

Quality Index for View Synthesis by Measuring Instance Degradation and Global Appearance

Leida Li , Member, IEEE, Yu Zhou , Jinjian Wu , Fu Li , and Guangming Shi , Senior Member, IEEE

Abstract—Virtual view synthesis plays a vital role in the application of multi-view and free-viewpoint videos. Depth-image-based rendering (DIBR) is the most commonly used approach in view synthesis, and many DIBR algorithms have been proposed. However, how to evaluate the quality of DIBR-synthesized images and benchmark the DIBR algorithms are still very challenging, which may hinder the further development of the view synthesis technique. Hence, an effective quality metric for evaluating the distortions in view synthesis is urgently needed. With this motivation, this paper presents a quality index for view synthesis by simultaneously measuring local Instance DEgradation and global Appearance (IDEA). Due to the imperfection of rendering algorithms, local geometric distortions are easily introduced around instance contours, causing instance degradation, which is the dominant distortion in synthesized views. In this work, image instances are first detected and local instance degradation is measured based on discrete orthogonal moments. Meantime, we propose to measure the global appearance of synthesized images based on the superpixel representation. By integrating both local and global aspects of the distortions, a more accurate quality model is built for view synthesis. Extensive experiments and comparisons have demonstrated the superiority of the proposed method in evaluating the quality of DIBR-synthesized images and benchmarking the performance of view synthesis algorithms.

Index Terms—View synthesis, quality assessment, depth-image-based rendering (DIBR), instance degradation, global appearance.

I. INTRODUCTION

THE wide applications of multi-view and free-viewpoint videos have boosted the development of virtual view synthesis [1]–[3]. Depth-Image-Based Rendering (DIBR) is the most popular view synthesis technique, which has significant

Manuscript received August 12, 2019; revised December 5, 2019; accepted March 5, 2020. Date of publication March 13, 2020; date of current version December 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grants 61771473, 61991451, 61672404, and 61379143, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20181354, in part by Six Talents Peaks High-level Talents in Jiangsu Province under Grant XYDXX-063, and in part by the Fundamental Research Funds of the Central Universities of China under Grant JC1904. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hantao Liu. (*Corresponding authors: Leida Li; Yu Zhou*)

Leida Li, Jinjian Wu, Fu Li, and Guangming Shi are with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ldli@xidian.edu.cn; jinjian.wu@mail.xidian.edu.cn; fuli@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

Yu Zhou is with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: zhouy7476@cumt.edu.cn).

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

Digital Object Identifier 10.1109/TMM.2020.2980185

influence on the quality of the synthesized views. Up to now, many DIBR algorithms have been proposed [4]–[7]. However, there is not a single DIBR algorithm that can work consistently well on all images with diversified visual contents. The accompanying problems are multifaceted. First, for a specific view, how to select a DIBR algorithm that can produce the best-quality synthesized view is difficult. Second, for a specific DIBR algorithm, how to determine the best parameters for a given image, which is typically ad-hoc, is aimless. For both tasks, an effective quality metric for view synthesis is highly desired, which can be used for both evaluating the quality of synthesized views and benchmarking DIBR algorithms.

Typically, DIBR consists of two stages, i.e. warping and rendering [8]. The warping operation is to first project the reference view to the 3D Euclidean space under the guidance of the depth information, and then re-project the 3D-space view to the target 2D view. In this stage, local disoccluded regions are usually produced in the target views. The subsequent rendering stage is employed to restore the disoccluded regions, producing the synthesized views. In practice, the disoccluded regions are usually present around the contours of instances in an image, and the imperfect rendering operation easily causes local geometric degradations to image instances. This is the dominated distortion type in view synthesis. Meantime, the forward/backward warping operations inevitably incur interpolation errors in the synthesized views, which further impacts the global appearance of the synthesized image. Therefore, it's necessary to design the quality model for DIBR-synthesized images from both local and global aspects.

In the literature, many traditional quality metrics have been proposed [9]–[15], which can be clarified into three types according to whether and how much the reference information is needed by the quality model, namely full-reference, reduced-reference, and no-reference. These quality metrics are mostly designed for the images with evenly distributed distortions. However, the dominated distortions in DIBR-synthesized images are locally distributed around the instance contours. This undoubtedly causes the fact that the traditional quality metrics underestimate the effect of the local distortions on image quality, further limiting their performance.

Recently, several quality metrics for DIBR-synthesized views have been proposed, which can be divided into three categories according to the representation level of the distortions. While these metrics have achieved notable success, challenges remain. We will detail each category of metrics and deeply analyze the factors that limit their performance.

- The first category is done by measuring the distortions in the synthesized image in a holistic manner, which is similar to the classic image quality metrics. In [16], the Morphological Wavelet decomposition was first utilized to decompose the reference view and the corresponding synthesized view into multiple scales. Subsequently, the Peak-Signal-to-Noise Ratio (PSNR) values at all scales were computed and integrated to produce the final quality score, producing the MW-PSNR metric. In [17], MW-PSNR was further improved by replacing the morphological wavelet decomposition with the Morphological Pyramids decomposition, producing the MP-PSNR metric. Later, both MW-PSNR and MP-PSNR were further improved by discarding the coarse-scale decomposition subbands when computing the quality score, producing the Reduced versions of MW-PSNR and MP-PSNR, which were called RMW-PSNR [16] and RMP-PSNR [18], respectively. Tian *et al.* [28] proposed a No-reference Image Quality assessment metric for Synthesized Views (NIQSV) by integrating the global luminance, contrast and saturation distortions. In [29], the low-level and mid-level structural distortions were combined to build the view synthesis quality metric. As aforementioned, both local and global distortions degrade the DIBR-synthesized images. However, the first category of view synthesis quality metrics only considers the global aspects of the distortions in the synthesized image, which tends to underestimate the local instance degradations.
- The second category of approaches emphatically measure local geometric distortions in the synthesized views. Gu *et al.* [19] proposed a quality metric based on local image description. The local geometric distortions were first detected by calculating the difference between the DIBR-synthesized image and the auto-regression reconstructed image. Next, the geometric distortions were quantified after integrating the visual saliency, producing the final quality score. In [27], an outlier detection based quality metric for synthesized images was proposed. The geometric and structural distortions were first identified using the median filtering. Afterwards, the quality score was calculated by quantifying both types of distortions. This category of metrics only evaluates the local distortions without considering the global appearance, which also leads to inaccurate evaluation.
- The third category is achieved by integrating the ideas in the first and second categories of approaches. Li *et al.* [30] proposed a view synthesis quality metric based on the Local Geometric distortions and global Sharpness (LOGS). Tian *et al.* [31] proposed to add two modules to severally evaluate the local black holes and the stretching distortions based on the global quality measure method, i.e. NIQSV [28], producing the NIQSV+ metric. This category of metrics takes into account both local and global aspects of the distortions, which also delivers better performances. However, it should be noted that in the existing metrics, the local geometric distortions are commonly measured using relatively low-level features, e.g. edges and

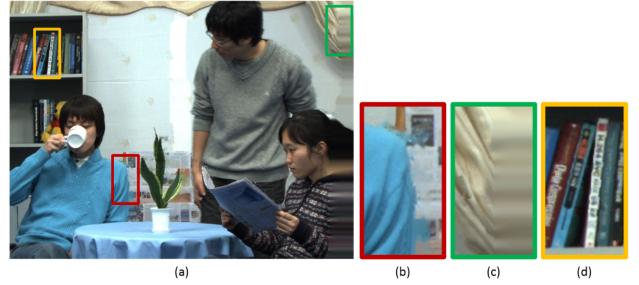


Fig. 1. Illustration of the distortions in view synthesis. Image (a): a DIBR-synthesized image; images (b)–(d): a disoccluded region with rendering distortions, a background region with blurring effect, and a background region with geometric deformation, respectively.

gradients. However, people tend to perceive an image in a higher level, especially from the salient objects (instances) in an image. As a result, although the DIBR process impairs all edges in the synthesized image, degradations around instance edges have the most significant impact on the quality of the image. In order to model the distortions more effectively, the local geometric distortions should be considered in a higher level. Furthermore, the displacement between the reference and target views also causes deformations in the whole image, which is not fully investigated in the existing metrics. Fig. 1 shows an example of distortions in a DIBR-synthesized image. Fig. 1(b) vividly shows the degradations around salient instance edges. Figs. 1(c) and 1(d) clearly show the distortions in the background regions in terms of image global appearance, including blurring effect and deformation.

Motivated by the above facts, this paper presents a quality metric for view synthesis by measuring the local Instance DEgradation and global Appearance (IDEA). The design principles of the IDEA method are as follows.

- For measuring the local instance degradation, an instance segmentation network trained on the COCO database [37] is employed to detect the instances from both the reference and synthesized views, producing the instance masks. Considering that the visually important distortion regions mainly locate at instance edges, edges of instance masks are first located and dilation operation is conducted to construct the local distortion region, from which the local instance degradations can be comprehensively investigated. Next, the instance degradations are quantified based on the Tchebichef moment [38], [39], which is effective for shape representation. The features of all instances are then averaged to produce the feature for describing the local instance degradations. Afterwards, the distance between the feature of the synthesized view and that of the reference view is calculated to produce the instance degradation score of the synthesized view.
- For the global appearance, a superpixel-based approach is proposed. This is motivated by the fact that the human visual system (HVS) tends to aggregate the regions with similar characteristics for image perception. Specifically, superpixel segmentation is first conducted on both the

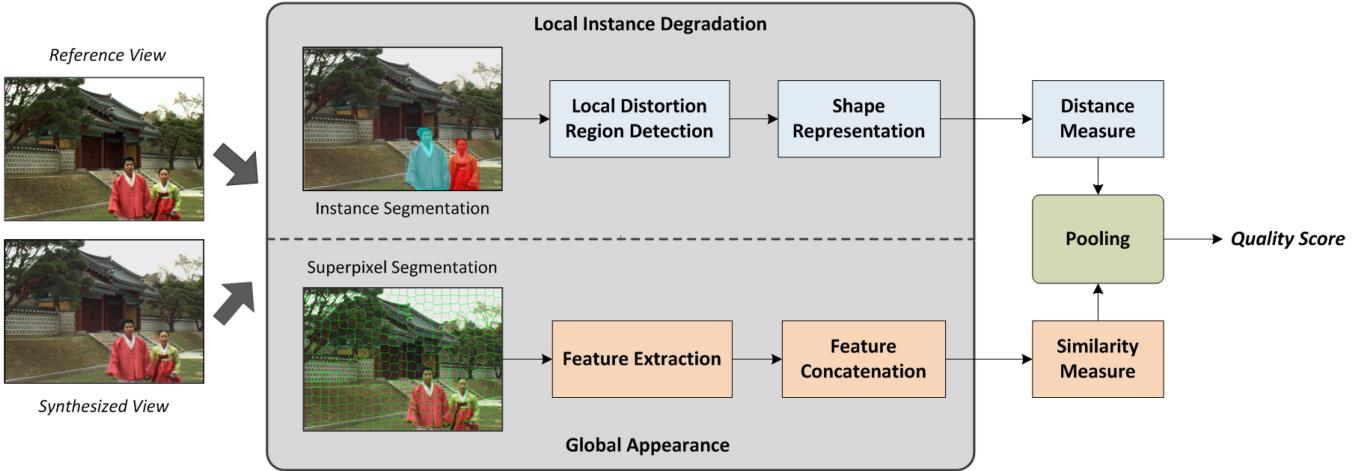


Fig. 2. Flowchart of the proposed method.

reference and synthesized views. Then two kinds of features are extracted on each superpixel to measure the blurring effect and deformation in global appearance, and the features on all the superpixels of the reference/synthesized views are concatenated to form an one-dimensional matrix. The similarity between the matrix of the reference view and that of the synthesized view is computed as the global appearance score.

- 3) Finally, the local instance score and the global appearance score are integrated to generate the overall quality score of the synthesized view.

The contributions of the proposed method are three-fold. First, two sets of features are designed to describe both the local and global distortions in the synthesized view, instead of merely measuring a single aspect of distortions as most existing metrics. Second, the local degradation is evaluated from a higher level, which is more consistent with the characteristics of paying more attention to salient objects of the HVS in quality perception. Third, extensive experiments conducted on two public view synthesis quality databases have demonstrated that the proposed IDEA method is superior to the state-of-the-art metrics in evaluating the quality of DIBR-synthesized images and benchmarking the performance of DIBR algorithms.

II. PROPOSED VIEW SYNTHESIS QUALITY METRIC

The design philosophy of the proposed metric is as follows. Both local and global aspects of the distortions in view synthesis are taken into account. The local aspect is evaluated in semantic level based on image instances. The global aspect is evaluated in appearance level based on superpixel segmentation. To be more specific, local instance degradation is measured based on instance shape change, and global appearance is measured based on blurring and deformation. After both aspects are evaluated, pooling is conducted to generate the overall quality score. The flowchart of the proposed method is shown in Fig. 2.

A. Instance Degradation Measurement

As discussed above, the visually important local distortions mainly locate at the instance boundaries, which can be seen

from Fig. 1(b). This motivates us to evaluate the local distortions from the instance level. To achieve this goal, we resort to instance segmentation approaches for locating the local distortion regions. In the literature, many instance segmentation methods have been reported. In this work, the state-of-the-art Mask RCNN model [36] pre-trained on the COCO database is employed for instance segmentation from both reference and synthesized views.

After instance segmentation, the instance masks can be obtained. In Fig. 3, images (b) and (g) show the segmentation results of the reference view (a) and the synthesized view (f), respectively. From the figure, we can see that 1) the Mask RCNN is effective for instance segmentation; 2) the rendering distortions around instance edges make the segmented instances slightly different between the reference and synthesized views, e.g. the edges around the arm and the left-side head of the man. To detect the local distortion regions, the instance edges are first detected based on the instance masks \mathbf{IM} ,

$$\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}, \quad (1)$$

where \mathbf{G}_x and \mathbf{G}_y are the horizontal and vertical gradients, which are computed as,

$$\mathbf{G}_x = \mathbf{g} * \mathbf{IM}, \quad \mathbf{G}_y = \mathbf{g}' * \mathbf{IM}, \quad (2)$$

where $*$ and $'$ denote the convolution and transpose operations, and \mathbf{g} is the gradient operator,

$$\mathbf{g} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}. \quad (3)$$

Since the instance edge map itself contains very limited information on the instance degradation, a neighboring region centered at the edge maps is needed to have a comprehensive understanding of the characteristics of the instance degradations. To this end, the morphological open operation is conducted on the edge map \mathbf{G} to locate the local distortion regions. The edge detection result and the detected distortion regions are shown in the third and fourth columns of Fig. 3, respectively. From Figs. 3(e) and 3(j), we can see that the local distortion regions

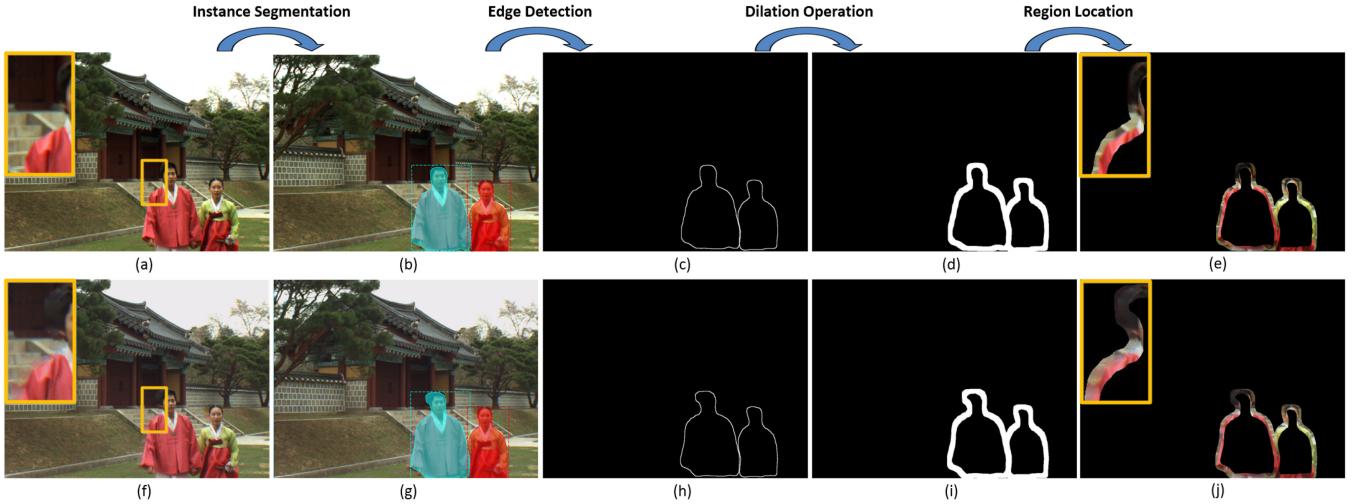


Fig. 3. Illustration of the proposed detection method for the local distorted regions. First column: A reference view and one corresponding synthesized view; Second column: Instance segmentation results using the MaskRCNN method [36]; Third column: Edge detection results on the instance masks; Fourth column: Edge inflation results; Last column: Regions in the initial images corresponding to the inflation results in the previous column.

causing the instance degradations can be fully covered in the detected regions.

The distortions in the local disoccluded regions are mainly present as the geometric distortions, which can be seen from Fig. 3 (a) and 3(f). Particularly, the local geometric distortions cause shape change of the instances in the reference and synthesized images. Discrete Tchebichef moments have been demonstrated to be effective for shape description [38]. Therefore, the Tchebichef moment is adopted [39] to quantify the instance degradations in this work. Specifically, the detected local distortion regions, as shown in Figs. 3(e) and 3(j), are first partitioned into blocks with equal size $p \times p$. Then the Tchebichef moments up to the $[(p-1)+(p-1)]$ th order of the i th block for the j th instance are calculated as

$$R_i^j = \begin{pmatrix} r_{00} & r_{01} & \cdots & r_{0(p-1)} \\ r_{10} & r_{11} & \cdots & r_{1(p-1)} \\ r_{20} & r_{21} & \cdots & r_{2(p-1)} \\ \vdots & & & \\ r_{(p-1)0} & r_{(p-1)1} & \cdots & r_{(p-1)(p-1)} \end{pmatrix}, \quad (4)$$

where r_{00} denotes the DC component of the moments, and the others are AC components.

The energy of all the AC components is calculated as

$$E_i^j = \left(\sum_{g,h=0}^{p-1} r_{g,h}^2 \right) - r_{00}^2. \quad (5)$$

Next, the sum of the energy of all the blocks belonging to the j th instance is computed as follows

$$E(j) = \sum_{i=0}^m E_i^j, \quad (6)$$

where m denotes the number of blocks for the j th instance.

The energy of all the instances are subsequently averaged as

$$E = \frac{1}{n} \sum_{j=0}^n E(j), \quad (7)$$

where n denotes the number of instances in an image.

In this work, the instance energies of the reference and synthesized views are denoted by E_r and E_s , respectively. Finally, the absolute distance between E_r and E_s is calculated as the local instance degradation score of the synthesized view,

$$Q_L = |E_r - E_s|, \quad (8)$$

where lower Q_L value represents better quality for the synthesized views.

B. Global Appearance Measurement

To measure the global appearance of a synthesized image, a superpixel-based method is proposed in this work. This is inspired by the fact that the HVS tends to aggregate the regions with similar characteristics for image perception, and superpixel is just the perceptually meaningful region composed of spatial neighboring pixels with similar intensity, color and adjacent spatial locations [32]. This merit makes superpixel an effective way for feature representation in many image processing tasks [33]–[35]. In this work, the Simple Linear Iterative Clustering (SLIC) [32] method is adopted for superpixel segmentation due to its computational efficiency and preservation of image boundaries. The SLIC superpixel segmentation method allows users to set the number of superpixels N and the compactness C . In this work, N and C are set to 420 and 25, respectively.

Fig. 4 shows an example of superpixel segmentation from a reference view and the corresponding synthesized view. To intuitively demonstrate the differences between superpixels between the reference and synthesized views, the superpixel edges of two views are superimposed and shown in image 4(c). From Fig. 4(c), it can be clearly observed that deformations occur in the

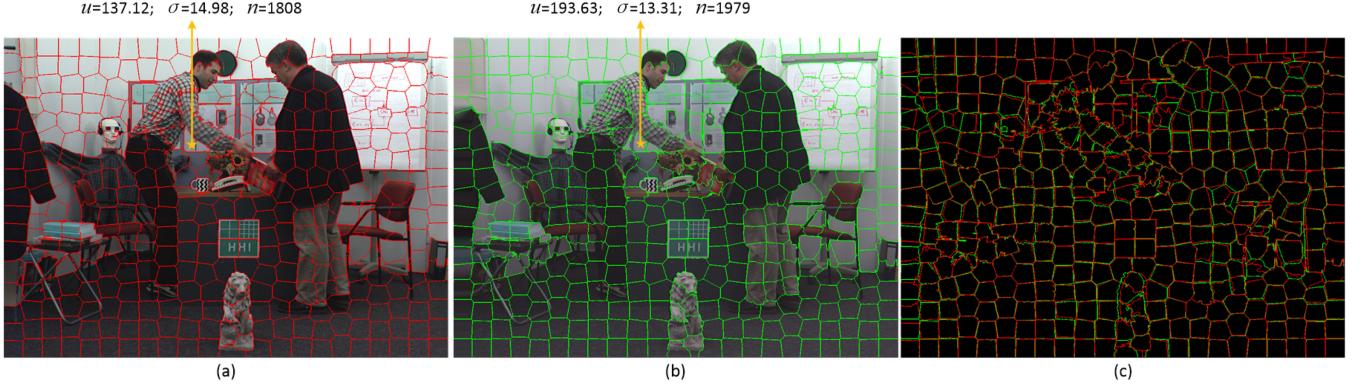


Fig. 4. Superpixel segmentation results of a reference view and a synthesized view in images (a) and (b), together with the three feature values of one pair of superpixels. Image (c): The superposition version of the superpixel edges in images (a) and (b) for intuitively demonstrating the deformations in global appearance.

reference and synthesized images, which can be well captured by superpixels. This characteristic has not been fully investigated before in the literature. In addition to the deformations, blurring effect is another common distortion in view synthesis [30]. Therefore, in this work, we evaluate the global appearance from blurring and deformation.

For blurring effect, we simply compute the mean u_i and standard deviation σ_i for the i th superpixel s_i

$$u_i = \frac{1}{n_i} \sum_{j \in s_i} \mathbf{I}(j), i \in [1, N], \quad (9)$$

$$\sigma_i = \sqrt{\frac{1}{n_i - 1} \left(\sum_{j \in s_i} (\mathbf{I}(j) - u_i) \right)^2}. \quad (10)$$

For geometric deformation, we simply use the number of pixels in a superpixel as feature representation, which is denoted by n_i .

Next, the three feature values of all superpixels are concatenated to form an one-dimensional feature vector \mathbf{M} ,

$$\mathbf{M} = \{u_1, \sigma_1, n_1, u_2, \sigma_2, n_2, \dots, u_N, \sigma_N, n_N\}. \quad (11)$$

For the reference view \mathbf{I}_r and the corresponding synthesized view \mathbf{I}_s , the feature matrices are denoted by \mathbf{M}_r and \mathbf{M}_s , respectively. Subsequently, the similarity between \mathbf{M}_r and \mathbf{M}_s is computed as

$$\mathbf{S} = \frac{2\mathbf{M}_r \mathbf{M}_s + c}{\mathbf{M}_r^2 + \mathbf{M}_s^2 + c}, \quad (12)$$

where c is a small constant to ensure numerical stability.

Finally, the global appearance score Q_G is calculated as,

$$Q_G = \frac{1}{3N} \sum_{k=1}^{3N} \mathbf{S}_k, \quad (13)$$

where higher Q_G value denotes better quality for the synthesized view.

C. Pooling

As aforementioned, lower Q_L value and higher Q_G represent better quality. Therefore, we design a simple pooling strategy

to integrate the instance degradation and global appearance for generating the final quality score as follows

$$Q = w_1 Q_G - w_2 Q_L, \quad (14)$$

where w_1 and w_2 denote the weights that represent the relative importance of global appearance Q_G and local instance degradation Q_L on the overall view synthesis quality evaluation. Higher Q value indicates better image quality.

III. EXPERIMENTAL RESULTS

A. Experimental Settings

The performance of the IDEA method is evaluated on two public view synthesis quality databases, e.g. IRCCyN/IVC DIBR image database [8] and IVY database [48].

- IRCCyN/IVC DIBR Image Database [8]. This database contains images with three different scenes, and for each scene there are four different viewpoints. So, 12 views can be generated using a single view synthesis algorithm. Seven DIBR algorithms (denoted by A1 [4], A2 [40], A3 [5], A4 [41], A5 [42], A6 [43], A7) are employed to synthesize the virtual views, producing 84 synthesized views in total. Since the A7 algorithm only includes the warping operation without the subsequent rendering, the views synthesized by A7 contains black holes and is not included in our experiments. Therefore, in our experiments the remaining 72 synthesized images generated using DIBR algorithms A1-A6 are adopted for performance evaluation. The Mean Opinion Score (MOS) of images are provided in this database, which was obtained using the Absolute Category Rating-Hidden Reference (ACR-HR) method. In other words, no explicit reference images are included in the subjective test. Similar to [21], [22], the initial MOS values are transformed to the Difference Mean Opinion Scores (DMOS) for the full-reference quality assessment [23],

$$DMOS = MOS_{syn} - MOS_{ref} + 5, \quad (15)$$

where MOS_{syn} and MOS_{ref} denote the MOS values of the synthesized view and the corresponding reference view,

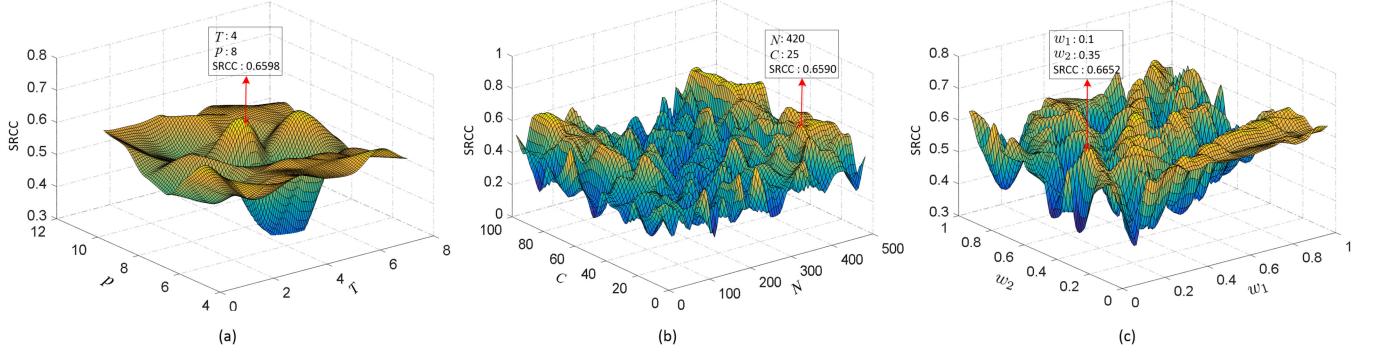


Fig. 5. Impact of three sets of parameters on the performance of the proposed method. Image (a): Impact of the open operation time T and the block size p on the performance of the IDM module; Image (b): Impact of the superpixel number N and the compactness C on the performance of the GAM module; Image (c): Impact of the weights w_1 and w_2 on the performance of the proposed method.

TABLE I
DETAILS OF THE IRCCyN/IVC AND IVY DATABASES

| Database | Syn. Views | Scenes | DIBR Algos | Score |
|----------------|------------|--------|------------|-------|
| IRCCyN/IVC [8] | 84 | 3 | 7 | MOS |
| IVY [48] | 84 | 7 | 4 | DMOS |

respectively. After the transformation, higher DMOS value represents better quality [21], [22].

- IVY Database [48]. This database consists of 84 stereo images, which are synthesized from 7 reference scenes using 4 DIBR view synthesis algorithms (denoted by B1 [44], B2 [45], B3 [46], B4 [47]). Totally, 12 viewpoint pairs are generated for each scene. The DMOS values generated using the double stimulus continuous quality scale method are provided with the database. Here, lower DMOS value indicates better quality. Besides, it should be noted that since the distortions in the left- and right-views are symmetric, the quality of two views are averaged as the final quality score in our experiments. Detailed information on the two databases is summarized in Table I.

B. Parameter Selection

There are three sets of parameters that have impact on the performance of the proposed method, namely the open operation time T and the block size p during calculating the local instance degradation, the superpixel number N and the compactness C during measuring the global appearance, and the weights w_1 and w_2 that separately determine the relative importance of the global and local quality. The parameters are optimized based on the SRCC value, which is a commonly used criterion for parameter optimization in the field of image quality assessment [49], [50]. Larger SRCC value indicates the higher monotonicity consistency between the predicted scores and the subjective scores. This optimization process is conducted on the IRCCyN/IVC database and the optimal parameters are directly tested on the IVY database.

1) *Impact of T and p :* Fig. 5(a) shows the performance of the instance degradation measurement (IDM) module with different

values of the open operation time T and the block size p . It can be observed from this figure that the performance of the IDM module varies with the changes of the T and p values, and the IDM module obtains the highest SRCC value when the values of T and p are set to 4 and 8, respectively.

2) *Impact of N and C :* For the global appearance measurement (GAM) module, there exist two parameters, i.e. the superpixel number N and the compactness C . In this part, we testify the impact of two parameters on the performance of the GAM module, which can be seen from Fig. 5(b). From this figure, we can see that both N and C have impacts on the performance of the GAM module, and when N and C are severally set to 420 and 25, the GAM module achieves the best performance.

3) *Impact of w_1 and w_2 :* The proposed method consists of the IDM and GAM modules, and they both contribute for the performance of the proposed method. w_1 and w_2 are the weights of GAM and IDM, respectively, reflecting their relative contributions for the proposed method. Fig. 5(c) illustrates the performance of the proposed method with different values of w_1 and w_2 . From the experimental results, we can observe that when w_1 and w_2 are set to 0.1 and 0.35, the proposed method obtains the optimal performance. This indicates that the proposed IDM module contributes more for the performance of the proposed method, which is consistent with the fact that the local rendering distortions causing the instance degradation is the dominant factor that affects the perceptual quality of the synthesized view.

In this work, the values of the three sets of parameters are set to the optimal ones.

C. Performance Evaluation

1) *Comparison With the State-of-the-Arts:* In this part, the performance of the proposed IDEA metric is compared with the state-of-the-arts, including five traditional quality metrics and eight existing view synthesis quality metrics. The traditional quality metrics are SSIM [9], VIF [10], FSIM [11], GSM [12], and LTG [13], and the view synthesis quality metrics are MP-PSNR [17], MW-PSNR [16], RMW-PSNR [16], LOGS [30], APT [19], NIQSV [28], NIQSV+ [31], and MNSS [20]. The performance evaluation is based on four commonly criteria, i.e.

TABLE II
PERFORMANCE OF THE PROPOSED METHOD AND THE STATE-OF-THE-ARTS ON THE WHOLE IRCCyN/IVC AND IVY DATABASES

| Method | Type | IRCCyN/IVC Database | | | | IVY Database | | | |
|----------------------|------|---------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| | | PLCC | SRCC | KRCC | RMSE | PLCC | SRCC | KRCC | RMSE |
| SSIM [9] | T | 0.5569 | 0.4093 | 0.2802 | 0.4686 | 0.5684 | 0.5662 | 0.4068 | 20.2826 |
| VIF [10] | T | 0.3209 | 0.2935 | 0.2061 | 0.5343 | 0.4013 | 0.3958 | 0.2685 | 22.5802 |
| FSIM [11] | T | 0.5116 | 0.3573 | 0.2416 | 0.4847 | 0.6118 | 0.5975 | 0.4223 | 19.4998 |
| GSM [12] | T | 0.5209 | 0.3963 | 0.2763 | 0.4816 | 0.5736 | 0.5805 | 0.4068 | 20.1934 |
| LTG [13] | T | 0.0349 | 0.1565 | 0.0950 | 0.5638 | 0.6214 | 0.6072 | 0.4337 | 19.3139 |
| MP-PSNR [17] | S | 0.7080 | 0.6216 | 0.4646 | 0.3984 | 0.6114 | 0.5954 | 0.4217 | 19.5066 |
| MW-PSNR [16] | S | 0.7454 | 0.6398 | 0.4717 | 0.3761 | 0.5240 | 0.5051 | 0.3528 | 20.9969 |
| RMW-PSNR [16] | S | 0.3341 | 0.5068 | 0.4008 | 0.5317 | 0.5224 | 0.5008 | 0.3540 | 21.0200 |
| LOGS [30] | S | 0.7243 | 0.6511 | 0.4849 | 0.3890 | 0.6442 | 0.6385 | 0.4509 | 18.8549 |
| APT [19] | S | 0.4223 | 0.5695 | 0.4495 | 0.5398 | 0.5240 | 0.4748 | 0.3389 | 20.9961 |
| NIQSV [28] | S | 0.4402 | 0.3130 | 0.2361 | 0.5066 | 0.4113 | 0.2717 | 0.1945 | 22.4706 |
| NIQSV+ [31] | S | 0.5226 | 0.5201 | 0.3653 | 0.4810 | 0.2191 | 0.2990 | 0.2037 | 24.0530 |
| MNSS [20] | S | 0.2912 | 0.4797 | 0.3078 | 0.5396 | 0.3845 | 0.2205 | 0.1474 | 22.7565 |
| Proposed IDEA | S | 0.7796 | 0.6652 | 0.4986 | 0.3533 | 0.6311 | 0.6132 | 0.4405 | 19.0379 |

PLCC, SRCC, KRCC and RMSE [30]. Among them, PLCC and RMSE are calculated to evaluate the prediction accuracy, SRCC and KRCC are used to evaluate the monotonicity. Higher values of PLCC, SRCC and KRCC, and lower RMSE value indicate the higher prediction accuracy and consistency [26]. Both PLCC and RMSE are computed following a five-parameter nonlinear mapping [24], [25],

$$f(x) = \psi_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\psi_2(x-\psi_3)}} \right) + \psi_4 x + \psi_5, \quad (16)$$

where x and $f(x)$ denote the predicted score and the mapped objective score respectively, and $\psi_1, \psi_2, \dots, \psi_5$ represent the fitting parameters. The aim of this function is to unify the objective scores calculated by different quality metrics into a common score space [53], [54].

Table II reports the experimental results on the IRCCyN/IVC and IVY databases, where the two best-performing metrics are marked in boldface. From this table, we can see that the IDEA method is among the top two on both databases. Particularly, on the IRCCyN/IVC database, the IDEA method shows obvious superiority to all the other metrics. On the IVY database, the performance of the IDEA method ranks the second, which is competitive to the LOGS metric. However, the PLCC value of the LOGS method on the IRCCyN/IVC database is much lower than that of the IDEA method, which indicates the advantage of the proposed IDEA method in prediction accuracy on the IRCCyN/IVC database. Therefore, the IDEA method has the best overall performance in evaluating the quality of synthesized views. Besides, the reason resulting in the slightly worse performance of the IDEA method on IVY database is that the instances that are correlated with quality of the “Aloe” scene images, i.e. aloes, are not detected. For clarity, Fig. 6 shows two examples of the “Aloe” scene images synthesized using two different DIBR algorithms [45], [47]. Therefore, for this scene,



Fig. 6. The examples of the “Aloe” scene images synthesized by two DIBR algorithms.

TABLE III
PERFORMANCE OF THE IDEA METHOD ON THE IVY DATABASE WITH AND WITHOUT THE ALOE SCENE IMAGES

| Scene | PLCC | SRCC | KRCC | RMSE |
|--------------|--------|--------|--------|---------|
| With Aloe | 0.6311 | 0.6132 | 0.4405 | 19.0379 |
| Without Aloe | 0.6773 | 0.6654 | 0.4613 | 18.2433 |

only the global appearance module is actually used for the quality assessment for all the “Aloe” images. To demonstrate this, we test the performance of the IDEA method for the other six scene images without the “Aloe” scene. The performance of the IDEA method on the IVY database with and without the “Aloe” scene images are summarized in Table III. It can be seen that when the “Aloe” scene images are not included, the performance of the IDEA method improves significantly. This problem can be alleviated by including more diversified visual contents and the corresponding annotations during the training of the instance segmentation model.

We further evaluate the performance of the IDEA method on real images with different scenes and qualities, which are



Fig. 7. Three images with different scenes and qualities in the IRCCyN/IVC database, together with their DMOS values, where the higher DMOS value indicates the better quality.

shown in Fig. 7. It can be observed that the qualities of the three images decrease from image (a) to image (c). So the quality scores predicted by objective metrics are also expected to follow the same trend. This task is quite challenging for image quality models, since the images have completely different visual contents. Fig. 8 shows histograms of the DMOS values and the objective scores predicted by the thirteen quality metrics and the proposed method. From the figure, we can see that only the proposed method can generate the correct predictions.

2) *Statistical Significance Analysis*: We further do the F-test to compare the statistical performance of the proposed IDEA method and the state-of-the-arts [51], [52]. To compare the statistical performance of a method X with that of a method Y , the F-test score is first calculated based on the RMSE values of the X and Y methods,

$$F_t = \left(\frac{\text{RMSE}_X}{\text{RMSE}_Y} \right)^2. \quad (17)$$

Subsequently, the threshold $F_{critical}$ for each database is computed using the MATLAB function *finv* based on the number of residuals and the confidence level [20], [51]. In this work, the confidence level with confidence level 90%. If the F_t value is larger than the $F_{critical}$ value, the method Y is believed to have significantly better statistical performance than the method X . If the F_t value is between $1/F_{critical}$ and $F_{critical}$, the method Y is believed to have competitive performance with the compared metric X . Otherwise, the method Y is believed to be significantly worse than metric X in the statistical performance. For the IRCCyN/IVC and IVY databases, their $F_{critical}$ values are 1.3549 and 1.3244, respectively.

The experimental results of the statistical performance are summarized in Table IV, where the value “1” indicates that the method Y is superior to the compared metric X , and value “0” represents the performances of the method Y and method X are competitive, while “-1” indicates the significantly better performance of the metric X . The results before and after the comma represent the ones on the IRCCyN/IVC database and the IVY database, respectively. From the experimental results, we can see that no metric performs significantly better than the IDEA method. Especially on the IRCCyN/IVC database, the IDEA method shows obvious superiority to most existing quality metrics. Besides, among the thirteen existing quality metrics, only three methods have competitive statistical performance

with the IDEA method on both database, i.e. MW-PSNR [16], MP-PSNR [17] and LOGS [30].

D. Impact of Different Superpixel Segmentation Approaches

In this part, we study the impact of different superpixel segmentation approaches on the performance of the proposed method. The methods compared with the SLIC method [32] are from [55] and [56]. Table V summarizes the experimental results. It can be observed from this table that the IDEA method using the SLIC approach [32] performs the best on both databases. Therefore, in our proposed method, the SLIC method is adopted for superpixel segmentation.

E. Impact of Different Edge Detection Operators

In the IDM module, the Sobel operator is used for mask edge detection. To study the performance impact of different edge detection operators, we further conduct an experiment to replace the Sobel operator with three other commonly used operators, including Canny [57], Prewitt [58] and Roberts [59]. Table VI summarizes the experimental results, from which we can observe that the proposed method with the Sobel operator achieves the best performance. Therefore, we employ this method for the mask edge detection.

F. Ablation Study

In order to further investigate the relative contributions of the two modules in the proposed metric, i.e. local instance degradation and global appearance, we conduct an ablation study. In this experiment, the two modules are used to predict the quality scores, respectively. The experimental results are summarized in Table VII.

Three findings can be obtained from the ablation study. First, compared with the experimental results of the state-of-the-arts presented in Table II, either the local instance degradation module or the global appearance module still shows obvious advantages to most of the existing quality metrics on both databases. This confirms the effectiveness of the two modules, which are both important for view synthesis quality modeling. Second, the local instance degradation module delivers better performances on both databases. This indicates that local distortions contributes more than the global appearance for the proposed method. The main reason causing this phenomenon is that the local instance degradation is the dominated factor that affects the quality of DIBR-synthesized images. Third, when the two modules are integrated in the proposed metric, the performance achieves significant gain. This demonstrates that the two modules are complementary and it is necessary to take into account both aspects in modeling the view synthesis quality.

G. Benchmarking DIBR Algorithms

In the practical application of the view synthesis techniques, one common issue is how to choose the most suitable DIBR algorithm that can achieve the best synthesis effect. A better DIBR algorithm generates virtual view with higher quality. Although the subjective test can reliably evaluate the quality of the synthesized

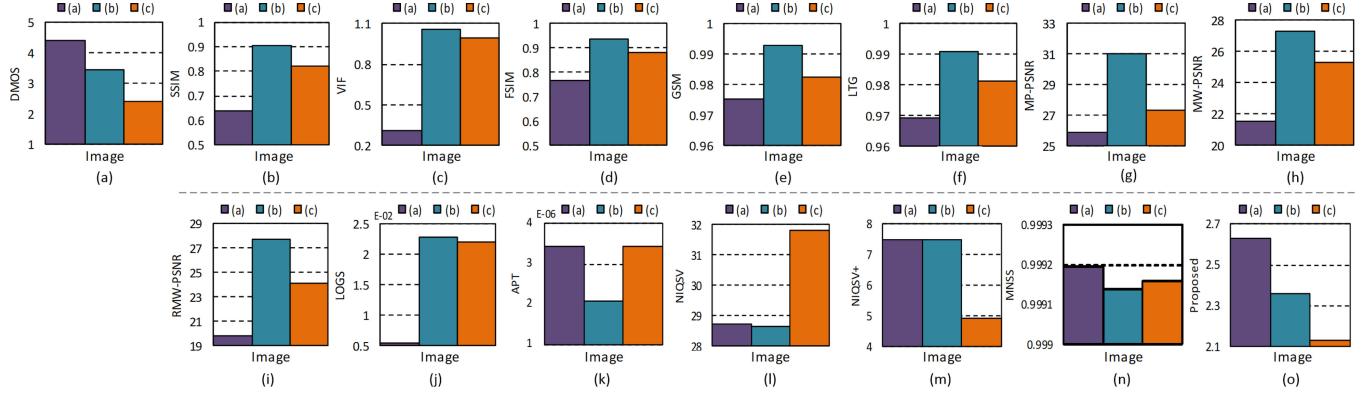


Fig. 8. DMOS values of the three images in Fig. 7 and the corresponding objective scores predicted by thirteen state-of-the-arts and the proposed method.

TABLE IV
STATISTICAL PERFORMANCE OF THE PROPOSED IDEA METHOD AND THE STATE-OF-THE-ARTS ON THE IRCCyN/IVC AND IVY DATABASES

| x \ y | SSIM | VIF | FSIM | GSM | LTG | MP-PSNR | MW-PSNR | RMW-PSNR | LOGS | APT | NIQSV | NIQSV+ | MNSS | IDEA |
|-------------|-------|------|------|-------|------|---------|---------|----------|-------|------|-------|--------|------|-------|
| SSIM | 0, 0 | 0, 0 | 0, 0 | 0, 0 | 1, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 0 | 0, 1 | 0, 0 | -1, 0 |
| VIF | 0, 0 | 0, 0 | 0,-1 | 0, 0 | 0,-1 | -1,-1 | -1, 0 | 0, 0 | -1,-1 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1,-1 |
| FSIM | 0, 0 | 0, 1 | 0, 0 | 0, 0 | 0, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 1 | 0, 1 | 0, 1 | -1, 0 |
| GSM | 0, 0 | 0, 0 | 0, 0 | 0, 0 | 1, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 0 | 0, 1 | 0, 0 | -1, 0 |
| LTG | -1, 0 | 0, 1 | 0, 0 | -1, 0 | 0, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 1 | -1, 1 | 0, 1 | -1, 0 |
| MP-PSNR | 1, 0 | 1, 1 | 1, 0 | 1, 0 | 1, 0 | 0, 0 | 0, 0 | 1, 0 | 0, 0 | 1, 0 | 1, 1 | 1, 1 | 1, 1 | 0, 0 |
| MW-PSNR | 1, 0 | 1, 0 | 1, 0 | 1, 0 | 1, 0 | 0, 0 | 0, 0 | 1, 0 | 0, 0 | 1, 0 | 1, 0 | 1, 0 | 1, 0 | 0, 0 |
| RMW-PSNR | 0, 0 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1, 0 |
| LOGS | 1, 0 | 1, 1 | 1, 0 | 1, 0 | 1, 0 | 0, 0 | 0, 0 | 1, 0 | 0, 0 | 1, 0 | 1, 1 | 1, 1 | 1, 1 | 0, 0 |
| APT | 0, 0 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1, 0 | -1, 0 | 0, 0 | -1, 0 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1, 0 |
| NIQSV | 0, 0 | 0, 0 | 0,-1 | 0, 0 | 0,-1 | -1,-1 | -1, 0 | 0, 0 | -1,-1 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1,-1 |
| NIQSV+ | 0,-1 | 0, 0 | 0,-1 | 0,-1 | 1,-1 | -1,-1 | -1, 0 | 0, 0 | -1,-1 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1,-1 |
| MNSS | 0, 0 | 0, 0 | 0,-1 | 0, 0 | 0,-1 | -1,-1 | -1, 0 | 0, 0 | -1,-1 | 0, 0 | 0, 0 | 0, 0 | 0, 0 | -1,-1 |
| IDEA | 1, 0 | 1, 1 | 1, 0 | 1, 0 | 1, 0 | 0, 0 | 0, 0 | 1, 0 | 0, 0 | 1, 0 | 1, 1 | 1, 1 | 1, 1 | 0, 0 |

TABLE V
PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT SUPERPIXEL SEGMENTATION APPROACHES

| Method | IRCCyN/IVC Database | | | | IVY Database | | | |
|-----------|---------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| | PLCC | SRCC | KRCC | RMSE | PLCC | SRCC | KRCC | RMSE |
| Ref. [55] | 0.7457 | 0.6434 | 0.4758 | 0.3772 | 0.6119 | 0.5971 | 0.4203 | 19.4689 |
| Ref. [56] | 0.7514 | 0.6509 | 0.4819 | 0.3688 | 0.6211 | 0.5989 | 0.4286 | 19.3507 |
| SLIC | 0.7796 | 0.6652 | 0.4986 | 0.3533 | 0.6311 | 0.6132 | 0.4405 | 19.0379 |

TABLE VI
PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT EDGE DETECTION OPERATORS

| Method | IRCCyN/IVC Database | | | | IVY Database | | | |
|--------------|---------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| | PLCC | SRCC | KRCC | RMSE | PLCC | SRCC | KRCC | RMSE |
| Canny [57] | 0.7676 | 0.6614 | 0.4926 | 0.3598 | 0.6222 | 0.6011 | 0.4308 | 19.2867 |
| Prewitt [58] | 0.7409 | 0.6305 | 0.4661 | 0.3800 | 0.6116 | 0.5860 | 0.4119 | 19.5043 |
| Roberts [59] | 0.7461 | 0.6400 | 0.4719 | 0.3755 | 0.6201 | 0.5999 | 0.4297 | 19.4084 |
| Sobel | 0.7796 | 0.6652 | 0.4986 | 0.3533 | 0.6311 | 0.6132 | 0.4405 | 19.0379 |

TABLE VII
CONTRIBUTION OF EACH COMPONENT IN THE PROPOSED METHOD

| Component | IRCCyN/IVC Database | | | |
|---------------------|---------------------|--------|--------|--------|
| | PLCC | SRCC | KRCC | RMSE |
| Local Instance | 0.7013 | 0.6598 | 0.4788 | 0.4022 |
| Global Appearance | 0.6369 | 0.6590 | 0.4646 | 0.4349 |
| Local+Global (IDEA) | 0.7796 | 0.6652 | 0.4986 | 0.3533 |

| Component | IVY Database | | | |
|---------------------|--------------|--------|--------|---------|
| | PLCC | SRCC | KRCC | RMSE |
| Local Instance | 0.6205 | 0.5986 | 0.4272 | 19.4001 |
| Global Appearance | 0.6033 | 0.5859 | 0.4111 | 19.5974 |
| Local+Global (IDEA) | 0.6311 | 0.6132 | 0.4405 | 19.0379 |

TABLE VIII
RANKINGS OF DIBR TECHNIQUES ACCORDING TO THE DMOS VALUES AND THE SCORES PREDICTED BY FOURTEEN OBJECTIVE QUALITY METRICS

| Method | IRCCyN/IVC Database | | | | | | IVY Database | | | |
|----------|---------------------|----------|----------|----------|----------|----------|--------------|----------|----------|----------|
| | A1 | A5 | A4 | A6 | A2 | A3 | B1 | B2 | B3 | B4 |
| DMOS | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 |
| SSIM | 6 | 3 | 2 | 4 | 5 | 1 | 1 | 2 | 3 | 4 |
| VIF | 6 | 2 | 3 | 4 | 5 | 1 | 1 | 4 | 2 | 3 |
| FSIM | 6 | 3 | 2 | 1 | 5 | 4 | 1 | 3 | 2 | 4 |
| GSM | 6 | 3 | 2 | 1 | 5 | 4 | 1 | 2 | 3 | 4 |
| LTG | 6 | 5 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| MP-PSNR | 6 | 2 | 1 | 3 | 5 | 4 | 1 | 2 | 3 | 4 |
| MW-PSNR | 6 | 2 | 1 | 3 | 4 | 5 | 1 | 3 | 2 | 4 |
| RMW-PSNR | 6 | 2 | 1 | 3 | 4 | 5 | 1 | 3 | 2 | 4 |
| LOGS | 6 | 5 | 3 | 2 | 1 | 4 | 4 | 1 | 2 | 3 |
| APT | 1 | 5 | 3 | 6 | 2 | 4 | 3 | 1 | 4 | 2 |
| NIQSV | 1 | 3 | 2 | 5 | 4 | 6 | 2 | 1 | 4 | 3 |
| NIQSV+ | 1 | 3 | 4 | 2 | 5 | 6 | 1 | 4 | 3 | 2 |
| MNSS | 4 | 1 | 6 | 5 | 3 | 2 | 1 | 2 | 4 | 3 |
| IDEA | 1 | 3 | 2 | 4 | 5 | 6 | 1 | 2 | 3 | 4 |

views, it is time-consuming and labor-expensive. Fortunately, the IDEA method can be a good substitute. In this part, we evaluate the performance of the IDEA metric for benchmarking DIBR algorithms. To be specific, the performances of the DIBR algorithms are measured from two perspectives, i.e. 1) from human ratings of synthesized images (subjective score provided by the datasets), and 2) from the objective scores predicted by objective quality metrics. In implementation, the average of the the subjective scores for all the images generated by each DIBR algorithm is first averaged. Afterwards, the average subjective scores associated with all the DIBR algorithms are used to generate a ranking, which indicates the ground-truth performance ranking of the algorithms. A similar ranking can be calculated based on the objective scores predicted by each objective quality metric. Then the two rankings are compared. For a good quality metric, the latter ranking is expected to be consistent with the ground-truth ranking.

The experimental results are listed in Table VIII, where the objective rankings inconsistent with the ground truth are highlighted in bold. From the table, we can see that on both the

TABLE IX
PERFORMANCE OF THE PROPOSED IDEA METHOD AND THE STATE-OF-THE-ARTS ON BENCHMARKING DIBR ALGORITHMS. PERFORMANCE VALUES ARE CALCULATED BASED ON SYNTHESIZED IMAGES WITH THE SAME SCENE CATEGORY

| Method | Type | IRCCyN/IVC | | IVY | |
|---------------|------|---------------|---------------|---------------|---------------|
| | | SRCC | KRCC | SRCC | KRCC |
| SSIM | T | 0.5164 | 0.3868 | 0.8571 | 0.7619 |
| VIF | T | 0.4390 | 0.3114 | 0.4286 | 0.3333 |
| FSIM | T | 0.5853 | 0.4163 | 0.6286 | 0.5238 |
| GSM | T | 0.5391 | 0.3822 | 0.7143 | 0.6190 |
| LTG | T | 0.5204 | 0.3820 | 0.8571 | 0.8095 |
| MP-PSNR | S | 0.4756 | 0.3677 | 0.5714 | 0.4762 |
| MW-PSNR | S | 0.4460 | 0.3434 | 0.7429 | 0.6667 |
| RMW-PSNR | S | 0.3644 | 0.2823 | 0.6857 | 0.5714 |
| LOGS | S | 0.5506 | 0.4114 | 0.5429 | 0.4762 |
| APT | S | 0.4617 | 0.3636 | 0.4286 | 0.2857 |
| NIQSV | S | 0.3123 | 0.2242 | 0.6000 | 0.4286 |
| NIQSV+ | S | 0.3982 | 0.2879 | 0.4000 | 0.3333 |
| MNSS | S | 0.2594 | 0.1852 | 0.4857 | 0.3333 |
| Proposed IDEA | S | 0.6452 | 0.4747 | 0.8857 | 0.8095 |

IRCCyN/IVC and IVY databases, our IDEA method can obtain the best consistency with the ground truth rankings. This demonstrates the superiority of the IDEA method in benchmarking the performances of different DIBR algorithms.

Considering the fact that one DIBR algorithm may have different performances in synthesizing the views with diversified scenes, we also test the performance of the proposed method in benchmarking the DIBR algorithms in terms of scene category. Specifically, the performances of the quality metrics are first tested on each image group according to the scene category. Then the average value of each criterion on all the image groups is calculated as the final performance criteria. For this experiment, the accuracy criteria SRCC and the monotonicity criteria KRCC are adopted for performance comparison, as they denote the monotonicity consistency.

Table IX summarizes the performance of the proposed IDEA method and the existing quality metrics on benchmarking DIBR techniques, where the best results are marked in boldface. From the results, we can see that the IDEA method achieves the best performance on both databases. Specifically, on the IRCCyN/IVC DIBR database, the FSIM method obtains the highest SRCC and KRCC values among all existing metrics. However, they are still much lower than those of the IDEA method. On the IVY database, the FSIM method no longer performs the best. Instead, the LTG method obtains the best performance among all existing quality metrics, which is still inferior to the IDEA method. This experiment demonstrates the superiority and stability of the proposed metric.

Fig. 9 shows the difference between the SRCC value of the proposed IDEA metric and the existing metric \mathcal{R} on each scene

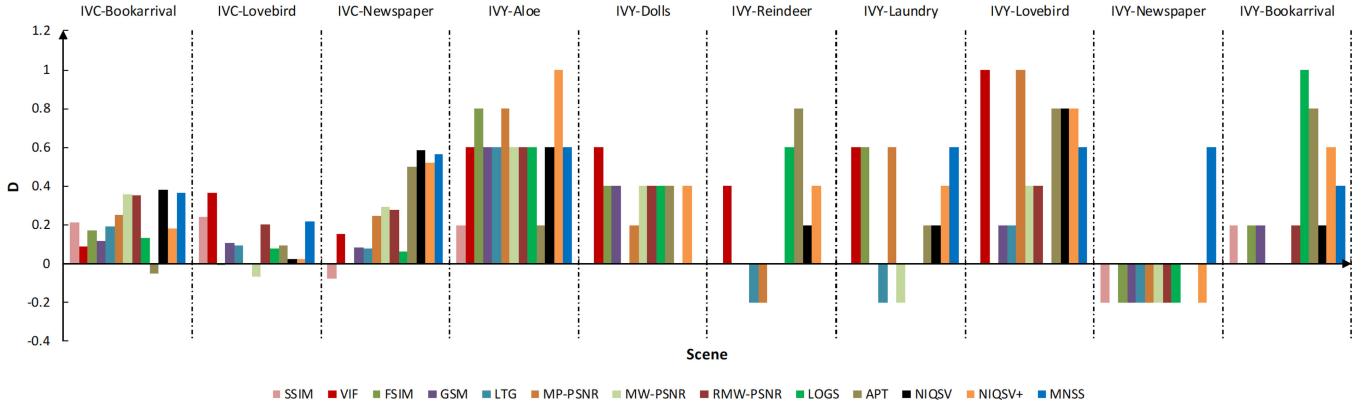


Fig. 9. The difference D between the SRCC value of the proposed method and the existing metric on each scene in the IRCCyN/IVC and IVY databases. The $Y-X$ above the figure indicates the X scene in the Y database.

group, which is denoted by D ,

$$D = \text{SRCC}_{\text{IDEA}} - \text{SRCC}_{\mathfrak{R}}. \quad (18)$$

It is easy to know that if the difference D on the $Y - X$ scene images is larger than zero, it means that the proposed metric performs better than metric \mathfrak{R} for the X scene in the Y database. If the D value is negative, metric \mathfrak{R} achieves better performance.

From Fig. 9, we know that the proposed method is superior to all the existing quality metrics on the IVY-Aloe, IVY-Dolls, IVY-Lovebird and IVY-Bookarrival scenes. For each of the three scenes of the IRCCyN/IVC database, only one metric obtains slightly better performance than the proposed method, but the best-performing metric on each scene shows obvious inferiority to the proposed method for the other two scenes. This demonstrates the consistent performance of the proposed method in benchmarking DIBR algorithms for diversified scenes. In IVY database, except for the Newspaper scene, the proposed metric achieves much better performance than the other metrics. Overall, the proposed IDEA metric achieves the best performance in benchmarking DIBR algorithms.

IV. CONCLUSION

In view synthesis, the synthesized images views are usually degraded by both local and global distortions. However, most existing view synthesis quality metrics are designed to measure a single aspect of the distortions. In this paper, we have proposed an effective view synthesis quality metric by simultaneously measuring local instance degradation and global appearance, which we call IDEA. The proposed local instance degradation module is achieved by first detecting the instance regions, followed by quantifying the instance shape change. The global appearance module is enlightened by the characteristics of the HVS in aggregating regions with similar characteristics for image perception. A key difference to the relevant state-of-the-art quality metrics is that the proposed metric is done in the instance level, which has more semantic information. Extensive experiments and comprehensive comparisons on two public view synthesis image quality databases have demonstrated the superiority of

the proposed method to the state-of-the-arts in both synthesized image quality assessment and benchmarking DIBR algorithms.

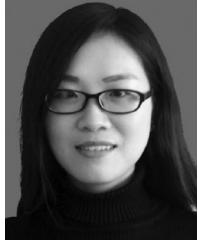
REFERENCES

- [1] F. Shao, Q. Z. Yuan, W. S. Lin, and G. Y. Jiang, "No-reference view synthesis quality prediction for 3-D videos based on color-depth interactions," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 659–674, Mar. 2018.
- [2] S. Li, C. Zhu, and M. T. Sun, "Hole filling with multiple reference views in DIBR view synthesis," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 1948–1959, Aug. 2018.
- [3] Y. Zhou *et al.*, "No-reference quality assessment for view synthesis using DOG-based edge statistics and texture naturalness," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4566–4579, Sep. 2019.
- [4] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3-D-TV," in *Proc. SPIE Conf. Stereoscopic Displays Virtual Reality Syst. XI*, 2004, pp. 93–104.
- [5] 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, 2009.
- [6] G. Luo, Y. Zhu, Z. Li, and L. Zhang, "A hole filling approach based on background reconstruction for view synthesis in 3 D video," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1781–1789.
- [7] C. Zhu and S. Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 82–93, Mar. 2016.
- [8] E. Bosc *et al.*, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J. Select. Top. Signal Process.*, vol. 5, no. 7, pp. 1332–1343, Nov. 2011.
- [9] 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.
- [10] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
- [15] M. A. Saad and A. C. Bovik, "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.
- [16] 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.

- [17] D. S. Stankovic, D. Kukolj, and P. L. Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. 3DTV Conf.*, 2015, pp. 1–4.
- [18] 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.
- [19] K. Gu *et al.*, "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.
- [20] K. Gu, S. Lee, H. T. Liu, W. S. Lin, and P. L. Callet, "Multiscale natural scene statistical analysis for no-reference quality evaluation of DIBR-synthesized views," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 127–139, Mar. 2020.
- [21] F. Battisti, E. Bosc, M. Carli, and P. L. Callet, "Objective image quality assessment of 3D synthesized views," *Sig. Process.: Image Commun.*, vol. 30, pp. 78–88, Jan. 2015.
- [22] Y. Zhou, L. D. Li, K. Gu, Y. M. Fang, and W. S. Lin, "Quality assessment of 3D synthesized images via disoccluded region discovery," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1012–1016.
- [23] G. Cermak, L. Thorpe, and M. Pinson, "Test plan for evaluation of video quality models for use with high definition TV content," *Video Qual. Experts Group*, Jun. 2009.
- [24] Y. B. Zhan, R. Zhang, and Q. Wu, "A structural variation classification model for image quality assessment," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1837–1847, Aug. 2017.
- [25] Y. Zhou *et al.*, "Blind quality index for multiply distorted images using border structure degradation and nonlocal statistics," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.
- [26] 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.
- [27] V. Jakhetiya *et al.*, "A highly efficient blind image quality assessment metric of 3D-synthesized images using outlier detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4120–4128, Jul. 2019.
- [28] S. 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.*, Jun. 2017, pp. 1248–1252.
- [29] Y. Zhou, L. D. Li, S. Y. Ling, and P. L. Callet, "Quality assessment of view synthesis using low-level and mid-level structural representation," *Signal Process.: Image Commun.*, vol. 74, pp. 309–321, 2019.
- [30] L. D. Li, Y. Zhou, K. Gu, W. S. Lin, and S. Q. 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.
- [31] S. S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV+: A no-reference synthesized view quality assessment metric," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1652–1664, Apr. 2018.
- [32] R. Achanta *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.
- [33] X. Jin and Y. Gu, "Superpixel-based intrinsic image decomposition of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4285–4395, Aug. 2017.
- [34] V. N. Gangapure, S. Nanda, and A. S. Chowdhury, "Superpixel based causal multisensor video fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1263–1272, Jun. 2018.
- [35] C. Peng, X. Gao, N. Wang, and J. Li, "Superpixel-based face sketch-photo synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 288–299, Feb. 2017.
- [36] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, Oct. 2017, pp. 2980–2988.
- [37] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vision*, Oct. 2014, pp. 740–755.
- [38] J. Flusser, T. Suk, and B. Zitova, *Moments and Moment Invariants in Pattern Recognition*. New York, NY, USA: Wiley, 2009.
- [39] R. Mukundan, S. H. Ong, and P. A. Lee, "Image analysis by Tchebichef moments," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1357–1364, Sep. 2001.
- [40] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph. Tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [41] K. Müller *et al.*, "View synthesis for advanced 3D video systems," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, 2009, Art. no. 438148.
- [42] P. Ndjiki-Nya *et al.*, "Depth image based rendering with advanced texture synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2010, pp. 424–429.
- [43] M. Köppel *et al.*, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 1809–1812.
- [44] S. Yoon, H. Sohn, Y. J. Jung, and Y. M. Ro, "Inter-view consistent hole filling in view extrapolation for multi-view image generation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 2883–2887.
- [45] ISO/IEC JTC1/SC29/WG11, "Reference software for depth estimation and view synthesis," Doc. M15377, Archamps, France, Apr. 2008.
- [46] I. Ahn, and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 614–626, Dec. 2013.
- [47] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
- [48] 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.
- [49] X. K. Min *et al.*, "Blind quality assessment based on pseudo-reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.
- [50] Y. B. Zhan and R. Zhang, "No-reference JPEG image quality assessment based on blockiness and luminance change," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 760–764, Jun. 2017.
- [51] L. D. Li *et al.*, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
- [52] G. C. Wang *et al.*, "Blind quality metric of DIBR-synthesized images in the discrete wavelet transform domain," *IEEE Trans. Image Process.*, vol. 29, pp. 1802–1814, Jan. 2020.
- [53] Y. M. Fang *et al.*, "Learning a no-reference quality predictor of stereoscopic images by visual binocular properties," *IEEE Access*, vol. 7, pp. 132649–132661, Sep. 2019.
- [54] Y. Zhou, L. D. Li, S. Q. Wang, J. J. Wu, and Y. Zhang, "No-reference quality assessment of DIBR-synthesized videos by measuring temporal flickering," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 30–39, Aug. 2018.
- [55] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vision*, Sep. 2010, pp. 211–224.
- [56] Y. H. Zhang, R. Hartley, J. Mashford and S. Burn, "Superpixels via pseudo-boolean optimization," in *Proc. Int. Conf. Comput. Vision*, Nov. 2011, pp. 1387–1394.
- [57] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
- [58] E. Onat, "FPGA implementation of real time video signal processing using sobel, roberts, prewitt and laplacian filters," in *Proc. 25th Signal Process. Commun. Appl. Conf.*, 2017, pp. 1–4.
- [59] L. G. Roberts, "Machine perception of three-dimensional solids," *Opt. Electro-Opt. Inf. Process.*, vol. 11, pp. 159–197, 1965.



Leida Li (Member, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xian, China, in 2004 and 2009, respectively. In 2008, he was a Research Assistant with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. From 2009 to 2019, he worked in the School of Information and Control Engineering, China University of Mining and Technology, as an Assistant Professor, an Associate Professor, and a Professor, respectively. He is currently a Professor with the School of Artificial Intelligence, Xidian University. His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as SPC for IJCAI 2019–2020, Session Chair for ICMR 2019, and PCM 2015, and TPC for AAAI 2019, ACM MM 2019–2020, ACM MM-Asia 2019, ACII 2019, PCM 2016. He is currently an Associate Editor for the *Journal of Visual Communication and Image Representation* and the *EURASIP Journal on Image and Video Processing*.



Yu Zhou received the B.Sc. and Ph.D. degrees from the China University of Mining and Technology, Xuzhou, China, in 2014 and 2019 respectively. She is currently an Assistant Professor with the School of Information and Control Engineering, China University of Mining and Technology. Her research interests include multimedia quality assessment and perceptual image processing.



Fu Li received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2010, respectively. From 2010 to 2011, he was with Microsoft Research Asia, Beijing, China. He is currently a Professor with the School of Artificial Intelligence, Xidian University. His research interests include depth sensing, nonlinear signal processing, high efficiency video coding architecture design, and very large scale integration design.



Jinjian Wu received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore, where he was a Postdoctoral Research Fellow from 2013 to 2014. From 2015 to 2019, he was an Associate Professor with Xidian University, where he had been a Professor since 2019. His research interests include visual perceptual modeling, biomimetic imaging, quality evaluation, and object detection. He was the recipient of the Best Student Paper Award at the ISCAS 2013. He was an Associate Editor for the *Journal of Circuits, Systems and Signal Processing*, the Special Section Chair for the IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP) 2017, and the Section Chair/Organizer/TPC member for the ICME 2014–2015, PCM 2015–2016, ICIP 2015, VCIP 2018, and AAAI 2019.



Guangming Shi (Senior Member, IEEE) received the B.Sc. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. He was a Visiting Professor at the University of Illinois and University of Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University. He is currently the Academic Leader on circuits and systems, Xidian University. He has authored and coauthored more than 200 papers in journals and conferences. His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and implementation of algorithms for intelligent signal processing. He was awarded the Cheung Kong Scholar Chair Professor by the Ministry of Education in 2012. He was the Chair for the 90th MPEG and 50th JPEG of the international standards organization, and Technical Program Chair for FSKD06, VSPC 2009, IEEE Pulse Code Modulation 2009, SPIE Visual Communications and Image Processing 2010, and IEEE International Symposium on Circuits and Systems 2013.