Multiscale Natural Scene Statistical Analysis for No-Reference Quality Evaluation of DIBR-Synthesized Views

Ke Gu, Junfei Qiao, *Member, IEEE*, Sanghoon Lee, *Senior Member, IEEE*, Hantao Liu, *Member, IEEE*, Weisi Lin, *Fellow, IEEE*, and Patrick Le Callet, *Fellow, IEEE*

***Abstract*—This paper proposes to blindly evaluate the quality of images synthesized via a depth image-based rendering (DIBR) procedure. As a significant branch of virtual reality (VR), superior DIBR techniques provide free viewpoints in many real applications, including remote surveillance and education; however, limited efforts have been made to measure the performance of DIBR techniques, or equivalently the quality of DIBR-synthesized views, especially in the condition when references are unavailable. To achieve this aim, we develop a novel blind image quality assessment (IQA) method via multiscale natural scene statistical analysis (MNSS). The design principle of our proposed MNSS metric is based on two new natural scene statistics (NSS) models specific to the DBIR-synthesized IQA. First, the DIBR-introduced geometric distortions damage the local self-similarity characteristic of natural images, and the damage degrees of self-similarity present particular variations at different scales. Systematically combining the measurements of the variations mentioned above can gauge the naturalness of the input image and thus indirectly reflect the quality changes of images generated using different DIBR methods. Second, it was found that the degradations in main structures of natural images at different scales remain almost the same, whereas the statistical regularity is destroyed in the DIBR-synthesized views. Estimating the deviation of degradations in main structures at different scales between one DIBR-synthesized image and the statistical model, which is constructed based on a large number of natural images, can quantify how a DIBR method damages the main structures and thus infer the image quality. Via trials, the two NSS-based features extracted above can well predict the quality of DIBR-synthesized images. Further, the two features come from distinct points of view, and we hence integrate them via a straightforward multiplication to derive the proposed blind MNSS metric, which achieves better performance than each component and state-of-the-art quality methods.**

***Index Terms*—Depth image-based rendering (DIBR), image quality assessment (IQA), blind/no-reference (NR), multiscale natural scene statistical analysis.**

I. Introduction

IN NUMEROUS practical systems such as remote education and surveillance, medical and entertainment applications, free viewpoint videos (FVVs) and the relevant technologies play extremely critical roles. On the basis of the depth image-based rendering (DIBR) techniques, new frames are created from existing adjacent frames. These techniques provide users with more flexible selection of direction and viewpoint, which allows users a wholly new generation of the scene via some rendering technologies and thus largely decreases the cost and complexity of camera setup [1]. When producing new frames that were once totally nonexistent, distortions (specifically geometric distortions) are inevitably introduced during this process [2]. Thus, a reliable quality evaluation method is highly desirable.

Quality evaluation techniques for various image and video applications have aroused much attention [3]–[7]. Structural computational models of HVS have been proposed to adequately solve the problems of image postprocessing and quality assessment [8]–[11]. In the last several decades, most of works were devoted to the typical artifacts, such as blur, noise, contrast change, and compression [12]–[19]. By contrast, the DIBR-introduced geometric distortion causes a considerably different visual degradation, as shown in Fig. 1. More specifically, the DIBR method is used to deploy a view and the corresponding per-pixel depth information to synthesize 'virtual' views from a slightly distinct viewing perspective. Generally, a new view is created along with 1) reprojecting the texture information into the 3D world via the per-pixel depth information and 2) producing the 'virtual' camera by projecting the 3D space points into the 2D image plane.

Manuscript received November 15, 2018; revised January 31, 2019; accepted February 7, 2019. This work was supported in part by the National Science Foundation of China under Grant 61703009 and Grant 61890930-5, in part by the China Association for Science and Technology through the Young Elite Scientist Sponsorship Program under Grant 2017QNRC001, in part by the Beijing Excellent Talents Funding through the Young Top-Notch Talents Team Program under Grant 2017000026833ZK40, in part by the Major Science and Technology Program for Water Pollution Control and Treatment of China under Grant 2018ZX07111005, and in part by the National Key Research and Development Project under Grant 2018YFC1900800-5. *(Corresponding author: Ke Gu.)*

K. Gu and J. Qiao are with the Beijing Advanced Innovation Center for Future Internet Technology, Faculty of Information Technology, Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com; junfeiq@bjut.edu.cn).

S. Lee is with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea (e-mail: slee@yonsei.ac.kr).

H. Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, U.K. (e-mail: liuh35@cardiff.ac.uk).

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

P. Le Callet is with IRCCyN UMR CNRS, Polytech Nantes, Luman Université, Université de Nantes, 6597 Nantes, France (e-mail: patrick.lecallet@univ-nantes.fr).

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/TBC.2019.2906768

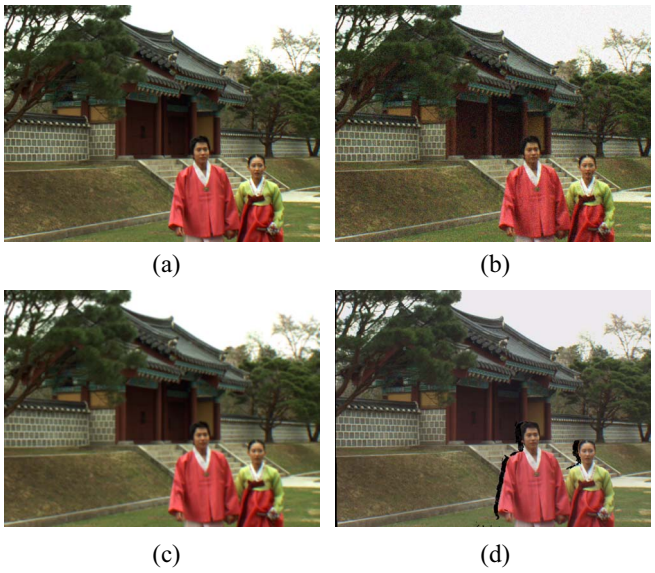


Fig. 1. Illustration of the differences between typical distortions (white noise and Gaussian blur) and the geometric distortion. (a) Reference. (b) White noise. (c) Gaussian blur. (d) Geometric distortion.

Based on the above process, the key target of DIBR is to transfer the occluded regions (mainly occurring at the contour of the foreground objects) in the original view to be visible in the 'virtual' view [2]. The depth information can be used to solve the occlusion problem to some extent but inevitably introduces geometric distortions due to the imperfect 3D reprojection [20], [21]. There exist the two main differences between geometric distortions and the majority of conventional distortions. The former represent one kind of local distortion and always occur at the contours of foreground objects, while the latter belong to the class of global distortions and often randomly appear anywhere. It is very likely that geometric distortions can more easily devastate the semantic structures and thus lead to a stronger influence on the image quality than typical artifacts. Experiments demonstrate that many IQA models, regardless of whether they are full reference (FR) [22]–[25] or no reference (NR) [26]–[29], cannot yield reliable quality scores when assessing DIBR-synthesized views. One possible reason why the IQA models mentioned above work ineffectively is that their design principles mainly rely on classical low-level visual features or their variants, which are sensitive to the typical distortions, whereas the new high-level features specific to the geometric distortion are not included.

Although there are many IQA models, limited efforts have been devoted to the explorations of DIBR-synthesized views to date. In [30], the view synthesis quality assessment (VSQA) method was designed by exploiting the impact of orientation, texture and contrast to modify the distortion/similarity map, which is computed from a synthesized view and its corresponding reference views. In [31], the 3D-synthesized view image quality metric (3D-SWIM) method was developed to conduct statistical comparisons of Haar wavelet subbands of the matched blocks between the reference and DIBR-synthesized images. In [32], [33], the morphological wavelet peak signal-to-noise ratio (MW-PSNR) and morphological pyramid peak signal-to-noise ratio (MP-PSNR) for evaluating the quality of 3D-synthesized views were developed; these metrics are based on morphological wavelet decomposition and morphological pyramid decomposition, respectively, followed by mean squared error pooling at several scales. In [34], a reduced version of MP-PSNR (MP-PSNR-RR) was proposed by merely leveraging detailed images from higher pyramid scales for visual quality estimation, which yields superior performance and less computational complexity compared with the original method. The aforementioned IQA metrics were designed specifically for 3D-synthesized image quality evaluation based on the entire or partial information of benchmark reference images, and this fact very possibly restricts their application scopes since original camera-capture images that correspond to the 'virtual' DIBR-synthesized images are usually not accessible in FVV systems. Furthermore, the camera-captured view and its associated depth map might be also distorted, and under such conditions, the quality of the above two maps would be importantly used to help evaluate the quality of a DIBR-synthesized image in terms of, e.g., texture and color distortions.

To address the above-mentioned problem, we directly develop a novel NR-IQA model used to blindly predict the quality of DIBR-synthesized views. The proposed blind IQA model is devised based on multiscale natural scene statistical analysis; hereinafter, we refer to it as MNSS. The design theory underlying our MNSS model relies on two newly explored natural scene statistics (NSS) models specific to the quality measurement of DIBR-synthesized images. The first NSS model originates from the subsequent two observations: one is that the DIBR-introduced geometric distortion damages the local self-similarity attribute of natural images; the other is that the damage degrees of self-similarity present particular variations at different scales. Through a valid combination of damage degrees at distinct scales, the first NSS-based feature can measure the naturalness of a DIBR-synthesized image and accordingly indirectly generate a reliable estimation of its quality. The second NSS model is built upon a prior that the statistical regularity of degradations in main structures at different scales, which come from natural images, will be evidently broken by DIBR methods. Estimating the distance of degradations in main structures at distinct scales between a DIBR-synthesized view and the statistical model constructed using massive natural images can reflect how a DIBR technology damages the main structures and is therefore adopted to predict the image quality. The two NSS-based features above were validated in the experimental results of a good ability to accurately assess the quality of DIBR-synthesized views. Considering that the two features are extracted from different perspectives, direct multiplication is thus applied to the two features to derive our NR MNSS metric.

The structure of the paper is outlined as follows. Section II presents the newly established two NSS models specifically devoted to DIBR-synthesized views and the associated blind MNSS metric. Section III conducts comparisons of our blind metric with state-of-the-art FR, RR and NR IQA methods.

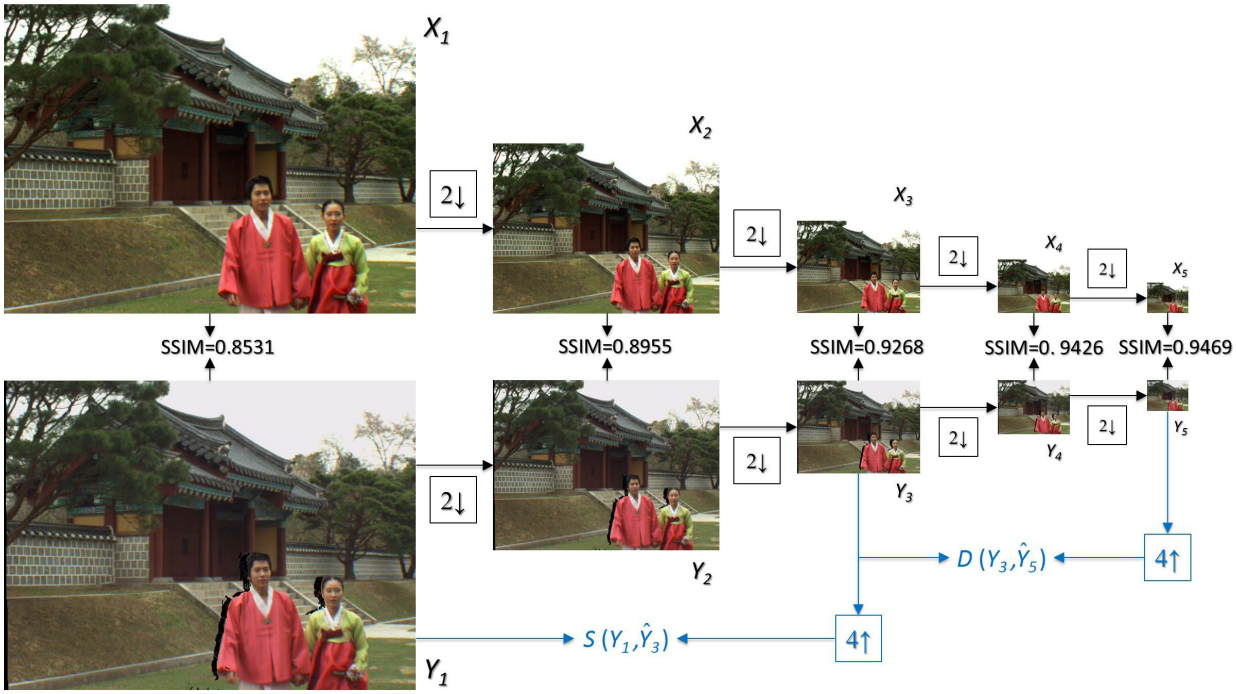


Fig. 2. Comparison of reference and DIBR-synthesized images at multiscales. 2↓ : downsampling by 2. 4↑ : upsampling by 4.

Section IV extends the MNSS to video quality assessment. Section V concludes the whole paper.

II. NSS-Based Blind Quality Assessment

In this section, we will first introduce the two novel NSS models, which are specific to DIBR-synthesized images, and then propose the blind IQA algorithm by fusing two features extracted based on the above two NSS models.

A. Self Similarity-Based NSS Model

As for natural images, the local self-similarity characteristic is an essential attribute and it has been broadly employed to many applications such as image description and compression [35], [36]. One can see from Fig. 1 that, in contrast to white noise or Gaussian blur which affects the global image self-similarity attribute, the geometric distortion merely changes the self-similarity characteristic in some typical local regions and meanwhile makes no influences on other areas. When we conduct the multiscale analysis on a reference image X and its associated DIBR-synthesized image Y, taking Fig. 1(a) and (d) for example, the distance between X and Y rapidly decreases with the scale reduced, as presented in Fig. 2. The top row refers to the reference image and the bottom row refers to the DIBR-synthesized image. The SSIM (Structural SIMilarity) values [12] accurately reflect the variation tendency stated above.[1] If the reference image X_i at the i-th down scaled version from X, where $i = \{1, 2, \ldots, 5\}$, is known, we can simply measure each distance between X_i and Y_i, denoted as $D(Y_i, X_i)$, and combine the five measures to infer the overall quality score of the DIBR-synthesized image Y. Note that X_1 and Y_1 are actually X and Y, respectively.

But in real applications, reference free viewpoint images are usually unavailable, so the DIBR technique is needed to generate virtual free viewpoint images. In other words, blind quality assessment of the DIBR-synthesized image is more close to the practical application scenarios. A straightforward way to address the above problem is to find an approximate alternative as reference. From Fig. 2, we find that the corrupted image Y_5 and the associated reference image X_5 have a very high SSIM value, which means they have a close distance and Y_5 can be approximately used as reference. Consider X_3 for illustration. We can derive the subsequent approximate relationship $D(Y_3, X_3) = D(Y_3, \hat{X}_5) \approx D(Y_3, \hat{Y}_5)$, where '$\hat{}$' is a perfect upsampling operator which is supposed to totally recover the detailed information lost owing to downsampling and keep the two inputs having the matchable size.[2] Here, $\hat{Y}_5$ is the upscaled version of Y_5 to the scale 3. Hence, we can compute $D(Y_i, \hat{Y}_5)$ to approximate $D(Y_i, X_i)$, where $\hat{Y}_5$ is the upscaled version of Y_5 to the scale i, $i = \{1, 2, \ldots, 5\}$.

We further treat the above issue from another viewpoint. $X_1, X_2, \ldots, X_5$ are reference images with the perfect quality. According to the SSIM value, we can derive the following quality rank: $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$. For further analysis, we fix Y_1 and Y_5 and set Y_i as a variable, where $i = \{1, 2, 3, 4, 5\}$. Suppose $S(Y_1, \hat{Y}_i)$ be a similarity measure between Y_1 and an upsampled version of Y_i to the scale 1. When $D(Y_i, \hat{Y}_5)$ rises, the similarity between Y_1 and $\hat{Y}_i$, i.e., $S(Y_1, \hat{Y}_i)$, increases, and vice versa. That is to say, $S(Y_1, \hat{Y}_i)$

[1]The greater SSIM value indicates the higher similarity and the closer distance between the two compared image signals.

[2]In general, we consider $D(Y_j, \hat{Y}_i)$ for illustration. $\hat{Y}_i$ is upsampled from the scale i to j and thus it has the same size with Y_j.

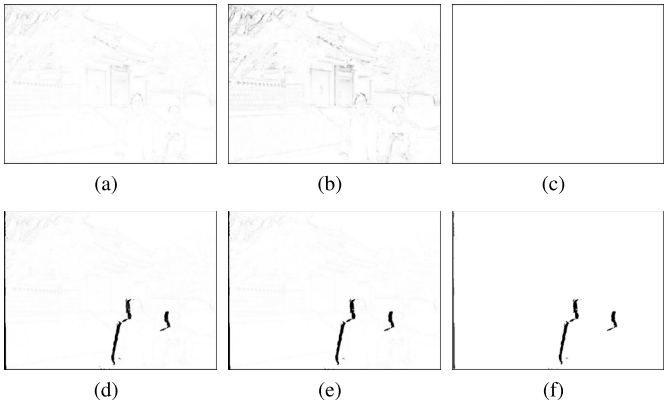


Fig. 3. Illustration of similarity maps: (a)-(c) $\bar{S}$, $\dot{S}$ and $\ddot{S}$ maps of the reference image Fig. 1(a); (d)-(f) $\bar{S}$, $\dot{S}$ and $\ddot{S}$ maps of the DIBR-synthesized image Fig. 1(d).

has the same changing trend with $D(Y_i, \hat{Y}_5)$. In the IQA, we pay more attention to the monotonicity between objective predictions and subjective ratings, e.g., SROCC [37], rather than precise estimations of subjective ratings. Therefore, we can consider $S(Y_1, \hat{Y}_i)$ to be an alternative of $D(Y_i, \hat{Y}_5)$ (or $D(Y_i, X_i)$), where $i = \{1, 2, 3, 4, 5\}$. Results verify that using $S(Y_1, \hat{Y}_i)$ is able to bring about an about 15% performance gain (in terms of SROCC) beyond using $D(Y_i, \hat{Y}_5)$. This phenomenon may be due to the use of SSIM-form measure instead of other distance measures or Y_1 has larger resolution and richer details as compared with Y_5. Hence, in this work we use the similarity measure.

Multiscale analysis is an important attribute of the human visual system (HVS), which has been widely used in many image processing fields, for example, quality evaluation [38] and saliency detection [39]. Following the multiscale analysis in [38], we fuse each similarity map and derive

$$\bar{S}_j = \prod_{i=1}^{N} [S_j(Y_1, \hat{Y}_i)]^{\gamma_i} \tag{1}$$

where $N = 5$ and j indicates the pixel index; $\{\gamma_1, \gamma_2, \ldots, \gamma_5\}$ are assigned in accordance to a psychophysical experiment as $\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ [38]. We employ the commonly used similarity metric (not SSIM) with three merits of symmetry, boundedness and unique maximum to be the distance measure:

$$S(Y_1, \hat{Y}_i) = \frac{2Y_1 \cdot \hat{Y}_i + \varepsilon}{Y_1^2 + \hat{Y}_i^2 + \varepsilon} \tag{2}$$

where ε is a constant number for avoiding division-by-zero. Note that Equation (2) is bounded belonging to the range of [0, 1] and it reaches to 1 when two inputs are exactly the same. Observing Equation (1), $\hat{Y}_1$ is indeed Y_1. So we can simplify it as

$$\bar{S}_j = \prod_{i=2}^{N} [S_j(Y_1, \hat{Y}_i)]^{\gamma_i}. \tag{3}$$

In the DIBR-synthesized image, e.g., Fig. 1(d), there exist some isolated noisy pixels, which have little influence on the quality perception. So we apply a small size median filter to

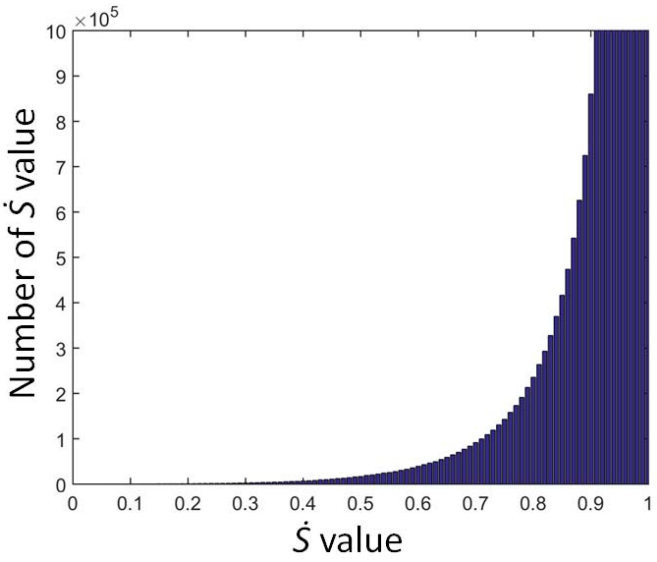


Fig. 4. Distribution of $\dot{S}$ values of about 50 million pixels from 300 natural images. The number of $\dot{S}$ values greater than 0.1 accounts for about 99.85% of all the pixels.

$\bar{S}$ and generate $\dot{S}$ in order to remove isolated noisy pixels. Furthermore, even for natural images, distortions that do not originally occur will be included into the fused map $\bar{S}$ since the bilinear interpolation method is not perfect upsampling technology and must introduce blur distortions. Therefore, we exert a thresholding on the filtered map $\dot{S}$ for extracting geometric distortion regions:

$$\ddot{S}_j = \begin{cases} 0, & \text{if } \dot{S}_j < T \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where T is a small constant threshold determined based on a large number of natural images. The newly generated blur distortions are beyond this threshold and therefore removed. By contrast, the geometric distortion regions are beneath this threshold and thus preserved. In Fig. 3, we show $\bar{S}$, $\dot{S}$ and $\ddot{S}$ of the reference and DIBR-synthesized views (i.e., Fig. 1(a) and (d)). Comparing Fig. 3(c) and (f), one can easily see that $\ddot{S}$ well extracts the geometric distortion areas.

Determining the T value depends on a new NSS regularity. Specifically, we randomly selected 300 natural images with high quality from the Berkeley image segmentation database [40] and compute their $\dot{S}_j$ maps. We assume that no geometric distortion regions are included in most of the chosen natural images, or in other words, the majority of (about 50 million) pixels should be higher than T. We plot the histogram of $\dot{S}$ values of all the pixels in Fig. 4. One can clearly see that the number of $\dot{S}$ values beyond 0.1 possesses about 99.85% of all the pixels and we thus approximately set T as 0.1.

Eventually, we blindly extract the first NSS-based feature for estimating the quality score of an input DIBR-synthesized image as

$$Q_1 = \frac{1}{L} \sum_{j=1}^{L} \ddot{S}_j \tag{5}$$

where L stands for the number of all the pixels in $\ddot{S}$. It is worthy to emphasize that the greater Q_1 value refers to less

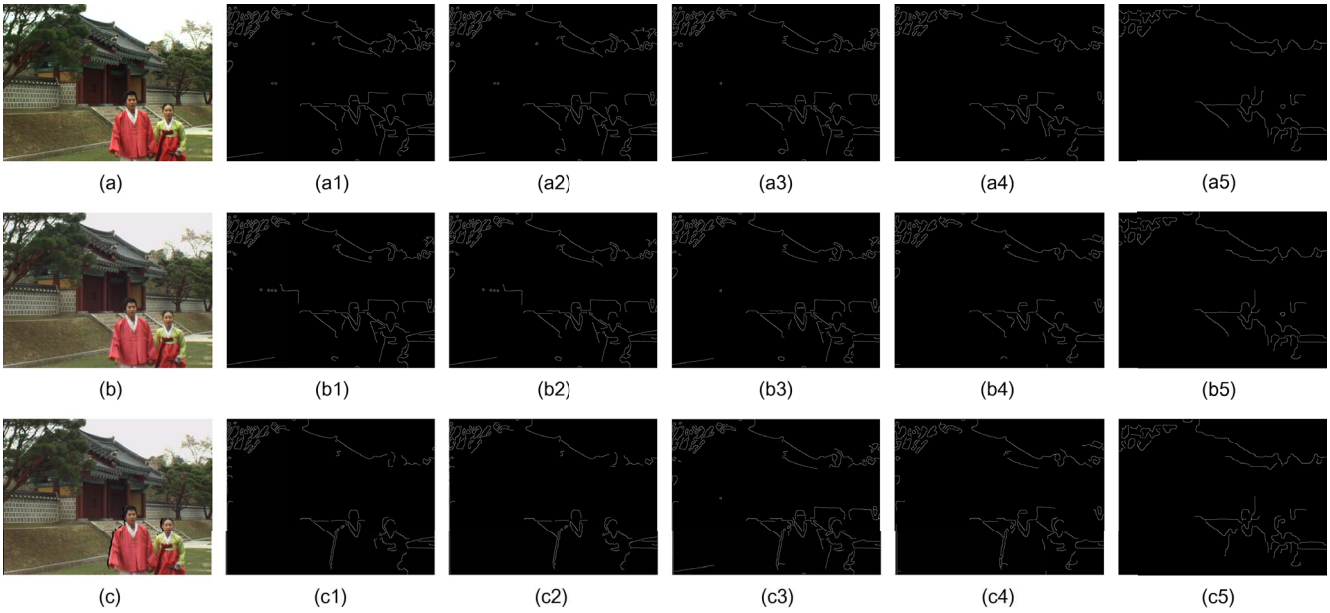


Fig. 5. Comparison of variations in main structures at multiscales: (a)-(c) camera-captured original image and two DIBR-synthesized images; (a1)-(a5), (b1)-(b5) and (c1)-(c5) main structures of (a), (b) and (c) at multiscales.

geometric distortion and larger subjective mean opinion score (MOS), and vice versa.[3]

B. Main Structure Consistency-Based NSS Model

Structures in images convey crucial visual information and are significant to scene analysis and understanding. Structure maps based on various measurements, such as covariance and gradient, were used in several classical IQA models, such as structural similarity index (SSIM) [12], feature similarity index (FSIM) [22], gradient magnitude standard deviation (GMSD) [23], etc, were developed by comparing the distance of structures between the reference and corrupted images. Damaging the structures, particularly the main structures, e.g., contours, leads to a considerable influence on visual quality of images and may even result in the loss of semantic information [42]. The most critical distortion of DIBR-synthesized images is the geometric distortion. During image rendering, occlusion is an inevitable problem because limited cameras are used for capturing views. This often creates some holes (i.e., geometric distortion) in the DIBR-synthesized images, especially at the contour of foreground objects, even though advanced inpainting and interpolation techniques are applied for repairing the holes. For illustration consider a representative sample shown in Fig. 1(d). It is evident that the DIBR-introduced geometric distortion damages the main structures of foreground objects (namely the contour of two persons).

We can simply make a direct comparison between a DIBR-synthesized image and its reference counterpart in terms of consistency of main structures, if possible. But reference free viewpoint images cannot be accessible in most cases due to the fact that there are only limited cameras for view capture. As thus, a new NSS model should be carefully developed for capturing the consistency of main structures. By comparing the difference of main structures of a DIBR-synthesized view with the NSS model established based on a large quantity of high-quality natural images, we can estimate the degradation level of main structures which is caused by the DIBR method and thereby infer the visual quality of the DIBR-synthesized image. The new NSS model was built based on the following observations and approximations.

First of all, we find that the distribution of main structures between a DIBR-synthesized image and its associated reference one converge to the consistency as the scale reduces. As shown in Fig. 5, (a)-(c) in order indicate a reference image and the two corresponding views synthesized by two typical DIBR methods, while, (a1)-(c5) provide the main structures computed by exploiting the well-known Canny edge operator at five scales. Owing to the disappearance of massive edges with the scale reduced, for a fair comparison, we adopt the popular and efficient bilinear upsampling method to rescale the downsampled image to its original size before finding the main structures. As seen, despite the noticeable difference in the main structure maps of (a)-(c) at the first original scale, i.e., (a1)-(c1), the three main structure maps at the last fifth scale, i.e., (a5)-(c5), are quite similar to each other. The distribution of main structures converges to the same as the scale reduces. By comparison, we can also find that, at the first scale, some differences in the main structure maps are caused due to the geometric distortion, for example, around the contour of the man's head, and these differences tend to disappear with the scale decreased.

With the above concern, the difference of degradations in main structures at multiscales between the DIBR-synthesized and reference images can be used to estimate the distortion

[3]Part of this subsection was described in [41]. How to determine T is a new contribution of this paper.

intensity and even the visual quality of the DIBR-synthesized view. A simple solution is to directly compare the difference of main structure maps at five scales, but there exists the mismatch problem caused by occlusion and thus a warping operation is needed for solving it. Of course, in this regard, the reference image must be available for adjusting the DIBR-synthesized view to make them have a well match; otherwise, using the pixel-based difference measure is meaningless. Our terminal target is to devise an IQA model without using any reference information, so we instead resort to another solution. More concretely, we compute the similarity of degradations[4] (not SSIM) in main structures between the DIBR-synthesized image and its reference one as the distance measure or visual quality:

$$Q_2(Y,X)=\frac{1}{U}\sum_{u=1}^{U}\frac{2\mathbf{m}_Y(u)\cdot\mathbf{m}_X(u)+\epsilon}{\mathbf{m}_Y(u)^2+\mathbf{m}_X(u)^2+\epsilon} \tag{6}$$

where ϵ is a very small constant to remove the big influence of division-by-zero; u indicates the scale where the similarity is measured; $U=5$ represents five scales; $\mathbf{m}_Y$ is a vector of degradations in main structures of Y, comparing the similarity of the two main structure maps at the U-th scale and the u-th scale defined by

$$\mathbf{m}_Y(u)=\sum_{v=1}^{V}M_U(v)\oplus M_u(v) \tag{7}$$

where M_u stands for the u-th main structure map; $\oplus$ is the XOR operator; v is the index of pixels in M_u; V means the number of all the pixels. There are only two values 1 or 0 at the location v, which separately mean whether or not a main edge is included. $\mathbf{m}_X$ has the similar definition for X. To make Q_2 have a good range of $[0,1]$, we separately normalize $\mathbf{m}_X$ and $\mathbf{m}_Y$ before Equation (6) is implemented.

Observing Equation (6), we easily find that the calculation of Q_2 still depends on partial information from the reference image, namely $\mathbf{m}_X$, which does not reach to the ideal no-reference condition even though only 5 numbers in $\mathbf{m}_X$ are required.[5] Therefore the next step is to find an alternative to approximate and thus replace $\mathbf{m}_X$, making the computation of Q_2 totally blind. We also used those 300 natural images from the Berkeley segmentation database [40] to obtain their associated vectors $\mathbf{m}_X$. For a vector $\mathbf{m}_X$, we can provide a 5-point curve. We plot the 300 curves together in Fig. 6 for easy comparison. We label five points of each curve with the color of blue and use five circles for highlighting. One can see that there exists an approximating statistical regularity in Fig. 6; or in other words, those 300 curves are close to each other and we may find a general five-point curve as prior instead of extracting from the reference image. As thus, we compute the median value of 300 values at five scales and derive the NSS vector $\mathbf{m}_P=\{1, 0.9919, 0.9520, 0.8108, 0\}$, namely the black dash line shown in Fig. 6. We provide four images for comparison and plot their curves. We label their quality scores

[4]Here we use the similarity measure to compare the differences for making the quality Q_2 value have the same order with Q_1.

[5]In fact, only 3 numbers are needed since the first and last numbers must be 1 and 0.

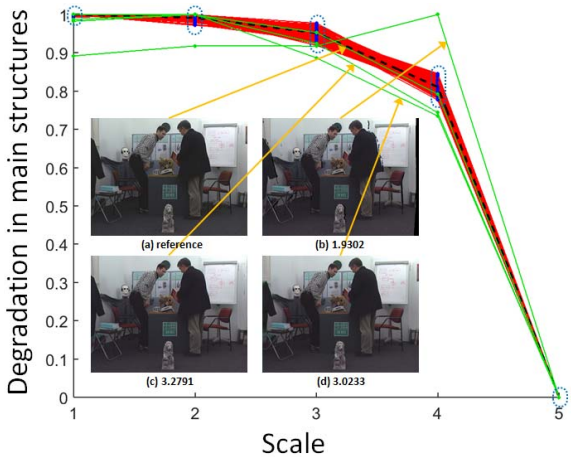


Fig. 6. Plot of the curves of degradations in main structures of 300 natural images at five scales. The black dash line is associated to the NSS vector.

in Fig. 6. As compared with the reference image (a), (b) has obvious black holes while (c) and (d) try to fill the holes. The difference between (c) and (d) is that the former one fills the holes better than the latter one. So, we can derive the quality rankings: (a) > (c) > (d) > (b), which is consistent with their quality scores. One can see from Fig. 6 that the better-quality image has the curve closer to the black dash line. In Fig. 6, how much difference can be found according to scale due to the geometric distortion of synthetic images from the NSS. On this basis, we are able to compute Q_2 in the completely no-reference condition by modifying Equation (6) to be

$$Q_2(Y)=\frac{1}{U}\sum_{u=1}^{U}\left(\frac{2\mathbf{m}_Y(u)\cdot\mathbf{m}_P(u)+\epsilon}{\mathbf{m}_Y(u)^2+\mathbf{m}_P(u)^2+\epsilon}\right)^{\gamma_u} \tag{8}$$

where $\{\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5\}$ are also set to be $\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$, akin to Equation (1). Note that the large Q_2 value means that the DIBR-synthesized image Y is close to its associated reference image X (or natural images) and therefore has high quality.

C. NSS-Based Blind Image Quality Estimation

We have developed two novel NSS models specific to the DIBR-synthesized IQA and obtained the two corresponding quality scores. Note that the first self similarity-based NSS model is based on the local characteristics of natural images for mainly detecting holes, while, the second main structure consistency-based NSS model is based on the global attributes of natural images for tackling the condition when holes are filled with similar values to neighboring pixels.According to this analysis, these two models may play complementary roles and thus they can be systematically incorporated together to generate a better blind quality assessment model. To specify, Q_1 and Q_2 have the consistent order and belong to $[0,1]$. So we can introduce a straightforward multiplication to combine them, which makes their weighted product still belongs to the range of $[0,1]$, and derive the final quality index Q_{MNSS} of the proposed blind MNSS metric:

$$Q_{\text{MNSS}}=Q_1^{\phi}\cdot Q_2 \tag{9}$$

where ϕ is a fixed positive weighting coefficient to balance the relative contribution of each component. Akin to Q_1 and Q_2, the larger the Q_{MNSS} value is, the higher quality the input DIBR-synthesized view has. More performance comparisons of Q_1, Q_2 and Q_{MNSS}, as well as the influences of stable constants on the performance of Q_{MNSS} will be discussed in the next section. And furthermore, in this work we employ the classical bilinear method and more advanced upsampling technologies will be considered in the future.

III. Validations and Discussions

This section concentrates on measuring and comparing the performance of the proposed MNSS metric with state-of-the-art IQA approaches on the image database specific to DIBR-synthesized IQA.

A. Experimental Protocol

Testing database: The IRCCyN/IVC database [2], which was established in the year of 2011 specifically for the DIBR-synthesized IQA task, is applied for examining the effectiveness of our proposed blind MNSS model. There include 96 images in the IRCCyN/IVC database, 12 reference images and their corresponding 84 DIBR-synthesized images based on 7 different DIBR methods. The absolute category rating-hidden reference (ACR-HR) method [43] is used for conducting the subjective test, and each observer scores the test item with a discrete category rating scale. Since this study is devoted to developing NR-IQA methods, we only focus on those 84 DIBR-synthesized views, which are mainly distorted by the geometric distortions.

Quality models: Up to twenty IQA models were collected for comparison. The first type is composed of seven FR IQA models, which are SSIM [12], visual signal-to-noise ratio (VSNR) [44], most apparent distortion (MAD) [13], information weighted SSIM (IW-SSIM) [45], feature similarity (FSIM) [22], gradient magnitude standard deviation (GMSD) [23], and perceptual similarity (PSIM) [42]. The second type includes three reduced-reference (RR) IQA models, reduced-reference entropic differencing (RRED) [46], fourier transform based quality measure (FTQM) [47], and orientation selectivity based visual pattern (OSVP) [48]. The third type consists of four NR IQA models, quality-aware clustering (QAC) [49], natural image quality evaluator (NIQE) [26], no-reference image quality metric for contrast distortion (NIQMC) [27], and integrated local NIQE (IL-NIQE) [28]. Note that about one half of the above-mentioned 14 IQA methods were developed in the past four years. The last type contains six state-of-the-art IQA models devoted to the DIBR-synthesized images, which are VSQA [30], 3D-SWIM [31], MW-PSNR [32], MP-PSNR [33], MP-PSNR-RR [34], and AR-plus thresholding (APT) [50]. Among them, the former fours are FR IQA models while the last twos are RR and NR IQA models.

Evaluation criteria: The correlation performances of our proposed blind MNSS and competitors are measured based on four broadly applied evaluation criteria, as suggested by the video quality expert group (VQEG) [51]. The first and second criteria are the spearman rank order correlation coefficient (SROCC) and the kendall's rank-order correlation coefficient (KROCC), which are both non-parametric tests for prediction monotonicity. By contrast, the SROCC focuses on computing the degree of association between two vectors, whereas, the KROCC inclines to evaluate the strength of dependence of two vectors. Further, the KROCC has stricter demands such as the vectors should be ordinal. The third criterion is the pearson linear correlation coefficient (PLCC) or commonly known as the linear correlation coefficient, which measures the correlation between two vectors in terms of prediction accuracy. Besides, the last one is the root mean square error (RMSE), which is used to predict the prediction consistency of two vectors. The first two criteria are directly applied to subjective quality scores ($\mathbf{q}_s$) and objective quality predictions ($\mathbf{q}_o$), whereas, the last two are conducted on $\mathbf{q}_s$ and converted objective quality predictions ($\tilde{\mathbf{q}}_o$), which is derived from $\mathbf{q}_o$ by decreasing its nonlinearity via a five-parameter nonlinear logistic equation:

$$\tilde{\mathbf{q}}_o = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp\left[\beta_2 (\mathbf{q}_o - \beta_3)\right]} \right) + \beta_4 \mathbf{q}_o + \beta_5 \quad (10)$$

where $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_5\}$ is a five-parameter vector, which is fitted using a nonlinear regression. It needs to stress two points: 1) the range of SROCC, KROCC and PLCC is $[0, 1]$ and that of RMSE is $[0, +\infty]$; 2) a good quality evaluation method inclines to obtain a large value of SROCC, KROCC and PLCC, but a small value of RMSE. More details concerning the criteria can be found in [11].

B. Performance Comparison

Comparison on evaluation criteria: We firstly calculate the performance indices of the proposed MNSS model in terms of the four evaluation criteria stated above. As reported in Table I, the MNSS model has attained inspiringly high values of SROCC, KROCC, PLCC and RMSE, respectively 0.770, 0.568, 0.785 and 0.412. The performance results of twenty competing IQA techniques are also illustrated in Table I for comparison. We highlight the best performing model in each type. As compared with classical and popular IQA methods in the first, second and third types, our proposed blind MNSS model has attained remarkably better performance than those IQA measures regardless of full-, reduced- and no-reference conditions. In fact, the optimal model among those 14 IQA models is the MAD, whose performance indices are 0.599, 0.437, 0.607 and 0.528, respectively. In terms of the SROCC, one of the most significant index in the four widely employed evaluation criterion, the proposed NR MNSS metric has given rise to the relative performance gain of 28.5% beyond the top FR MAD algorithm.

Furthermore, in comparison to the fourth type of six IQA methods which were developed specific to DIBR-synthesized images, despite the fact that reference image information is required and exploited in those metrics, the proposed MNSS model is noticeably superior to them. Considering the SROCC, our MNSS leads to a big relative performance gain of 7.5% beyond the best performing APT metric, which obtains 0.716

TABLE I
PERFORMANCE AND EFFICIENCY (SECOND) COMPARISON OF OUR BLIND MNSS MODEL WITH 20 CLASSICAL AND STATE-OF-THE-ART IQA METHODS ON THE IRCCYN/IVC DATABASE. WE HIGHLIGHT THE BEST PERFORMING METHOD IN EACH TYPE IN BOLD FONT

Metric	SROCC	KROCC	PLCC	RMSE	Time
SSIM	0.436	0.287	0.485	0.582	0.069
VSNR	0.385	0.275	0.437	0.598	0.526
MAD	**0.599**	**0.437**	**0.607**	**0.528**	3.516
IW-SSIM	0.405	0.273	0.583	0.540	1.046
FSIM	0.414	0.273	0.582	0.541	0.370
GMSD	0.332	0.239	0.407	0.608	0.019
PSIM	0.412	0.284	0.504	0.575	0.156
RRED	0.309	0.210	0.407	0.608	1.028
FTQM	**0.554**	**0.386**	**0.562**	**0.550**	0.929
OSVP	0.382	0.250	0.476	0.585	0.186
QAC	0.310	0.193	0.351	0.623	0.224
NIQE	0.373	0.242	0.437	0.598	0.523
NIQMC	0.425	0.280	0.362	0.621	3.950
IL-NIQE	**0.526**	**0.360**	**0.493**	**0.579**	5.811
VSQA	0.523	0.367	0.574	0.545	0.666
3D-SWIM	0.615	0.449	0.658	0.501	10.43
MW-PSNR	0.575	0.437	0.562	0.550	0.082
MP-PSNR	0.622	0.483	0.617	0.523	0.734
MP-PSNR-RR	0.663	0.538	0.677	0.489	0.473
APT	**0.716**	**0.577**	**0.731**	**0.455**	168.8
Q_1 (Pro.)	0.652	0.525	0.636	0.513	0.228
Q_2 (Pro.)	0.522	0.352	0.514	0.571	0.348
MNSS (Pro.)	**0.770**	**0.568**	**0.785**	**0.412**	0.526

of SROCC, 0.577 of KROCC, 0.731 of PLCC, and 0.455 of RMSE respectively.

Finally, we also compare the proposed MNSS metric with its two components, namely Q_1 and Q_2. As shown in Table I, according to SROCC, KROCC, PLCC and RMSE indices, Q_1 has acquired the values of 0.652, 0.525, 0.636 and 0.513, while, Q_2 has obtained the values of 0.522, 0.352, 0.514 and 0.571. These two blind quality measures are of comparable performance to those above state-of-the-art IQA algorithms particularly devoted to the DIBR-synthesized images, which demonstrates the two features' effectiveness. Additionally, we also find that systematically combining these two measures, i.e., the proposed MNSS model, can bring about remarkable performance gain, about 18.1% beyond Q_1 and 47.5% beyond Q_2 in terms of the SROCC. This reflects the necessity of incorporating Q_1 and Q_2 and indirectly prove their complementarity in the DIBR-synthesized IQA.

Comparison on statistical significance: Besides the above four evaluation criteria used for numerical comparison, more attention was paid to the statistical significance comparison recently. F-test is a typical statistical test, which follows a variance-based hypothesis and reveals additional information about the relative performance of distinct IQA methods [52]. The hypothesis which the F-test is based on supposition that the residual differences of subjective quality ratings and the corresponding objective quality predictions follow a Gaussian distribution, which, however cannot be perfectly satisfied in several cases. According to the Central Limit Theorem, the residual difference's distribution can be supposed to have a good approximation to the Gaussian shape when a large set of sample points are included, and thus the F-test is somewhat reasonable and can be applied in this study to compare the variances of two objective IQA models' prediction residuals. By determining whether the two sample sets have the same distribution, the F-test is able to make a statistically sound judgment about superiority or inferiority of an objective IQA model against another one. We implement the F-test on the proposed MNSS model and 17 high-performance competitors (including 15 competing IQA methods and two components), and the results are listed in Table II. A value '0' means that the two IQA methods are statistically equivalent, while, '$+1$' and '-1' separately mean that the method in the column is statistically superior and inferior to that in the row. Clearly, our proposed MNSS model is statistically better than all the IQA models tested.

Furthermore, we also check the statistical significance of six representative models, which include the best performing method in each of four types and two components in the MNSS, with the IQA models considered. One can see that, despite that there exist differences of those testing IQA models in numerical performance such as RMSE and PLCC, the most of statistical significance results between each pair of them are zero, namely, statistically indistinguishable.

Comparison on scatter plots: One of the most commonly used comparison strategy in IQA research is the scatter plot, which can show some intuitive illustrations and comparisons of different IQA models. As illustrated in Fig. 7, we show the scatter plots of MOS versus 10 representative objective IQA models tested on the IRCCyN/IVC database. Those models considered are composed of: 1) FR SSIM, FR MAD, FR PSIM, RR FTQM, and NR IL-NIQE; 2) DIBR-specific FR MW-PSNR, FR MP-PSNR, NR Q_1, NR Q_2, and NR MNSS. As can be observed, the proposed MNSS model derives the sample points with better linearity and convergence, which illustrates its superiority in blindly evaluating the quality of DIBR-synthesized images.

Comparison on parameter influence: Robustness of the used parameters is also a critical index to check the effectiveness of a quality metric. Typically, a good IQA model should have slight influences on performance when the used parameters vary. We first examine the impacts of the stability constants, i.e., ε in Equation (2) and ϵ in Equation (8), on the performance of our proposed MNSS model. Here we merely concentrate on the representative PLCC and SROCC values. Results are provided in Fig. 8. For the readers' conveniences, we label the maximum (minimum) PLCC and SROCC values, which are respectively 0.785 (0.750) and 0.770 (0.731). This means that the proposed MNSS algorithm has a good robustness to the stability constants. Moreover, we carefully observe the Fig. 8 and find that ε has very few influences but ϵ almost has no impacts. Next, how the number

TABLE II
STATISTICAL SIGNIFICANCE COMPARISON OF TESTING IQA MODELS BASED ON THE F-TEST. M_1: SSIM; M_2: VSNR; M_3: MAD; M_4: FSIM; M_5: GMSD; M_6: PSIM; M_7: RRED; M_8: FTQM; M_9: OSVP; M_A: NIQE; M_B: NIQMC; M_C: IL-NIQE; M_D: MW-PSNR; M_E: MP-PSNR; M_F: APT; M_G: Q_1; M_H: Q_2; M_I: THE PROPOSED MNSS

Method	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_A	M_B	M_C	M_D	M_E	M_F	M_G	M_H	M_I
M_3	0	0	–	0	0	0	+1	0	0	0	+1	0	0	0	-1	0	0	-1
M_8	+1	0	0	0	0	0	+1	–	0	+1	+1	+1	0	0	-1	0	+1	-1
M_C	0	0	0	0	0	0	0	+1	0	0	0	–	0	0	-1	+1	0	-1
M_F	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	–	+1	+1	-1
M_G	+1	0	0	0	0	+1	+1	0	0	+1	+1	+1	0	0	-1	–	+1	-1
M_H	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	-1	+1	–	-1
M_I	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	–

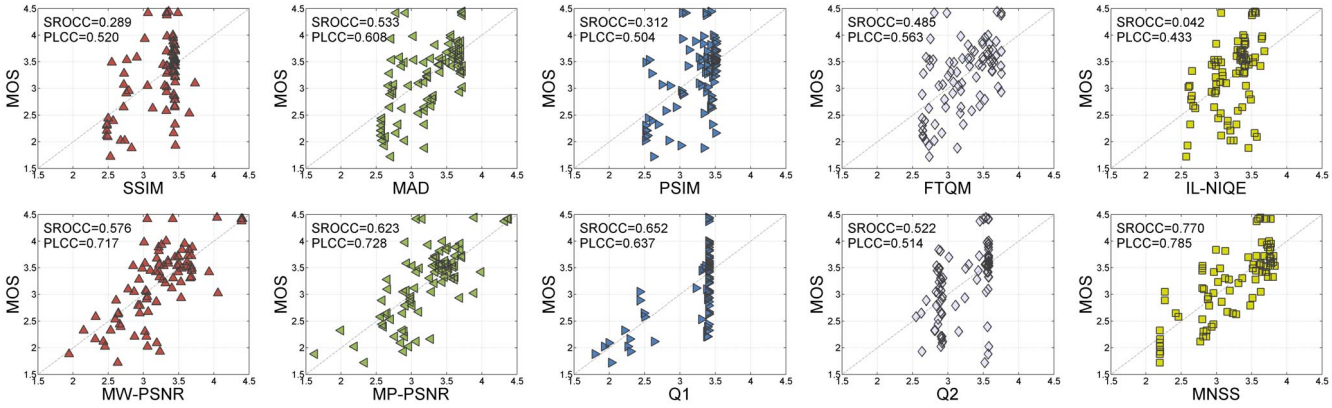


Fig. 7. Scatter plots of MOS versus 10 representative objective IQA models tested on the IRCCyN/IVC database. We label two typical indices (i.e., SROCC and PLCC) for each model and use the black diagonal dash line to provide the benchmark perfect prediction.

of scale affects the performance is also checked and, as it grows from 2 to 5, the PLCC (SROCC) values are separately 0.616 (0.466), 0.617 (0.506), 0.657 (0.498), and 0.785 (0.770). One can find that the performance is sensitive to the number of scale, which is possibly because the two NSS models do not work when the number of scale is not sufficiently large.

Comparison on implantation efficiency: Implementation speed, or equivalent computational complexity, is also a very critical factor for judging the performance of a quality metric. Hence, we compute the implementation time of the proposed MNSS model with the entire IRCCyN/IVC database, in which the DIBR-synthesize views are of the resolution 1024×768. We carry out the test of implantation efficiency using the MATLAB R2015a and a laptop with 2.50GHz CPU processor and 8.00GB memory. The average result of implementation costs of all tested images is about 0.5 second, as reported in Table I. The average implementation costs of 20 IQA metrics considered in this research are also illustrated in Table I for comparison. As seen, our proposed MNSS model works quite efficiently. In fact, the MNSS model is operated with a series computing. Note that we can independently compute Q_1 and Q_2 values, and moreover, in Q_1 (or Q_2), we can also simultaneously conduct the similarity measure for each scale before combination. With these concerns, the parallel computing is very probably introduced to largely elevate its implementation efficiency for meeting the real-time requirements in practical applications.

IV. APPLICATION TO VIDEO QUALITY ASSESSMENT

Video, due to its supplement of motion information in the temporal domain, is one of the most commonly used media for information acquisition. Further, we can regard image as a special case of video when no motion information is involved, or in other words, all the frames in one video sequence are totally the same. So the video quality assessment (VQA) is of important research and application values for its capability of helping to monitor and control the quality of video under acquisition, compression, transmission and so on [53]. We try to extend the MNSS model proposed above to a blind VQA model by taking temporal information into consideration.

A. Proposed VQA Metric

We design a temporal pooling to combine the quality score of each frame, which is estimated by the proposed blind MNSS model. Based on some motivations, which come from classical and popular temporal pooling strategies as stated above and from spatial pooling schemes summarized in [54] such as saliency-based and quality-based spatial pooling, this work introduces a new efficient and effective singular frame-aware

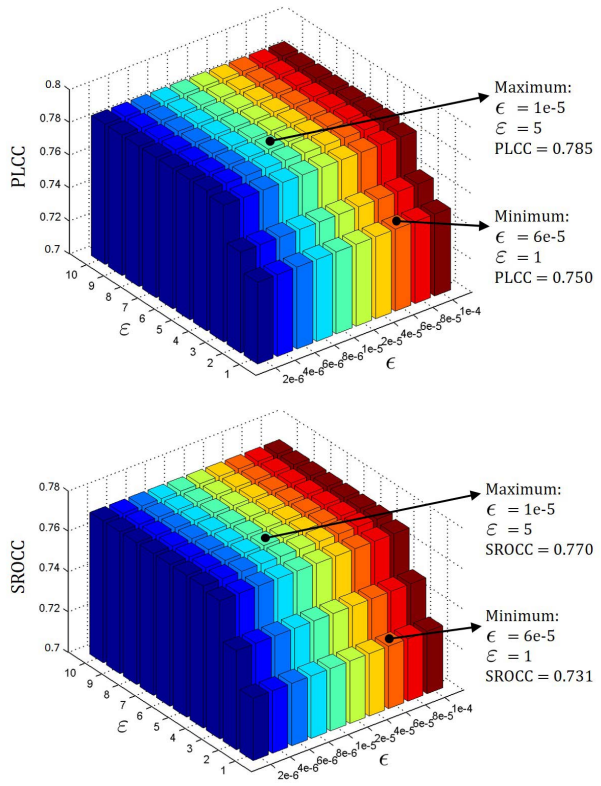


Fig. 8. PLCC and SROCC values with distinct stability constants.

temporal pooling for extending the MNSS model to the video quality measure of DVBR-synthesized views. The singular frames are defined from two perspectives: one each for frame quality and frame complexity. Apparently, the frames of low visual quality usually attract more human attention than other frames, and therefore we include this as an important factor to decide singular frames.

The frame (or image) complexity is abstract and hard to be endowed with a clear definition although it is an essential attribute of images. Here we simply exploit the magnitude of information contained in a frame to define its complexity - the more information one frame contains, the higher complexity it has, and vice versa. According to the free energy-based brain theory [55] and some analyses, the local autoregressive (AR) model has been successfully used to compute image complexity [56]. For an image of size 1024×768, the time consumed to estimate image complexity with the local AR model is about 156 seconds, which considerably decreases the application values in real-time video processing systems. Towards alleviating the time-consuming problem, we replace the local AR model with an efficient compression-backed computational model [57]. Generally speaking, as compared with the intensity of an input stimulus, humans pay more attention to its variations, especially in the study of quality assessment. For illustration consider the following example. Luminance structures and contrast structures are extracted in the benchmark SSIM method. The former one is associated to the intensity mean, while, the latter one is associated to the intensity variations. In light of correlation with subjective quality scores, contrast structures have brought about greater performance than luminance structures. With this concern, we similarly concentrate on the variations of frame complexity used for temporal pooling.

Based on the above analyses, we assume that the singular frames, which have unordinary image complexity and low visual quality, usually attract more human attention than other frames. Specifically, we denote a video sequence as $\mathbf{v} = \{v_1, v_2, \ldots, v_n\}$, where n is the number of the frames, and its frame quality scores derived using Equation (9) as $\mathbf{f} = \{f_1, f_2, \ldots, f_n\}$ and its image complexity estimations as $\mathbf{c} = \{c_1, c_2, \ldots, c_n\}$. We more care about image complexity variations, which is denoted as $\tilde{\mathbf{c}} = \{\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n\} = \{|c_1 - \bar{c}|^\theta, |c_2 - \bar{c}|^\theta, \ldots, |c_n - \bar{c}|^\theta\}$ where $\bar{c}$ is the expectation of $\mathbf{c}$ and θ is a parameter empirically assigned as 3. First, we generate a new vector $\mathbf{t} = \tilde{\mathbf{c}}/\mathbf{f} = \{\tilde{c}_1/f_1, \tilde{c}_2/f_2, \ldots, \tilde{c}_n/f_n\}$ and rank this vector from large to small and record the former $m\%$ indices, i.e., $\{i_1, i_2, \ldots, i_m\}$, as the singular frames. Next we define frame quality scores and image complexity variations of the singular frames to be $\mathbf{f}_m = \{f_{i_1}, f_{i_2}, \ldots, f_{i_m}\}$ and $\tilde{\mathbf{c}}_m = \{\tilde{c}_{i_1}, \tilde{c}_{i_2}, \ldots, \tilde{c}_{i_m}\}$. Second, we yield the overall quality score Q_{MNSSV} of one video stream using weighted averaging pooling:

$$Q_{\text{MNSSV}} = \frac{\sum_{k=1}^{m} f_{i_k} \cdot \tilde{c}_{i_k}}{\sum_{k=1}^{m} \tilde{c}_{i_k}}. \tag{11}$$

Note that the larger Q_{MNSSV} score means the better quality prediction and vice versa.

B. Performance Evaluation

To examine the proposed MNSSV metric for blind VQA, the IRCCyN/IVC DVBR database was applied [59], which contains 102 video sequences with 6 seconds and resolution 1024×768 between 15 and 30 frames per second. Note that the design principle of the proposed blind MNSSV metric is the two newly established NSS models specific to the geometric distortions, so we exclude the 18 compression-related sequences. Three different multiview plus depth (MVD) video sequences, i.e., 'Book Arrival' (1024×768, 16 cameras with 6.5cm spacing), 'Lovebird1' (1024×768, 12 cameras with 3.5 cm spacing) and 'Newspaper' (1024×768, 9 cameras with 5 cm spacing), are included and processed with seven DIBR technologies to generate four novel viewpoints for each vide sequence. The authors gathered the individual votes and MOS scores based on the ACR-HR experiment. Three blind quality assessment models are included for comparison. Two of them are respectively NIQEV and IL-NIQEV, which are developed by incorporating the proposed temporal pooling into NIQE and IL-NIQE, while, the rest one is the recently proposed NR-VQA metric, VIIDEO [60]. In what follows, we will carry out three types of comparisons, one each for evaluation criteria, statistical significance, and implantation efficiency.

First of all, we examine the performance of the proposed blind MNSSV metric and report the results in Table III. As seen, the performance indices have reached to a high level, namely 0.631 of SROCC, 0.457 of KROCC, 0.632 of PLCC, and 0.382 of RMSE respectively. As compared with three

TABLE III
ACCURACY AND EFFICIENCY (SECOND/FRAME) COMPARISON OF FOUR NR MODELS ON THE IRCCYN/IVC DVBR DATABASE

Metric	SROCC	KROCC	PLCC	RMSE	Time
NIQEV	0.482	0.332	0.431	0.449	0.491
IL-NIQEV	0.384	0.266	0.403	0.445	8.835
VIIDEO	0.293	0.211	0.331	0.483	6.539
MNSSV (Pro.)	0.631	0.457	0.632	0.382	0.538

competing NR quality methods, our MNSSV metric has lead to a relative performance gain over 25% in light of SROCC. Secondly, we conduct the statistical significance comparison between the proposed MNSSV metric of each competitors considered using the F-test. Results reveal that MNSSV is statistically superior to the whole three testing quality models. Lastly, we compute the average computational time of each testing quality metric consumed to each video frame. One can see from Table III, for a video frame of resolution 1024×768, the proposed MNSSV model just needs half a second, close to the time of our MNSS model for computing one image of the same resolution. Similarly, we can introduce the parallel computing towards speeding up its implementation.

C. Discussion

The MNSSV metric is of good performance in predicting the quality of DVBR-synthesized views, but in comparison, its performance indices are far less than the MNSS metric in the DIBR-synthesize IQA. In our opinion, this problem is mainly caused by the subsequent two reasons. The first is that the MNSSV metric is an extended version of MNSS, which is devised based on two NSS models established on natural images, with a temporal pooling scheme. Therefore, the MNSSV will be improved by introducing novel NSS models built on natural video streams. The second reason lies in that, as for one frame in a DVBR-synthesized video sequence, less than one tenth of second is supplied for observation and quality evaluation, while, by contrast, human beings can take sufficient time, e.g., 2 to 4 seconds, to view and assess a DIBR-synthesized image. Hence, some video attention methods will be considered to be incorporated for advancing the proposed blind MNSSV metric. Besides, it is noted that, as compared with NFERM [10] and PPPC [61] which were developed with pixel-based and global-based NSS models, the proposed MNSS and MNSSV metrics are patch-based. In the future, we plan to properly combine pixel-, patch- and global-based NSS models to devise a better blind IQA model of DVBR-synthesized views.

V. CONCLUSION

In this paper, we have proposed a novel training-free NR quality assessment model of DIBR-synthesized views based on two newly established NSS models. The first NSS model is developed based on the self-similarity of natural images for extracting local statistical feature, and the second one is based on the main structure consistency of natural images for extracting global statistical feature. We further systematically combine the two features above to propose the final MNSS method. Via experiments, the local and global features were proved of high performance in assessing the quality of DIBR-synthesized images, while, the proposed blind MNSS metric has achieved noticeably greater correlation performance than its two components and state-of-the-art IQA methods. The code of our model will be released at https://kegu.netlify.com.

REFERENCES

[1] F. Battisti and P. Le Callet, "Quality assessment in the context of FTV: Challenges, first answers and open issues," *IEEE COMSOC MMTC Commun. Front.*, vol. 11, no. 2, pp. 22–27, Mar. 2016.
[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] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 341–353, Jun. 2018.
[4] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "Toward a blind quality metric for temporally distorted streaming video," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 367–378, Jun. 2018.
[5] A. Doumanoglou *et al.*, "Quality of experience for 3-D immersive media streaming," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 379–391, Jun. 2018.
[6] F. Battisti, M. Carli, P. Le Callet, and P. Paudyal, "Toward the assessment of quality of experience for asymmetric encoding in immersive media," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 392–406, Jun. 2018.
[7] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 518–523, Jun. 2018.
[8] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process. Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.
[9] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "Efficient image deblocking based on postfiltering in shifted windows," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 122–126, Jan. 2008.
[10] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
[11] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.
[12] 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.
[13] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Mar. 2010, Art. no. 011006.
[14] L. 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.
[15] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1–13, Jun. 2016.
[16] L. Li *et al.*, "Perceptual quality evaluation for image defocus deblurring," *Signal Process. Image Commun.*, vol. 48, pp. 81–91, Oct. 2016.
[17] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.
[18] B. Hu *et al.*, "No-reference quality assessment of compressive sensing image recovery," *Signal Process. Image Commun.*, vol. 58, pp. 165–174, Oct. 2017.
[19] H. Cai, L. Li, Z. Yi, and M. Gong, "Towards a blind image quality evaluator using multi-scale second-order statistics," *Signal Process. Image Commun.*, vol. 71, pp. 88–99, Feb. 2019.
[20] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. SPIE Conf. Stereoscopic Displays Virtual Real. Syst. X*, May 2004, pp. 93–104.
[21] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Process. Image Commun.*, vol. 24, nos. 1–2, pp. 65–72, Jan. 2009.

[22] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[23] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[24] 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.

[25] L. Li *et al.*, "Sparse representation-based image quality index with adaptive sub-dictionaries," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3775–3786, Aug. 2016.

[26] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[27] K. Gu *et al.*, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.

[28] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[29] Y. Zhou *et al.*, "Blind quality index for multiply distorted images using biorder structure degradation and nonlocal statistics," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.

[30] P.-H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," in *Proc. Electron. Imag. Int. Soc. Opt. Photon.*, Feb. 2012, pp. 8288–8256.

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

[32] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological wavelets," in *Proc. IEEE Int. Workshop Qual. Multimedia Exp.*, Jan. 2015, pp. 1–6.

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

[34] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "Multi-scale synthesized view assessment based on morphological pyramids," *J. Elect. Eng.*, vol. 67, no. 1, pp. 1–9, Jan. 2016.

[35] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 661–674, Nov. 1984.

[36] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, Sep. 2001.

[37] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[38] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[39] C. Kim and P. Milanfar, "Visual saliency in noisy images," *J. Vis.*, vol. 13, no. 4, pp. 1–14, Mar. 2013.

[40] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jul. 2001, pp. 416–423.

[41] K. Gu, J.-F. Qiao, P. Le Callet, Z. Xia, and W. Lin, "Using multiscale analysis for blind quality assessment of DIBR-synthesized images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 745–749.

[42] K. Gu, L. Li, H. Lu, X. Min, and W. 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] *Subjective Video Quality Assessment Methods for Multimedia Applications*, Int. Telecommun. Union, Geneva, Switzerland, 2008.

[44] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[45] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[46] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[47] M. Narwaria, W. Lin, I. V. McLoughlin, S. Emmanuel, and L.-T. Chia, "Fourier transform-based scalable image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364–3377, Aug. 2012.

[48] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.

[49] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 995–1002.

[50] 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. 2018.

[51] VQEG. (Mar. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.vqeg.org/

[52] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[53] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[54] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "Video quality pooling adaptive to perceptual distortion severity," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 610–620, Feb. 2013.

[55] K. Friston, "The free-energy principle: A unified brain theory?" *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010.

[56] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sep. 2014.

[57] K. Gu *et al.*, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

[58] H. Attias, "A variational Bayesian framework for graphical models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 2000, pp. 209–215.

[59] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, "Visual quality assessment of synthesized views in the context of 3D-TV," in *Proc. 3D-TV Syst. Depth Image Rendering*, 2013, pp. 439–473.

[60] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan. 2016.

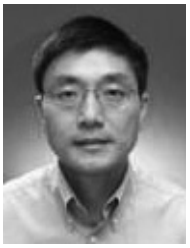
[61] K. Gu, J. Qiao, and X. Li, "Highly efficient picture-based prediction of PM2.5 concentration," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3176–3184, Apr. 2019.

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

Junfei Qiao (M'11) received the B.E. and M.E. degrees in control engineer from Liaoning Technical University, Fuxin, China, in 1992 and 1995, respectively, and the Ph.D. degree from Northeast University, Shenyang, China, in 1998. He was a Post-Doctoral Fellow with the School of Automatics, Tianjin University, Tianjin, China, from 1998 to 2000. He joined the Beijing University of Technology, Beijing, China, where he is currently a Professor. He is the Director of the Intelligence Systems Laboratory. His current research interests include neural networks, intelligent systems, self-adaptive/learning systems, and process control systems. He is a Reviewer for over 20 international journals, such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He is a member of the IEEE Computational Intelligence Society.

Sanghoon Lee (SM'12) received the B.S. degree in electrical engineering from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1991, and the Ph.D. degree in electrical engineering from the University of Texas at Austin, Austin, TX, USA, in 2000. From 1991 to 1996, he was with Korea Telecom, Seongnam, South Korea. From 1999 to 2002, he was with Lucent Technologies, Murray Hill, NJ, USA, on 3G wireless and multimedia networks. In 2003, he joined the Department of Electrical and Electronics Engineering, Yonsei University, as a Faculty Member, where he is currently a Full Professor. His current research interests include image/video quality assessment, computer vision, graphics, cloud computing, multimedia communications, and wireless networks. He was a recipient of the 2015 Yonsei Academic Award from Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society. He was the Technical Program Co-Chair of the International Conference on Information Networking 2014 and the Global 3-D Forum in 2012 and 2013, respectively, and the General Chair of the 2013 IEEE IVMSP Workshop. He has been serving as the Chair of the IEEE P3333.1 Quality Assessment Working Group since 2011. He has been serving on the Technical Committee of the IEEE IVMSP Technical Committee since 2014 and the IEEE Multimedia Signal Processing since 2016. He also served as an Editor for the *Journal of Communications and Networks* from 2009 to 2015 and a special issue Guest Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014. He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS since 2014, and the *Journal of Electronic Imaging* since 2015.

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is serving for the IEEE MMTC, as the Chair of the Interest Group on Quality of Experience for Multimedia Communications. He is an Associate Editor of the IEEE TRANSACTIONS ON HUMAN–MACHINE SYSTEMS and the IEEE TRANSACTIONS ON MULTIMEDIA.



Weisi Lin (F'16) received the Ph.D. degree from Kings College London. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with over 340 refereed papers published in international journals and conferences. He has been on the Editorial Board of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and JVCI. He has been elected as APSIPA Distinguished Lecturers in 2012 and 2013. He served as the Technical-Program Chair for Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo in 2013, and the International Workshop on Quality of Multimedia Experience in 2014. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

Patrick Le Callet (F'19) received the M.Sc. and Ph.D. degrees in image processing from the Ecole Polytechnique de l'Universite de Nantes, Nantes, France. He was with the Department of Electrical Engineering, Technical Institute of the University of Nantes, Nantes, as an Assistant Professor from 1997 to 1999 and a full-time Lecturer from 1999 to 2003. Since 2003, he teaches with the Department of Electrical Engineering and the Department of Computer Science, École Polytechnique de l'Universite de Nantes (Engineering School), where he is currently a Full Professor. Since 2006, he has been the Head of the Image and Video Communication Laboratory with CNRS IRCCyN, a group of over 35 researchers. He is currently involved in research dealing with the application of human vision modeling in image and video processing. His current research interests include 3-D image and video quality assessment, watermarking techniques, and visual attention modeling and applications. He has co-authored over 200 publications and communications and holds 13 international patents in the above areas. He served as an Associate Editor of several journals, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS SYSTEM AND VIDEO TECHNOLOGY, the *EURASIP Journal on Image and Video Processing*, and *SPIE Electronic Imaging*.