MW-MSE. The reduced MW-PSNR (RMW-PSNR) was also proposed [6] by only using wavelet subbands from higher decomposition scales.
6) Morphological Pyramid Peak Signal-to-Noise Ratio (MP-PSNR) [7]. MP-PSNR was improved from MW-PSNR, where morphological wavelets were replaced by morphological pyramids. Similar to RMW-PSNR, reduced MP-PSNR (RMP-PSNR) was also defined by only using detail images from higher pyramid scales to calculate the quality score.
7) 3D Synthesized view Image quality Metric (3DSwIM) [9]. Both the reference and synthesized images were first partitioned into blocks. Registration was conducted through block matching. Then the matched blocks were decomposed by the Haar wavelet, and the histogram distributions were computed. Finally, the histogram distance was calculated as the overall quality score.
8) Critical Binocular Asymmetry (CBA) [10] metric. Critical areas were first detected using the synthesized left-, right- view images and the corresponding disparity maps. Then the average SSIM scores were computed only in the critical areas for both views. Finally, average pooling was conducted between the quality scores of left- and right- view images, producing the overall CBA score. CBA is mainly designed for characterizing the binocular asymmetry property of human eyes.
9) Synthesized Image Quality Evaluator (SIQE) [11]. SIQE was based on the cyclopean perception theory and divisive normalization transform. Divisive normalization was utilized to model the statistical characteristics of the original views, based on which the statistical model of the cyclopean image was obtained. The statistical model of the synthesized image was also estimated using the divisive normalization. Finally, the Bhattacharyya distance was calculated as the quality score.
10) DIBR-Synthesized image Quality Metric (DSQM) [12]. Block-wise matching was first conducted between the original image (left view or right view) and the DIBR-synthesized image. Then phase congruency features in the corresponding blocks were extracted and compared to measure the distortions in the synthesized blocks. The final quality score was obtained by averaging the results of all blocks.
11) Autoregression Plus Thresholding (APT) [13]. The idea is to detect the geometric distortions in DIBR process using autoregression-based image representation. Visual saliency is also incorporated to adapt to the characteristics of the human perception.
12) Ryu's method [14]. The impact of texture and depth on view synthesis was first analyzed. Then the quality model was proposed based on synthesis intolerance. This method focuses on depth distortions.
13) Reduced-Reference synthesized Virtual View Quality Metric (RR-VVQM) [15]. The characteristics of depth distortions were first analyzed. Then the correlations between the decoded depth and color images were employed to design the depth quality. The final RR-VVQM model was obtained by integrating all depth quality scores.

Among the aforementioned DIBR image quality metrics, models [2], [5]–[8] do not explicitly consider the geometric distortions, which is crucial for the evaluation of DIBR-based view synthesis. In 3DSwIM [9], displacement is handled by block matching. However, the matched blocks do not necessarily indicate the disoccluded regions. So it is not very effective for measuring local geometric distortions in disoccluded regions, which are the dominated distortions in synthesized images. While notable advances have been achieved in the aforementioned DIBR quality indices, new advanced models are still highly demanded for more effective quality evaluation of view synthesis.

The rendering process is mainly affected by geometric distortions and sharpness change [3], [6], [9]. With this inspiration, this work presents a new metric for the quality assessment of DIBR-synthesized images by measuring LOcal Geometric distortions and global Sharpness (LOGS). LOGS consists of three modules, including visual image generation, local disoccluded region detection and evaluation, and global sharpness evaluation. The first module is to adapt to the impact of viewing distance on quality perception. In the second module, SIFT-flow based warping is first proposed to detect the disoccluded regions. Then the size-weighted distortion strength is computed to measure the quality of disoccluded regions. In the third module, global sharpness is evaluated based on a proposed sharpness metric. The overall score is obtained by pooling the scores of geometric distortions and sharpness. Performance evaluation is

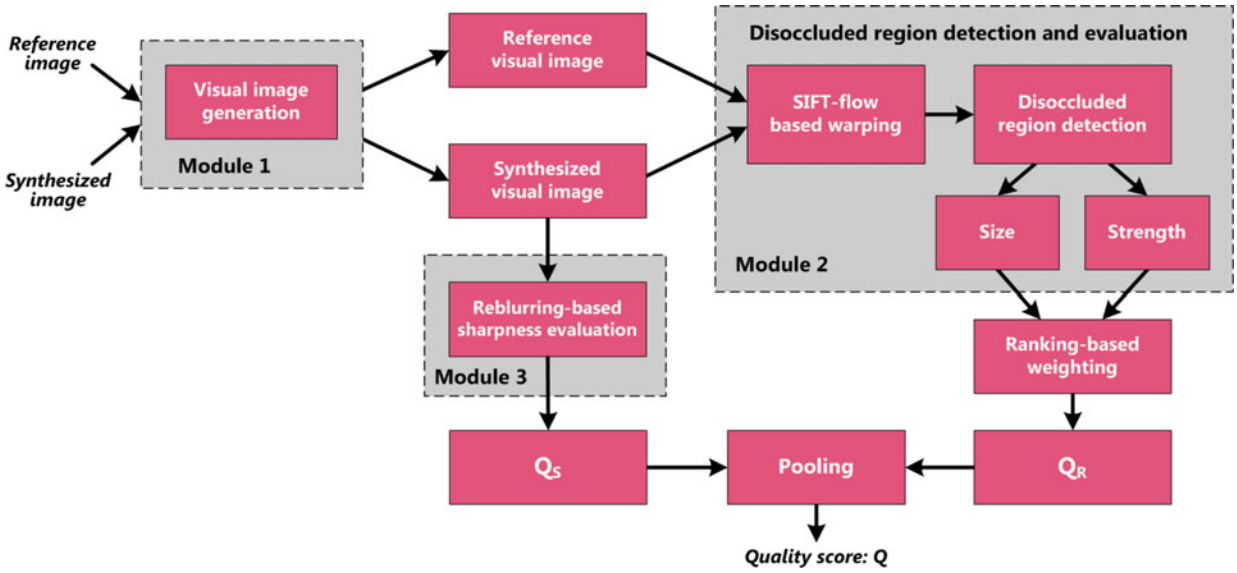

Fig. 2. Flowchart illustration of the proposed quality model for DIBR-synthesized images.

carried out on four DIBR image/video databases. The simulation results demonstrate that the proposed metric outperforms both traditional 2D and the existing DIBR quality indices. As applications, the proposed method is further used for improving existing image quality metrics and benchmarking DIBR algorithms, and very encouraging results are achieved.

II. Proposed View Synthesis Quality Metric

In DIBR, the quality of synthesized views are mainly influenced by geometric distortions in disoccluded regions and degradation of image sharpness, which are caused by warping and rendering. Based on these observations, the underlying idea of the proposed method is to evaluate local geometric distortions and global sharpness in synthesized images. The diagram of the proposed metric is illustrated in Fig. 2, which is composed of three modules, i.e., visual image generation, disoccluded region detection and evaluation, and global sharpness evaluation.

A. Visual Image Generation

The perceptual quality of an image depends on various factors. Apart from the distortion itself, viewing distance also has significant impact on the perceived quality [17]. In order to intuitively show this phenomenon, an example is illustrated in Fig. 3. In the figure, two patches from DIBR-synthesized images are viewed at different viewing distances. It is obvious that when images are viewed at a closer distance, the perceivable pixel size becomes larger [18]. In this case, the viewer can distinguish pixel differences more easily, and image local distortions can be easily perceived, resulting in poorer quality. On the contrary, with the increase of viewing distance, the perceivable pixel size becomes smaller, and it is more difficult for the viewer to identify local distortions.

The above phenomenon can be interpreted as follows. For different viewing distances, the images projected onto the retina are also different, resulting in different perceived quality [18].

Fig. 3. Examples of perceived quality by viewers at different viewing distances. With the increase of viewing distance, local distortions become less visible.

The human visual system has the capacity to adaptively adjust the resolution of an image when projected onto the retina [18]. In other words, the perceived image on the retina is actually a scaled version of the original one. For clarity, we use the term *visual image* to denote the image perceived by the retina. Further, we use *visual resolution* to denote the resolution of a visual image, and use *spatial resolution* to denote the resolution of the original image to be evaluated. It is not hard to understand that the visual resolution is determined by the spatial resolution and viewing distance simultaneously [19]. Since the HVS perceives image quality based on visual images, more accurate quality prediction can be expected if such visual images are generated for building the quality model. Towards this direction, recently several works have been done to investigate the impact of visual image generation (in terms of viewing distance and image

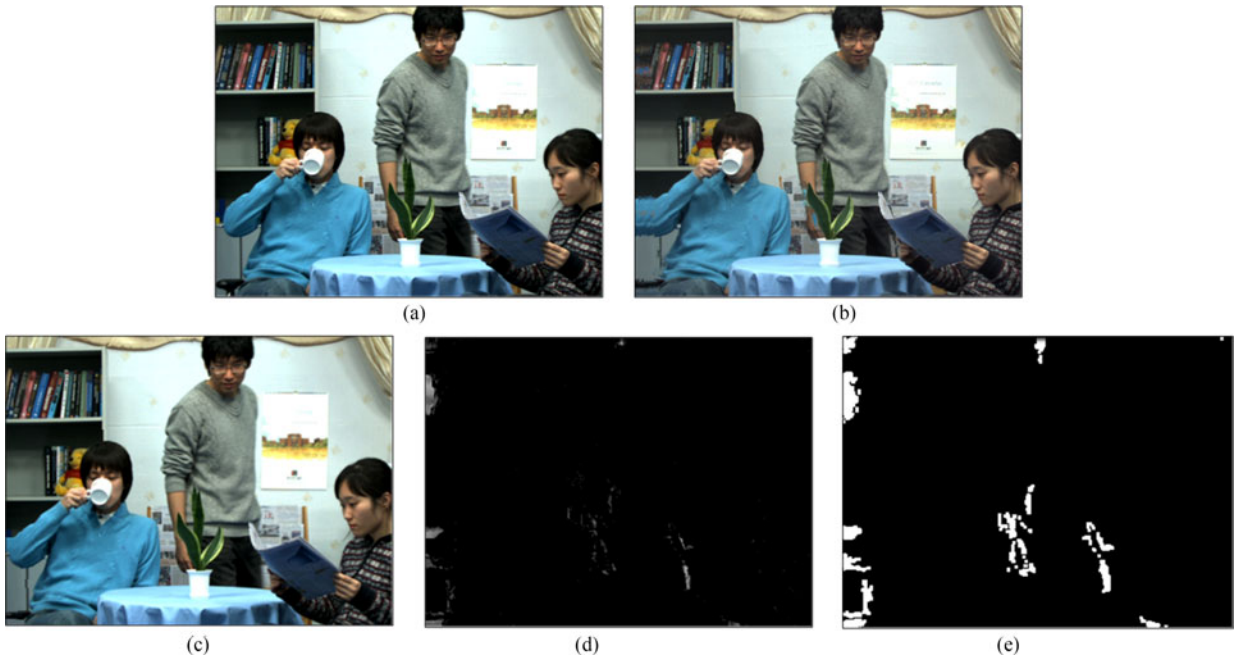

Fig. 4. Illustration of the disoccluded region detection process. (a) Original image, (b) Synthesized image, (c) Warped original image, (d) Difference map between (b) and (c), (e) Binary detection map after dilation operation.

resolution) on visual quality evaluation [17], [19]–[21]. In this paper, we employ the Self-Adaptive Scale Transform (SAST) model [20] to generate the visual images.

In the SAST model [20], the visual image is generated by resizing the original image by a scaling factor z, which is defined as:

$$z = \sqrt{\frac{\mathcal{A}}{\tilde{\mathcal{A}}}} = \sqrt{\frac{h \times w}{\tilde{h} \times \tilde{w}}} = \sqrt{\frac{1}{4 \tan\left(\frac{\theta_h}{2}\right) \times \tan\left(\frac{\theta_w}{2}\right)} \times \left(\frac{h}{d}\right)^2 \times r}, \quad (1)$$

where $\mathcal{A}$ and $\tilde{\mathcal{A}}$ represent the spatial resolution of the original image and visual resolution of the visual image, which are $h \times w$ and $\tilde{h} \times \tilde{w}$ respectively; θ_h and θ_w denote the view angles in horizontal and vertical directions; d is the viewing distance; and $r = w/h$. More details on the SAST model can be found in [20].

Upon obtaining the scaling factor z, the original reference and synthesized images are resized to produce the reference visual image and synthesized visual image, respectively. The subsequent quality evaluation stages are all based on the visual images. In our implementation, the bicubic interpolation algorithm is adopted in generating the visual images.

B. Disoccluded Region Detection and Evaluation

In DIBR, the warping operation usually causes disoccluded regions. Then, inpainting and interpolation techniques are employed to fill these regions, namely rendering. Both inpainting and interpolation share the similar principle that neighboring pixles are utilized to produce an approximation of a target pixel. During this process, artifacts occur in the disoccluded regions, which are the dominated distortions in synthesized images. In this part, we first propose a disoccluded region detection method. Then the quality of disoccluded regions is evaluated using size-weighted distortion strength.

1) Disoccluded Region Detection: The warping operation causes displacements between the reference and synthesized images [6]. To detect the disoccluded regions, we first align the geometric displacements. In this work, this is achieved using SIFT-flow [22]. SIFT-flow is a robust image matching method, which is based on the estimation of dense correspondence of SIFT descriptors between two images. The SIFT-flow map between the reference view $\mathbf{I}_{ref}$ and synthesized view $\mathbf{I}_{syn}$ is first estimated. Then the reference image $\mathbf{I}_{ref}$ is warped to the synthesized image $\mathbf{I}_{syn}$ with the guidance of the SIFT-flow map, producing the warped reference image $\mathbf{I}^w_{ref}$. Then the absolute difference between $\mathbf{I}_{syn}$ and $\mathbf{I}^w_{ref}$ is computed:

$$\mathbf{M} = \left|\mathbf{I}_{syn} - \mathbf{I}^w_{ref}\right|. \quad (2)$$

In order to exclude weak changes in the difference map that are not visible to human eyes, a threshold is employed to refine the difference map:

$$\mathbf{M}(x,y) = \begin{cases} 0, & \text{if } \mathbf{M}(x,y) < T, \\ \mathbf{M}(x,y), & \text{otherwise}, \end{cases} \quad (3)$$

where $x \in [1, M], y \in [1, N]$, and $M \times N$ represents the image resolution. In this work, T is empirically set to 50. Furthermore, with the consideration that small holes may be present in the detected disoccluded regions, image dilation operation is also conducted to obtain the final regions.

Fig. 4 intuitively illustrates the detection process with a pair of reference and synthesized images. Two detection maps are ob-

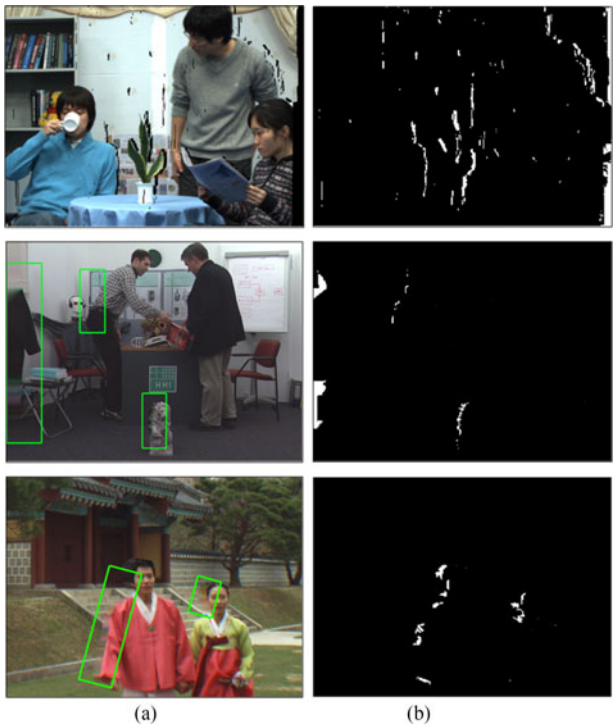

Fig. 5. Examples of disoccluded region detection. (a) Synthesized images. (b) Detected regions.

tained, including the original difference map and a binary detection map. In this work, the difference map is used to evaluate the distortion strength in disoccluded regions, and the binary map is used to compute the sizes of these regions. Fig. 5 further shows several examples of disoccluded region detection. It is observed that the detection result correctly captures the disoccluded regions. In our method, disoccluded regions are extracted based on the SIFT-flow method [22]. In our implementation, the average time to detect the disoccluded regions from a DIBR-synthesized image with size 1024×768 in the IRCCyN/IVC database [2] is 8.2 seconds, which is based on a computer with Intel i5 CPU at 3.2 GHz, 8 GB RAM, and Windows 7 operation system.

In (3), the empirical threshold T is set to 50. In implementation, we have found that by setting $T = 50$ the disoccluded region detection results are satisfactory for the common images. Although in the "Visual Image Generation" stage the original images are resized to generate the visual images, it does not affect the disoccluded region detection too much. Fig. 6 shows an example, where the detected regions from the original synthesized images and the corresponding visual images are given. It can be clearly seen from the figure that the resizing operation has little impact on the disoccluded region detection results. In implementation, the same conclusion has been achieved for images with different visual contents. Therefore, the empirical threshold is applicable to all conditions.

2) Quality Evaluation of Disoccluded Regions: It is obvious that a larger disoccluded region with heavier distortion is more annoying to human eyes. With this consideration, we evaluate the quality of disoccluded regions from two aspects, i.e., region size and distortion strength.

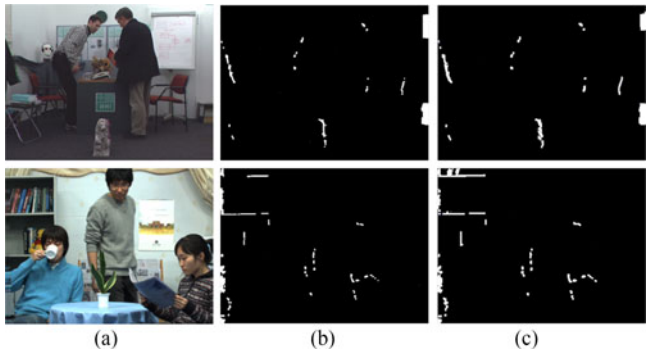

Fig. 6. Detected disoccluded regions with threshold $T = 50$. (a) Synthesized images. (b) Detected regions from the original synthesized image. (c) Detected regions from the corresponding visual images.

Region size: In practice, a synthesized image may contain multiple isolated disoccluded regions. In this work, we first conduct a segmentation to obtain individual disoccluded regions, based on which we estimate the overall quality.

With the binary detection map, this can be done easily using Matlab function *bwboundaries*. The disoccluded regions are denoted by $\{\mathbf{\Omega}_i, i = 1, 2, \cdots, K\}$, where K is the total number of disoccluded regions. In this paper, the size of a disoccluded region is defined as the number of the pixels within it. Here, the sizes of disoccluded regions are denoted by $\{s_i, i = 1, 2, \cdots, K\}$.

Distortion strength: For different DIBR algorithms, the distortion levels in disoccluded regions are different, depending on both image content and rendering method. In this paper, the distortion strength of a specific disoccluded region is defined as the mean value of the region in the whole difference map $\mathbf{M}$:

$$e_i = \frac{1}{s_i} \sum_{(x,y)\in\mathbf{\Omega}_i} \mathbf{M}(x, y), \tag{4}$$

where $i = 1, 2, \cdots, K$.

Quality of disoccluded regions: With the aim to obtain an overall score for disoccluded regions, we propose to use ranking-based weighting [23]. Specifically, region sizes are used to generate weights, which are then adopted to weight the distortion strength values.

The sizes of the disoccluded regions are first sorted in ascending order. Then the weight of a disoccluded region is defined as:

$$w_i = \log_2 \left(1 + \frac{\mathrm{Rank}_i}{K}\right), i = 1, 2, \cdots, K, \tag{5}$$

where Rank_i denotes the ranking index of the ith region. Since $\mathrm{Rank}_i \in \{1, 2, \cdots, K\}$, it is easy to know that $w_i \in (0, 1]$, and a larger region will be assigned a bigger weight.

The quality score of disoccluded regions is calculated as:

$$Q_{\mathrm{R}} = \frac{\sum_{i=1}^{K} w_i e_i}{\sum_{i=1}^{K} w_i}. \tag{6}$$

A high $\mathrm{Q_R}$ value indicates a low quality image with heavy distortions in disoccluded regions.

C. Global Sharpness Evaluation

In additional to the distortions in local disoccluded regions, the warping and rendering in DIBR also cause blurring effect in the synthesized images, especially around the transitions of background/foreground and disoccluded regions [6], [9]. Here, we propose a new method for evaluating global sharpness, which is achieved by measuring the textural distance between a synthesized view and its intentionally-blurred version. The proposed method can be simply explained by an intuitive observation that the extent of similarity between an image and its intentionally-blurred version is mostly influenced by its original sharpness. Furthermore, a general brain theory called "free energy" reveals that, given an input image signal, it will be first separated into predicted orderly part and unpredicted disorderly part by human brain. In [24], [25], it was found that the latter unpredicted part has a strong relevance with and has been successfully deployed to image sharpness assessment and more. From the perspective of practical applications, several recent works concerning multimedia signal processing were developed based on the above-described "intentionally-blurring" strategy for enhancing images [26] or highlighting possible interesting regions [27].

1) Reblurring: A synthesized image $\mathbf{I}_{syn}$ is intentionally blurred based on a Gaussian smoothing filter:

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}, \tag{7}$$

where σ denotes the standard deviation. In this work, a Gaussian filter of size 3×3 and $\sigma=5$ is employed. The intentionally-blurred version of a synthesized image is denoted by $\mathbf{I}'_{syn}$.

2) Sharpness Evaluation: The sharpness of a synthesized image is measured block-wisely. First, a synthesized image $\mathbf{I}_{syn}$ and its reblurred version $\mathbf{I}'_{syn}$ are partitioned into non-overlapping blocks of size $k\times k$, with a total number of Z blocks, where

$$Z=\left\lfloor\frac{M}{k}\right\rfloor\cdot\left\lfloor\frac{N}{k}\right\rfloor. \tag{8}$$

The textural complexity of a block is represented by its variance σ^2 [28]. Then the overall sharpness score of a synthesized image is defined as the average textural distance of all blocks:

$$Q_{\mathrm{S}}=\frac{\sum_{i=1}^{i=Z}\sqrt{|\sigma_{1i}^2-\sigma_{2i}^2|}}{Z}, \tag{9}$$

where σ_{1i}^2 and σ_{2i}^2 represent the standard deviations of the ith blocks in the synthesized image and the intentionally-blurred version, respectively. A high Q_{S} value indicates a high quality image with good sharpness.

It should be noted that the global sharpness evaluation module in the proposed method is not unique. Besides the proposed reblurring-based approach, edge/gradient representation can also be explored for evaluating image sharpness. For example, we can first detect image edges based on the commonly used edge operators, such as Sobel, Roberts, Prewitt and Laplacian of Gaussian (LoG). Then we can use the average edge width to represent the perceived sharpness, since blurring typically causes spread of image edges [29]. In addition, the exiting NR image sharpness metrics can also be adopted in the proposed quality model. However, after extensive experiments and comparisons we find that the proposed reblurring-based approach achieves the best performance, which will be detailed in the experiment section.

D. Pooling

With the quality scores of local disoccluded regions (Q_{R}) and global sharpness (Q_{S}), it is desirable to integrate them to produce an overall score. Q_{R} is a distortion measure, which means that a high Q_{R} value indicates that the distortion in disoccluded regions is severe, and thus image quality is bad. By contrast, the score Q_{S} is a quality measure, which means that a high Q_{S} value indicates that the global sharpness of the synthesized image is good, and thus image quality is good. With these considerations, we propose the following simple pooling strategy to define the final quality score of a synthesized image:

$$Q=\frac{Q_{\mathrm{S}}^{\alpha}}{Q_{\mathrm{R}}^{\beta}+c}, \tag{10}$$

where the parameters α and β are used to balance the relative contributions of local disoccluded regions and global sharpness, and c is a very small constant, which is employed to avoid numerical instability. In implementation, we empirically set $\alpha=1.5$, $\beta=2.3$, and $c=1\times 10^{-6}$. A higher Q value indicates a better quality.

III. Experimental Results

A. Evaluation Protocols

Four DIBR-synthesized image/video databases are adopted for performance evaluation, including three DIBR image databases IRCCyN/IVC [2], IVY [10] and MCL-3D [30], and one DIBR video database SIAT-3D [31]. These databases consist of reference views and the corresponding synthesized views obtained by different DIBR algorithms, together with the associated subjective scores.

1) IRCCyN/IVC [2]. This dataset consists of 12 reference images and 84 synthesized ones by 7 different DIBR algorithms, which are denoted by A1-A6 [32]–[37] and A7 (warping without rendering). In the subjective experiment, Absolute Category Rating-Hidden Reference (ACR-HR) is utilized, and the subjective scores are represented using Mean Opinion Score (MOS). Similar to [9], in order to use the subjective scores for full-reference quality assessment, the Difference Mean Opinion Scores (DMOS) are first generated based on the MOS values [38].
2) IVY [10]. The IVY database consists of 84 stereo images, which are obtained from 7 reference image sets. For each set, 12 synthesized image pairs are generated from 3 views using 4 DIBR view synthesis algorithms, respectively. The subjective experiment is conducted using the Double Stimulus Continuous Quality Scale (DSCQS) method, and the subjective scores are represented by DMOS. It should be noted that in IVY database, each synthesized view has two reference images, i.e., left- and right- views.

So in our experiment, we first compute the quality scores for both views. Then the average value is calculated as the final quality score.

3) MCL-3D [30]. This database consists of 693 stereoscopic image pairs from 9 image-plus-depth sources. The image resolution is either 1024×768 or 1920×1080. Six kinds of common distortions are added to the texture/depth image before rendering. Distortions introduced by the rendering operation are also considered, which are produced by four different rendering algorithms. Pair-wise comparison was used in the subjective experiment and MOS was calculated as the subjective score. Similar to the IVY database, in experiment we employ the mean values of the left- and right- views as the final score.
4) SIAT-3D [31]. It contains 140 DIBR-synthesized videos with resolution 1024×768 or 1920×1088 from ten sequences. For each sequence, 14 combinations of texture/depth quantizations were injected, which were generated to simulate the compression distortions. DMOS was calculated as the subjective score. For this database, we first compute the quality scores frame by frame. Then the average value is calculated as the final score.

Although the four databases are all related to DIBR, they are somewhat different in terms of the distortion characteristics. The IRCCyN/IVC database [2] contains images generated by different DIBR algorithms using uncompressed texture and depth images. So the distortions are mainly caused by the imperfect rendering, which is the main objective of the proposed LOGS metric. The IVY database [10] mainly contains distortions caused by the asymmetry in left and right views. By comparison, MCL-3D [30] and SIAT-3D [31] databases contain more practical distortions, including both compression artifacts from texture/depth coding and different DIBR algorithms.

For performance evaluation, four popular criteria are employed, namely Pearson Linear Correlation Coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC), Kendall's Rank Correlation Coefficient (KRCC) and RMSE. To be specific, PLCC and RMSE are calculated to evaluate the prediction accuracy, while SRCC and KRCC are employed for measuring prediction monotonicity. They are computed following a five-parameter nonlinear mapping:

$$f(x) = \tau_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2 (x - \tau_3)}} \right) + \tau_4 x + \tau_5, \qquad (11)$$

where x denotes the predicted socre, $f(x)$ denotes the corresponding subjective score, and $\tau_1, \tau_2, \cdots, \tau_5$ are the fitting parameters.

B. Parameter Selection

Before conducting comprehensive experiments, several parameters should be determined, including the parameters of the Gaussian filter (filter size and standard deviation σ in (7)), block size in sharpness evaluation, and the weights α, β in the final pooling. In this work, these parameters are determined based on the IRCCyN/IVC database. It should be noted that the proposed method is learning-free, so it does not involve a training stage. What we do here is to determine the optimal model parameters. Once determined, they are directly used in the subsequent experiments for all four databases.

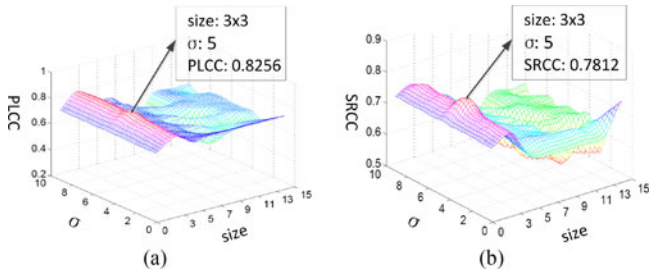

Fig. 7. Impact of filter size and standard deviation of the Gaussian filter on the performance of the proposed metric on IRCCyN/IVC database. (a) PLCC, (b) SRCC.

TABLE I
IMPACT OF BLOCK SIZE $k \times k$ ON THE PERFORMANCE OF THE PROPOSED METRIC IN IRCCYN/IVC DATABASE

Block size	PLCC	SRCC	KRCC	RMSE
4×4	0.8109	0.7691	0.5733	0.3862
8×8	**0.8256**	**0.7812**	**0.6071**	**0.3601**
16×16	0.8201	0.7793	0.6002	0.3654
32×32	0.8199	0.7736	0.5984	0.3680
64×64	0.8146	0.7705	0.5831	0.3749

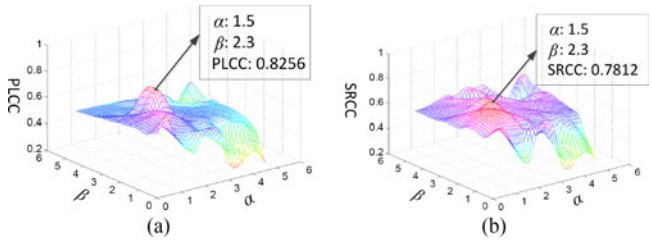

Fig. 8. Impact of different weights in the pooling on the performance of the proposed metric on IRCCyN/IVC database. (a) PLCC, (b) SRCC.

Fig. 7 illustrates the performance of our metric with different settings of the Gaussian filter. It is observed that the standard deviation does not influence the performance too much. By contrast, filter size has more impact on the overall performance. Our metric achieves the best performance when the filter size is 3×3 and $\sigma = 5$. Therefore, these values are adopted in our experiments.

Table I lists the performance values of the proposed metric with different block sizes in global sharpness evaluation. It can be seen that block size has little influence on the results. Slightly better result can be achieved when the block size is 8×8. Therefore, we use block size 8×8 in this paper.

Fig. 8 shows the performance of the proposed metric on IRCCyN/IVC database when different combinations of α and β are used. It is clear that the proposed metric delivers the best performance when $\alpha = 1.5$ and $\beta = 2.3$. Therefore, these parameters are adopted.

TABLE II
EXPERIMENTAL RESULTS OF TRADITIONAL 2D AND DIBR QUALITY INDICES ON IRCCYN/IVC AND IVY DIBR IMAGE DATABASES

Metric	Type	IRCCyN/IVC Database [2]				IVY Database [10]			
		PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
PSNR	2D	0.5098	0.4628	0.3321	0.5728	0.6518	0.6515	0.4739	18.6967
SSIM [16]	2D	0.5644	0.5245	0.3650	0.5496	0.5684	0.5662	0.4068	20.2826
FSIM [39]	2D	0.5828	0.4161	0.2738	0.5410	0.6118	0.5975	0.4223	19.4998
GSM [40]	2D	0.5246	0.4215	0.2750	0.5668	0.5736	0.5805	0.4068	20.1934
LTG [41]	2D	0.5311	0.4139	0.2900	0.5642	0.6218	0.6077	0.4337	19.3059
ADD-SSIM [23]	2D	0.6470	0.5611	0.4141	0.5077	0.7012	0.7080	0.5425	16.4119
Bosc [2]	DIBR	0.5843	0.4905	0.3414	0.5403	0.6196	0.6046	0.4246	19.3497
VSQA [5]	DIBR	0.5742	0.5233	0.3673	0.5451	0.5012	0.5034	0.3118	22.5648
MW-PSNR [6]	DIBR	0.5622	0.5757	0.4378	0.5506	0.5240	0.5051	0.3528	20.9969
RMW-PSNR [6]	DIBR	0.5744	0.6245	0.4960	0.5450	0.5224	0.5008	0.3540	21.0200
MP-PSNR [7]	DIBR	0.6174	0.6227	0.4833	0.5238	0.5842	0.5707	0.4005	20.0081
RMP-PSNR [8]	DIBR	0.6772	0.6634	0.5382	0.4899	0.5310	0.5033	0.3471	20.8894
3DSwIM [9]	DIBR	0.6584	0.6156	0.4496	0.5011	0.3338	0.3374	0.2249	23.2375
SIQE [11]	DIBR	0.5284	0.4492	0.3269	0.5653	0.3855	0.3764	0.2524	22.7462
DSQM [12]	DIBR	0.7650	0.7067	0.5382	0.4288	0.7003	0.6915	0.5181	17.5734
Proposed LOGS	DIBR	**0.8256**	**0.7812**	**0.6071**	**0.3601**	**0.7313**	**0.7256**	**0.5865**	**10.1250**

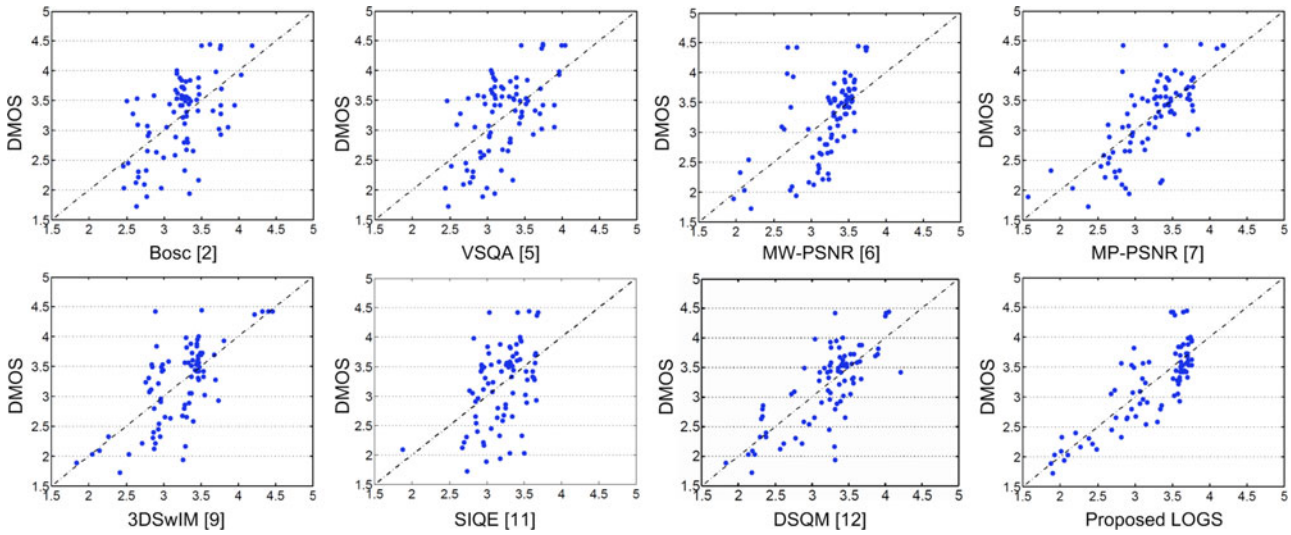

Fig. 9. Scatter plots of eight quality indices designed for view synthesis on IRCCyN/IVC database.

C. Performance Evaluation

In this subsection, we evaluate the performance of our method comprehensively and compared with the state-of-the-art quality models. The compared traditional 2D quality metrics include PSNR, SSIM [16], FSIM [39], GSM [40], LTG [41] and ADD-SSIM [23], while the existing DIBR quality metrics include the models reported in [2], [5]–[9], [11], [12], which have been introduced in the introduction. Table II summarizes the experimental results on IRCCyN/IVC and IVY databases, where the best performance values are marked in boldface.

It is observed from Table II that in both databases, the 2D quality metrics and the DIBR quality metrics are not that effective in evaluating the quality of DIBR-based view synthesis. In IRCCyN/IVC, the best performance of 2D metric is achieved by ADD-SSIM [23]. However, the PLCC, SRCC and KRCC values are only 0.6470, 0.5611 and 0.4141, which are not satisfactory. Compared to 2D metrics, the current DIBR quality metrics achieve notable performance improvements. Among them, the best performance is delivered by DSQM [12]. By comparison, the proposed method delivers the best performances, and it outperforms the existing metrics significantly. In IVY database, similar results have been observed. This demonstrates the advantages of the proposed metric.

Fig. 9 further shows the scatter plots of the subjective versus the objectives scores generated by different DIBR quality indices on IRCCyN/IVC database. It is clear in the figure that for the proposed method, the points are gathered more closely to the curve. This indicates that the predicted scores are more consistent with ground truth.

Table III further lists the results of the DIBR quality metrics on MCL-3D [30] and SIAT-3D [31], two more practical and challenging databases. It is known that the proposed metric still produces the best performance on SIAT-3D database, which is significantly better than the other DIBR quality metrics. For the MCL-3D database, RMW-PSNR [8] achieves the best

TABLE III
PERFORMANCES OF DIBR QUALITY METRICS ON MCL-3D AND SIAT-3D DATABASES

Metric	Type	MCL-3D [30]				SIAT-3D [31]			
		PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
Bosc [2]	DIBR	0.4453	0.4079	0.3044	2.3295	0.6067	0.5872	0.4153	0.1020
VSQA [5]	DIBR	0.5078	0.5120	0.3917	2.9175	0.4012	0.4136	0.3589	0.1152
MW-PSNR [6]	DIBR	0.8012	0.8099	0.6063	1.5568	0.5745	0.5545	0.3875	0.1050
RMW-PSNR [6]	DIBR	**0.8239**	**0.8308**	**0.6299**	**1.4743**	0.5757	0.5101	0.3595	0.1049
MP-PSNR [7]	DIBR	0.8169	0.8231	0.6206	1.5007	0.5681	0.5289	0.3727	0.1056
RMP-PSNR [8]	DIBR	0.8173	0.8250	0.6224	1.4993	0.5640	0.5318	0.3731	0.1059
3DSwIM [9]	DIBR	0.6519	0.5683	0.4088	1.9729	0.5677	0.5462	0.3798	0.1053
SIQE [11]	DIBR	0.6734	0.6976	0.5004	1.9233	0.3627	0.3041	0.2089	0.1195
DSQM [12]	DIBR	0.6995	0.6980	0.5086	1.8593	0.4001	0.3994	0.2864	0.1195
Proposed LOGS	DIBR	0.7612	0.7577	0.5629	1.6874	**0.6481**	**0.6342**	**0.4711**	**0.0972**

TABLE IV
PERFORMANCES OF THE PROPOSED METRIC WITH DIFFERENT VISUAL IMAGE GENERATION METHODS IN IRCCYN/IVC DATABASE

Method	PLCC	SRCC	KRCC	RMSE
No preprocessing	0.7321	0.7156	0.4823	0.4625
OSS [19]	0.5408	0.5505	0.4120	0.5377
AHC [21]	0.5898	0.5602	0.4164	0.5601
SAST [20]	**0.8256**	**0.7812**	**0.6071**	**0.3601**

TABLE V
PERFORMANCES OF THE PROPOSED METRIC WITH DIFFERENT SHARPNESS ASSESSMENT METHODS IN IRCCYN/IVC DATABASE

Sharpness Module	PLCC	SRCC	KRCC	RMSE
Marziliano [29]	0.7899	0.7586	0.5123	0.4211
JNB [43]	0.7985	0.7626	0.5450	0.4021
CPBD [44]	0.7742	0.7403	0.5238	0.4452
S3 [45]	0.7863	0.7612	0.5506	0.4266
FISH [46]	0.7933	0.7633	0.5410	0.4111
LPC [47]	0.7842	0.7605	0.5241	0.4278
MLV [48]	0.7933	0.7612	0.5326	0.4186
SVC [49]	0.7875	0.7601	0.5124	0.4250
BIBLE [50]	0.7996	0.7698	0.5612	0.3997
ARISM [51]	0.8021	0.7793	0.5762	0.3897
SPARISH [52]	0.7902	0.7781	0.5634	0.4005
Proposed (edge width with Sobel)	0.7705	0.7502	0.5090	0.4631
Proposed (edge width with Roberts)	0.7839	0.7610	0.5171	0.4281
Proposed (edge width with Prewitt)	0.7781	0.7486	0.5064	0.4402
Proposed (edge width with LoG)	0.7806	0.7553	0.5101	0.4390
Proposed	**0.8256**	**0.7812**	**0.6071**	**0.3601**

performance. This is mainly due to the fact that the distortions in MCL-3D database mainly distribute evenly in the whole image. Therefore, the PSNR-induced metrics, e.g., MW-PSNR, RMW-PSNR, MP-PSNR and RMP-PSNR, deliver better performances than the other metrics. In MCL-3D, the performance of the proposed metric is only inferior to these PSNR-induced metrics, and it is much better than those of other metrics. Based on the results in Tables II and III, we know that the proposed LOGS metric delivers the best overall performance.

D. Evaluation of Visual Image Generation Methods

In this paper, the SAST [20] method is used to generate the visual images in the first module. In practice, several alternative approaches have been proposed for quality assessment. Here, we investigate the performance of our metric using another two approaches, namely Optimal Scale Selection (OSS) [19] and Adaptive High-frequency Clipping (AHC) [21]. Table IV summarizes the experimental results. In the table, 'No preprocessing' represents the presented model without integrating the visual image generation module.

It is known from Table IV that our metric achieves the best performance when the SAST method is used. Using OSS [19] or AHC [21] in the proposed method does not bring performance improvements. This may be due to the fact that the spatial domain-based SAST method is able to highlight disoccluded regions via appropriately downsampling the images to be processed [23]. In comparison, wavelet decomposition-based AHC and OSS metrics are more designed and fine-tuned for typical types of distortions such as additive noise, Gaussian blur and JPEG compression, but not for the disoccluded areas caused by geometric distortions. Hence, both OSS and AHC implement remarkably worse than the SAST technique used in this paper. It is also known from the table that very promising results are obtained even if the first module is not used, which are already better than most of the other metrics. This further demonstrates the effectiveness of the features used in the proposed metric.

E. Evaluation of Sharpness Evaluation Approaches

Sharpness is a key issue in DIBR-based view synthesis quality evaluation. So far, several effective image sharpness evaluation models have been reported. In this subsection, we test the performance of the proposed method when different sharpness metrics are integrated in the proposed method, including Marziliano [29], JNB [43], CPBD [44], S3 [45], FISH [46], LPC [47], MLV [48], SVC [49], BIBLE [50], ARISM [51] and SPARISH [52]. Furthermore, we also include several edge width based metrics as aforementioned in Section II-C. Specifically, four commonly used edge detection operators are investigated, including Sobel, Roberts, Prewitt and LoG. Table V summarizes the simulation results.

It is observed from Table V that when these sharpness metrics are adopted in the proposed method, the performances are

TABLE VI
PERFORMANCES OF DIFFERENT COMPONENTS IN THE PROPOSED METHOD IN IRCCYN/IVC DATABASE

Component	PLCC	SRCC	KRCC	RMSE
Region	0.7612	0.6948	0.4752	0.4839
Sharpness	0.5218	0.4231	0.3114	0.5678
All	**0.8256**	**0.7812**	**0.6071**	**0.3601**

TABLE VII
AVERAGE TIME FOR PREDICTING THE QUALITY SCORE OF AN 1024×768 IMAGE IN IRCCYN/IVC DATABASE BY DIFFERENT DIBR IMAGE QUALITY METRICS

Metric	Time (s)	Metric	Time (s)
Bosc [2]	0.13	RMP-PSNR [8]	0.51
VSQA [5]	0.87	3DSwIM [9]	8.70
MW-PSNR [6]	0.09	SIQE [11]	0.74
RMW-PSNR [6]	0.09	DSQM [12]	4.36
MP-PSNR [7]	1.29	LOGS	9.51

still very promising, and they already outperform most of the existing 2D and DIBR quality metrics, which can be observed from Table I. When the edge-based sharpness metrics are utilized, the performances are only comparable. By comparison, the proposed reblurring-based sharpness module produces the best performance.

F. Evaluation of Components in the Proposed Metric

The proposed metric evaluates the quality of DIBR-synthesized images from two aspects, i.e., quality of local disoccluded regions and quality of global sharpness. In order to investigate their relative importance in the proposed metric, we test each component separately in the IRCCyN/IVC database. Table VI summarizes the relevant experimental results.

It is observed from the table that both components achieve very promising results when they are used separately. The quality of disoccluded regions alone outperforms most of the existing techniques, which is very encouraging. This further confirms that distortions in the local disoccluded regions are important and should be measured properly. In contrast, sharpness is relatively less important in the overall performance. By combining them, the best performance can be achieved, which is significantly better than other metrics. This further demonstrates the rationality of the proposed metric.

G. Computational Complexity

In this part, we compare the computational complexities of the DIBR-based view synthesis quality models. Specifically, the average time for predicting the scores of DIBR-synthesized images with resolution 1024×768 in IRCCyN/IVC database [2] is calculated to measure the computation complexity. In implementation, we used a computer with Intel i5 CPU at 3.2 GHz, 8 GB RAM, and Windows 7 OS. Table VII summarizes the experimental results of ten DIBR quality metrics.

It is known from Table VII that most of the metrics are fast to compute. By contrast, 3DSwIM [9] and the proposed LOGS are more computationally demanding. This mainly attributes to the registration operation between the reference and synthesized images in both methods. In 3DSwIM [9], block-based registration was employed to guarantee that the best matching blocks are compared. In the proposed metric, the SIFT-flow-based matching and warping is utilized to correct the geometric displacements between the reference and synthesized images, based on which the disoccluded regions are detected. In implementation, this process is computationally expensive and takes up most of the time. In the future, we will try to explore more efficient registration techniques to reduce the computational cost of the proposed method.

IV. APPLICATIONS

A. Improving Existing Quality Metrics for DIBR Images

The proposed LOGS metric can be used as an universal module to improve the existing quality metrics for DIBR-synthesized images. To investigate the effectiveness of LOGS as a performance-boosting module, we integrate LOGS into the existing quality metrics (both traditional 2D and DIBR) and test their performances on IRCCyN/IVC database. Specifically, the new score of an improved metric is defined as:

$$Q_{\mathrm{improved}} = Q_{\mathrm{x}} \times Q_{\mathrm{LOGS}}, \tag{12}$$

where x denotes a specific quality metric, such as SSIM, FSIM, DSQM; Q_{improved}, Q_{x} and Q_{LOGS} denote the quality scores of the improved metric, original metric and LOGS, respectively. Table VIII summarizes the performances of the improved metrics, together with a statistic of the performance gains relative to their original values in percentage. Here, the best results and highest performance gains are marked.

It is easily observed from Table VIII that notable improvements can be achieved for all the tested quality metrics, regardless of 2D metrics and DIBR metrics. The biggest performance gains in terms of PLCC, SRCC and KRCC are 55.92%, 83.72% and 109.20%, respectively, which are very impressive. The best performance after incorporating the LOGS module is achieved by DSQM [12], which outperforms all the existing metrics. This further demonstrates that the features used in the proposed method can efficiently capture the DIBR-related distortions in the synthesized images.

B. Benchmarking DIBR Algorithms

In practical view synthesis applications, a natural issue is how to select a DIBR algorithm that can produce the best synthesized images. Although this can be done by subjective test, it is labor-expensive and cannot be embedded into real-time systems. Our metric can provide an alternative way. Generally, a better DIBR algorithm produces synthesized images with better quality, and corresponds to high quality scores. In this part, we benchmark the DIBR algorithms using both subjective and objective scores. In IRCCyN/IVC database, there are three image sequences, and each sequence has four synthesized images. For a specific DIBR

TABLE VIII
PERFORMANCES OF THE EXISTING IMAGE QUALITY METRICS AFTER INCORPORATING LOGS, TOGETHER WITH STATISTICS OF THE PERFORMANCE IMPROVEMENTS (IN PERCENTAGE)

Metric	Type	PLCC		SRCC		KRCC		RMSE	
		Value	Gain	Value	Gain	Value	Gain	Value	Gain
LOGS*PSNR	2D	0.7949	↑ **55.92%**	0.7522	↑ 62.53%	0.5670	↑ 70.73%	0.4040	↓ 29.47%
LOGS*SSIM	2D	0.7992	↑ 41.60%	0.7549	↑ 43.93%	0.5670	↑ 55.34%	0.4002	↓ 27.18%
LOGS*FSIM	2D	0.8064	↑ 38.37%	0.7589	↑ 82.38%	0.5728	↑ **109.20%**	0.3937	↓ 27.23%
LOGS*GSM	2D	0.8102	↑ 54.44%	0.7588	↑ 80.02%	0.5682	↑ 106.62%	0.3903	↓ **31.14%**
LOGS*LTG	2D	0.8101	↑ 52.53%	0.7603	↑ **83.72%**	0.5692	↑ 96.31%	0.3904	↓ 30.80%
LOGS*ADD-SSIM	2D	0.8101	↑ 25.22%	0.7604	↑ 35.52%	0.5693	↑ 37.48%	0.3903	↓ 23.12%
LOGS*Bosc	DIBR	0.7006	↑ 19.90%	0.7032	↑ 43.36%	0.5018	↑ 46.98%	0.4513	↓ 16.66%
LOGS*VSQA	DIBR	0.6901	↑ 20.18%	0.6733	↑ 28.66%	0.5314	↑ 44.68%	0.4903	↓ 10.05%
LOGS*MW-PSNR	DIBR	0.7888	↑ 40.31%	0.7381	↑ 28.21%	0.5480	↑ 25.17%	0.4092	↓ 25.68%
LOGS*RMW-PSNR	DIBR	0.7866	↑ 36.94%	0.7328	↑ 17.34%	0.5411	↑ 9.09%	0.4111	↓ 24.57%
LOGS*MP-PSNR	DIBR	0.7941	↑ 28.62%	0.7445	↑ 19.56%	0.5572	↑ 15.29%	0.4047	↓ 22.74%
LOGS*RMP-PSNR	DIBR	0.7952	↑ 17.42%	0.7459	↑ 12.44%	0.5578	↑ 3.64%	0.4037	↓ 17.60%
LOGS*3DSWIM	DIBR	0.7780	↑ 18.17%	0.7209	↑ 17.11%	0.5451	↑ 21.24%	0.4183	↓ 16.52%
LOGS*SIQE	DIBR	0.6542	↑ 23.81%	0.5789	↑ 28.87%	0.5225	↑ 59.83%	0.4905	↓ 13.23%
LOGS*DSQM	DIBR	**0.8280**	↑ 8.24%	**0.8157**	↑ 15.42%	**0.6293**	↑ 16.93%	**0.3533**	↓ 17.61%

The best results and highest performance gains are marked in boldface.

TABLE IX
RANKING OF SEVEN DIBR ALGORITHMS ACCORDING TO DMOS VALUES AND PREDICTED SCORES OBTAINED BY TEN DIBR-ORIENTED METRICS IN THREE SEQUENCES

Sequence	DIBR	DMOS	Bosc [2]	MP-PSNR [7]	MW-PSNR [6]	VSQA [5]	3DSwIM [9]	RMW-PSNR [6]	RMP-PSNR [8]	SIQE [11]	DSQM [12]	LOGS
Book arrival	A1 [32]	1	**2**	**6**	**7**	1	1	**7**	**6**	**2**	1	1
	A2 [33]	2	**3**	**5**	**5**	**3**	2	**5**	**5**	**4**	**4**	2
	A5 [36]	3	**4**	**2**	**1**	**4**	3	**1**	**2**	**6**	3	3
	A6 [37]	4	**6**	**3**	**3**	**2**	**5**	**3**	**3**	**7**	**5**	4
	A4 [35]	5	5	**1**	**2**	5	**4**	**2**	**1**	5	**6**	5
	A3 [34]	6	**1**	**4**	**4**	6	6	**4**	**4**	**3**	**2**	6
	A7	7	7	7	**6**	7	7	**6**	7	**1**	7	7
Lovebird	A4 [35]	1	1	1	1	1	1	1	1	6	**3**	1
	A5 [36]	2	**5**	2	**3**	**3**	**3**	**3**	**3**	5	**4**	2
	A6 [37]	3	**4**	3	**2**	**2**	**2**	**2**	**2**	7	**5**	3
	A2 [33]	4	**7**	**5**	**5**	**5**	4	**5**	4	4	**2**	**6**
	A1 [32]	5	**2**	**6**	**6**	**7**	5	**6**	**6**	1	**6**	5
	A3 [34]	6	6	**4**	**4**	**4**	6	**4**	**5**	3	**1**	**4**
	A7	7	**3**	7	7	**6**	7	7	7	2	7	7
Newspaper	A1 [32]	1	**7**	**6**	**6**	1	1	**6**	**6**	**4**	**4**	1
	A3 [34]	2	**3**	**3**	**3**	2	2	**3**	**3**	**6**	**3**	2
	A5 [36]	3	**1**	**2**	**2**	**6**	3	**2**	**2**	**2**	**1**	3
	A4 [35]	4	4	**1**	**1**	**3**	4	**1**	**1**	**3**	**5**	4
	A2 [33]	5	**2**	**4**	5	5	**6**	5	5	5	**6**	5
	A6 [37]	6	6	**5**	**4**	**4**	**5**	**4**	**4**	**1**	**2**	6
	A7	7	**5**	7	7	7	7	7	7	7	7	7

Ranking "1" represents the best performance and "7" means the worst performance.

algorithm, the mean DMOS value of the four images in a sequence is used to measure its performance, which is used as ground truth. At the same time, the mean objective score is also computed. Then the consistency between the subjective rankings and objective rankings is estimated. A good metric is expected to produce performance rankings consistent with subjective rankings. Table IX summarizes the experimental results, where 1 to 7 correspond to the best to worst performances, and inconsistent rankings are marked in boldface.

From Table IX, we have the following two findings. (1) Among the ten DIBR quality metrics, the proposed LOGS achieves the best consistency with the subjective rankings. Specifically, for "Book arrival" and "Newspaper" sequences, the rankings of the proposed metric are exactly the same with the ground truth. For "Lovebird" sequence, there are only two inconsistent rankings for the proposed method. These results are much better than other DIBR quality metrics. It is also noted that the second best metric is 3DSwIM [9], which also considers the geometric displacements between the reference and synthesized images. This further indicates that geometric distortions should be considered in the quality assessment of DIBR-synthesized images. (2) Due to the diversity of image

contents, a single DIBR algorithm is unlikely to produce the best synthesis results for all sequences. For example, DIBR algorithm A1 [32] performs the best in "Book arrival" and "Newspaper" sequences, but only ranks the fifth in "Lovebird" sequence. From this perspective, the proposed method can be used for automatic algorithm selection in DIBR.

V. Conclusion

In this work, we have been dedicated to the objective quality evaluation of DIBR-synthesized images, which is an important problem in Free Viewpoint Videos. A modular framework has been proposed to simultaneously evaluating local geometric distortions and global sharpness in the synthesized images. A visual image generation stage has been employed to mimic the characteristic of the HVS in perceiving image quality. A SIFT-flow based warping method is then proposed to detect the disoccluded regions, based on which the size and distortion strength are combined to generate a quality score for local geometric distortions. We have also proposed a reblurring-based method to measure the global sharpness of synthesized images. A combination of them has been proved very effective for the quality evaluation of DIBR-based view synthesis, and the experimental results outperform those of the existing metrics. We have also used the proposed metric for both improving existing quality metrics and benchmarking DIBR approaches, and very promising results have been obtained for both applications.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions for improving the quality of this paper. They would also like to thank Prof. M. Shahid Farid from Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy, for providing the codes of SIQE [11] and DSQM [12] to conduct comparisons in the revision of this manuscript, as well as for the valuable discussions.

References

[1] Y. C. Fan, P. K. Huang, and D. W. Shen, "3DTV depth map reconstruction based on structured light scheme," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2013, pp. 835–838.

[2] 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, Sep. 2011.

[3] 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, pp. 1–4.

[4] M. Solh, G. AlRegib, and J. M. Bauza, "3VQM: A vision-based quality measure for DIBR-based 3D videos," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2011, pp. 1–6.

[5] P. H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," *Electron. Imag. Int. Society Opt. Photon.*, vol. 8288, pp. 8288–8256, Feb. 2012.

[6] D. S. Stankovic, D. Kukolj, and P. L. Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *Proc. IEEE Int. Workshop Quality Multimedia Experience*, Jan. 2015, pp. 1–6.

[7] D. S. Stankovic, D. Kukolj, and P. L. Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. True Vis.-Capture, Transm. Display 3D Video*, Oct. 2015, pp. 1–4.

[8] D. S. Stankovic, D. Kukolj, and P. L. Callet, "Multi-scale synthesized view assessment based on morphological pyramids," *J. Elect. Eng.*, vol. 67, no. 1, pp. 3–11, Jan. 2016.

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

[10] Y. J. Jung, H. G. Kim, and Y. M. Ro, "Critical binocular asymmetry measure for 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.

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

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

[13] K. Gu *et al.*, "Model-based referenceless quality metric of 3D synthesized images using local image description," *IEEE Trans. Image Process.*, 2017, doi: 10.1109/TIP.2017.2733164, to be published.

[14] S. Ryu, S. Kim, and K. Sohn, "Synthesis quality prediction model based on distortion intolerance," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 585–589.

[15] T. H. Le, V. T. Long, D. T. Duong, and S. W. Jung, "Reduced reference quality metric for synthesized virtual views in 3DTV," *ETRI J.*, vol. 38, no. 6, pp. 1114–1123, Dec. 2016.

[16] 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.

[17] R. G. Fang, D. P. Wu, and L. Q. Shen, "Evaluation of image quality of experience in consideration of viewing distance," in *Proc. China Summit Int. Conf. Signal Inf. Process.*, 2015, pp. 653–657.

[18] H. Kim, J. Y. Kim, T. G. Oh, and S. H. Lee, "Blind sharpness prediction for ultra-high-definition video based on human visual resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 951–964, May 2017, doi: 10.1109/TCSVT.2016.2515303.

[19] K. Gu, M. Liu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.

[20] K. Gu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Self-adaptive scale transform for IQA metric," in *Proc. IEEE Int. Conf. Symp. Circuits Syst.*, May 2013, pp. 2365–2368.

[21] K. Gu *et al.*, "Adaptive high-frequency clipping for improved image quality assessment," in *Proc. IEEE Conf. Vis. Commun. Image Process.*, Nov. 2013, pp. 1–6.

[22] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2008, vol. 5304, pp. 28–42.

[23] K. Gu *et al.*, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.

[24] G. T. Zhai, X. L. Wu, X. K. Yang, W. S. Lin, and W. J. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[25] K. Gu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.

[26] S. Q. Wang *et al.*, "Guided image contrast enhancement based on retrieved images in cloud," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 219–232, Feb. 2016.

[27] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1–13, Jun. 2016.

[28] L. D. Li, Y. Zhou, J. J. Wu, W. S. Lin, and H. L. Li, "GridSAR: Grid strength and regularity for robust evaluation of blocking artifacts in JPEG images," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 153–163, Jul. 2015.

[29] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process., Image Commu.*, vol. 19, no. 2, pp. 163–172, Feb. 2004.

[30] R. Song, H. S. Ko, and C.-C. Jay 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, Sep. 2015.

[31] X. K. Liu *et al.*, "Subjective and objective video quality assessment of 3D synthesized views with texturedepth compression distortion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4847–4861, Dec. 2015.

[32] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. SPIE Conf. Stereoscopic Displays Virtual Reality Syst. X*, Jan. 2004, vol. 5291, pp. 93–104.

[33] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph., Tools*, vol. 9, no. 1, pp. 23–34, 2004.

[34] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3-D warping using depth information for FTV," *Signal Process., Image Commun.*, vol. 24, pp. 65–72, Jan. 2009.
[35] K. Muller *et al.*, "View synthesis for advanced 3-D video systems," *J. Image Video Process.*, vol. 2008, Nov. 2008, Art. no. 438148.
[36] P. N. Nya *et al.*, "Depth image based rendering with advanced texture synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2010, pp. 424–429.
[37] M. Koppel *et al.*, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1809–1812.
[38] G. Cermak, L. Thorpe, and M. Pinson, "Test plan for evaluation of video quality models for use with high definition TV content," Video Quality Experts Group, Jun. 2009.
[39] L. Zhang, L. Zhang, X. Q. 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.
[40] A. M. Liu, W. S. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
[41] K. Gu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 506–510.
[42] K. Gu, L. D. Li, H. Lu, X. K. Min, and W. S. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.
[43] 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.
[44] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.
[45] C. T. Vu, T. D. Phan, and D. M. Chandler, "S3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 934–945, Mar. 2012.
[46] 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.
[47] R. Hassen, Z. Wang, and M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.
[48] K. Bahrami and A. C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 751–755, Jun. 2014.
[49] Q. B. Sang, H. X. Qi, X. J. Wu, C. F. Li, and A. C. Bovik, "No-reference image blur index based on singular value curve," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1625–1630, Oct. 2014.
[50] L. D. Li *et al.*, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.
[51] K. Gu, G. T. Zhai, W. S. Lin, X. K. Yang, and W. J. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.
[52] L. D. Li *et al.*, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.

Leida Li (M'14) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. From February to June 2008, he was a visiting Ph.D. student in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan. From January 2014 to January 2015, he was a Visiting Research Fellow, and from July 2016 to July 2017 a Senior Research Fellow in the Rapid-Rich Object Search Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Full Professor in the School of Information and Control Engineering, China University of Mining and Technology, China. His research interests include multimedia quality assessment, information hiding, and image forensics.

Yu Zhou received the B.S. degree in 2014 from China University of Mining and Technology, Xuzhou, China, where she is currently working toward the Ph.D. degree in the School of Information and Control Engineering. Her research interests include multimedia quality assessment and perceptual image processing.

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 the reviewer for more than 20 SCI journals, such as IEEE T-IP, T-MM, T-CSVT, T-BC, T-CYB, T-NNLS, T-IE, J-STSP, SPL, Access, Information Sciences, Neurocomputing, SPIC, JVCI, DSP, MTAP, ELL, etc. His research interests include image quality assessment, visual saliency, contrast enhancement, machine learning, and air quality prediction.

Weisi Lin (M'92–SM'98–F'16) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from Kings College, London University, London, U.K., in 1992. He was involved in teaching and research with Zhongshan University, Shantou University, Shantou, China, Bath University, Bath, U.K., the National University of Singapore, the Institute of Microelectronics, Singapore, and the Institute for Infocomm Research, Singapore. He was the Laboratory Head of Visual Processing and the Acting Department Manager of Media Processing, Institute for Infocomm Research. He is currently a Full Professor in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

Shiqi Wang received the B.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2008, and the Ph.D. degree in computer application technology from Peking University, Beijing, China, in 2014. From March 2014 to March 2016, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From April 2016 to April 2017, he was with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He is currently an Assistant Professor in the Department of Computer Science, City University of Hong Kong, Hong Kong. He has proposed more than 30 technical proposals to ISO/MPEG, ITU-T, and AVS standards. His research interests include image/video compression, analysis, and quality assessment.