

# Model-Based Referenceless Quality Metric of 3D Synthesized Images Using Local Image Description

Ke Gu, Vinit Jakhetiya, Jun-Fei Qiao, *Member, IEEE*, Xiaoli Li, Weisi Lin, *Fellow, IEEE*, and Daniel Thalmann

**Abstract**—New challenges have been brought out along with the emerging of 3D-related technologies, such as virtual reality, augmented reality (AR), and mixed reality. Free viewpoint video (FVV), due to its applications in remote surveillance, remote education, and so on, based on the flexible selection of direction and viewpoint, has been perceived as the development direction of next-generation video technologies and has drawn a wide range of researchers’ attention. Since FVV images are synthesized via a depth image-based rendering (DIBR) procedure in the “blind” environment (without reference images), a reliable real-time blind quality evaluation and monitoring system is urgently required. But existing assessment metrics do not render human judgments faithfully mainly because geometric distortions are generated by DIBR. To this end, this paper proposes a novel referenceless quality metric of DIBR-synthesized images using the autoregression (AR)-based local image description. It was found that, after the AR prediction, the reconstructed error between a DIBR-synthesized image and its AR-predicted image can accurately capture the geometry distortion. The visual saliency is then leveraged to modify the proposed blind quality metric to a sizable margin. Experiments validate the superiority of our no-reference quality method as compared with prevailing full-, reduced-, and no-reference models.

**Index Terms**—Quality assessment, no-reference, depth image-based rendering, image description, autoregression, saliency.

## I. INTRODUCTION

RECENTLY, Free Viewpoint Videos (FVVs) and 3D television have received considerable attention owing to its widespread applications in several areas such as

Manuscript received September 3, 2016; revised January 14, 2017, March 24, 2017, and May 5, 2017; accepted July 14, 2017. Date of publication July 28, 2017; date of current version November 3, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61703009, Grant 61473034, and Grant 61673053, in part by the Nova Programme Interdisciplinary Cooperation Project under Grant Z161100004916041, in part by Singapore MoE Tier 1 Project under Grant M4011379 and Grant RG141/14, and in part by the Singapore MoE Tier 2 Project under Grant M4020355.020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet. (*Corresponding author: Ke Gu*.)

K. Gu, J.-F. Qiao, and X. Li are with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn; junfeiq@bjut.edu.cn; lixiaolibjut@bjut.edu.cn).

V. Jakhetiya is with the Department of Computer Science Engineering, Bennett University, Greater Noida 201310, India (e-mail: vinit.jakhetiya@bennett.edu.in).

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

D. Thalmann is with the BeingThere Centre, Institute for Media Innovation, Nanyang Technological University, Singapore 639798 (e-mail: daniel.thalmann@ntu.edu.sg).

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/TIP.2017.2733164

remote surveillance, medical applications, remote education, entertainment, and more. In FVVs, it is required to generate new viewpoints with the help of neighboring multiple views and these new viewpoints are generated using the Depth-Image-Based-Rendering (DIBR) techniques, which allows the rendering of an entirely new representation of the scene and thereby saves large cost and complexity of camera set up [1]. Typically, there include distortions in these synthesized new views, especially the geometric distortion, which has different characteristics compared with the distortions that occur in natural images [2]. A poor DIBR-synthesized view may create annoying artifacts in the whole FVVs, but current solutions fail in accurately capturing those artifacts. With this concern, it is desired to conduct more elaborately subjective assessment and have an efficient and effective objective assessment metric to judge the quality of DIBR-synthesized views [1].

In last few decades, many Image Quality Assessment (IQA) metrics were devised in the literature for natural images. IQA metrics can be divided into three classes, namely Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) or blind. In these three categories, complete, partial and no information of the reference image is required. Such as, Wang *et al.* proposed an FR IQA model called Structural SIMilarity (SSIM) [3] based on the luminance, contrast and structural similarity between a distorted image and its reference image. Using the SSIM as the bed-stone, Wang *et al.* further came up with the Information Weighted SSIM (IW-SSIM) metric with a weighting strategy, which is designed based on the statistical information theory and the natural scene statistics (NSS) model [4]. In the same line, Gu *et al.* recently improved the SSIM with a valid pooling scheme and proposed the Analysis of Distortion Distribution-based SSIM (ADD-SSIM) metric [5].

During the past some years, researchers have realized the strong relationship between free energy principle and quality assessment. Wu *et al.* [6], Zhai *et al.* [7], and Gu *et al.* [8] deployed this principle in the design of FR, RR and NR IQA methods, respectively. Generally speaking, the free energy principle reveals that the human vision separates an image into two parts, which have different impacts on the overall perception of visual quality. According to typical low-level features, e.g. image gradient, Zhang *et al.* [9], Gu *et al.* [10], and Gu *et al.* [11] proposed highly effective FR IQA methods on commonly encountered distortion categories. In practice, the development of high-performance FR models is benefit

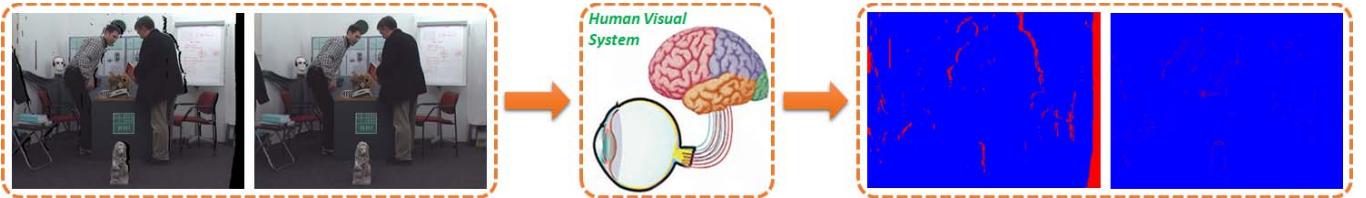


Fig. 1. Quality perception by the human visual system. The left two images are two DIBR-synthesized images respectively under weak and strong geometry distortion corruption. The DIBR-synthesized image with weak geometry distortion is processed by filling holes based on inpainting techniques. Via the human visual system (middle), the left DIBR-synthesized images can be classified into geometry distortion and non-geometry distortion regions, as illustrated in the right two images. We label these two regions with different colors.

for devising reliable blind quality metrics. Instead of human ratings, Gu *et al.* [12] first used the quality scores computed by FR models to label a large number of training samples, and then used the labelled training samples to learn the features and derive the NR metric. The blind IQA models developed based on this strategy can better avoid the overfitting problem introduced in training the features.

Several recent studies were conducted to analyze the statistical characteristics of natural images. Representative works are [13]–[15]. Besides these, the Natural Image Quality Evaluator (NIQE) [16] and Integrated Local NIQE (IL-NIQE) [17], which are free of learning and human scored images, have aroused more attention. These two NR metrics estimate the distances of a question image and a group of natural images in different angles, e.g. the statistics of mean subtracted contrast normalized (MSCN) coefficients, image gradient, log-gabor filter responses, etc, to predict the visual quality of the question image. These two metrics were proved to perform very efficiently for capturing the main occurring distortions in natural images and consequently yielding the faithful quality measures of natural images. Besides, some special IQA studies were concerned in recent works, including DIBR-synthesized images [2], multiple distortions [18], [19], foggy images [20], contrast distortions [21], tone mapping [22], [23], stereoscopic images [24], screen content images [25], etc.

Whether for typical distortions or for particular distortions as mentioned above, well-established NR metrics have been developed except the problem of DIBR-synthesized IQA. In reality, via extensive experiments, we found that FR models, even though specifically designed for judging the quality of DIBR-synthesized images, are unable to deliver high performance. Aiming at this, Bosc *et al.* [2] made the first attempt to explore why IQA methods based on existing theories and models fail on DIBR-synthesized views, and in the meantime pointed out the necessity of new IQA algorithms for DIBR-synthesized views. Such as Conze *et al.* [26] devised an SSIM-based VSQA metric, in which the SSIM model is modulated with the orientation, texture and contrast backed weighting maps. Battisti *et al.* [27] proposed an approach based on the comparison of statistical features extracted from wavelet subbands of original and distorted DIBR-synthesized views. In this approach a registration step is included to insure shifting-resilience before features comparison. Sandić-Stanković *et al.* deployed morphological wavelet decomposition for the quality assessment of DIBR-synthesized views and named it MW-PSNR [28]. Likewise, the same authors devised an IQA model by replacing the morphological

wavelet decomposition with the morphological pyramid decomposition, MP-PSNR, for better performance [29], [30]. Lately, Sandić-Stanković *et al.* further proposed MP-PSNR-reduc by modifying MP-PSNR to attain larger performance and faster calculation [31]. Although these specifically designed algorithms implement better than the algorithms designed for typical distortions, their performance is still not up to the mark.

In present IQA metrics, we observe two important issues highly related to the DIBR-synthesized IQA. First, the IQA algorithms designed for natural images cannot capture the geometric distortion that is the prevailing artifact appearing in DIBR-synthesized images. Second, existing IQA methods proposed for DIBR-synthesized views in the literature, even if only a few, were all based on the entire reference image; that is, no NR IQA models has been developed yet. It needs to stress that, in real application scenarios, reference DIBR-synthesized image is not accessible, so NR IQA metrics have much higher values than FR methods. This is why we focus our attention on the quality assessment of DIBR-synthesized images without original references. The main contribution of this paper is to devise an effective blind IQA algorithm using the newly developed NSS model. Based on autoregression (AR)-based local image description, the NSS model suits the characteristics of the DIBR-synthesized images and is able to capture the geometric distortions in DIBR-synthesized views. We called the proposed metric “AR-plus thresholding” (APT). The performance of the proposed APT algorithm is validated on the DIBR-synthesized image database and experiments illustrate that our blind APT achieves the best result beyond state-of-the-art FR, RR and NR IQA approaches.

The remainder of the paper is organized below. Section II briefly analyzes the issues occurred in the existing IQA models for assessing the quality of DIBR-synthesized images and introduces the proposed new NSS model and NR APT metric. Experiments are performed on the IRCCyN/IVC database [2] to validate the superiority of our method than state-of-the-art competitors. Concluding remarks are given in Section IV.

## II. PROPOSED ALGORITHM

In this work, we propose a referenceless quality assessment metric for DIBR-synthesized images using the local AR modeling followed by the thresholding. The key problem, which highly degrades the quality of DIBR-synthesized images, is the geometry distortion that usually introduces a remarkable

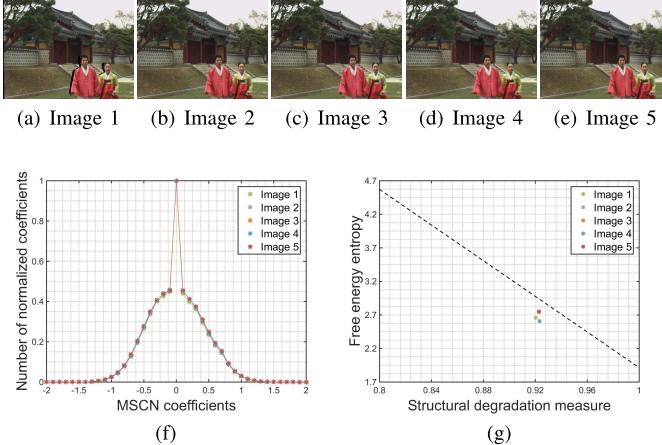


Fig. 2. Comparison of NSS models for DIBR-synthesized images: (a)-(e) five DIBR-synthesized images which have quite different quality scores and are associated to the same scene; (f) the NSS model used in [15]; (g) the NSS model used in [8].

destruction of naturalness attribute of an image. As shown in Fig. 1, the human visual system (HVS) can effortlessly and faithfully separate the geometry distortion regions from non-geometry distortion regions in a DIBR-synthesized image. The left image is a typical DIBR-synthesized image under strong geometry distortion corruption. Via the HVS, the left DIBR-synthesized image can be classified into geometry distortion and non-geometry distortion regions, as illustrated in the right image in Fig. 1. We label these two regions with red and blue colors, respectively, for readers' conveniences.

To address the above-described issue, a natural way is to straightforwardly take current NSS models into consideration. Unfortunately, these NSS models were developed to capture the structural distortion in natural images and cannot capture the geometric distortion. In order to give more insight about these arguments, in our research, two typical NSS models in [8] and [15] are applied to model DIBR-synthesized images. In Figs. 2(a)-(e), five DIBR-synthesized images that are chosen from the IRCCyN/IVC database [2] are presented. These five DIBR synthesized images have quite different quality scores. Fig. 2(f) shows the statistic histogram of MSCN coefficients deployed in [15]. This NSS model reveals that the histogram of MSCN coefficients of a natural image exhibits a Gaussian-like distribution, which will be destroyed by the addition of distortions. For example, Gaussian blur distortion, reshapes the distribution of MSCN coefficients towards a Laplacian distribution. One can observe that these five histograms are quite close to each other; in other words, this NSS model cannot be used to distinguish the quality of these five DIBR-synthesized images.

The results derived by exploiting the NSS model in [8] on DIBR-synthesized images are illustrated in Fig. 2(g). This regulation suggests that, for natural images, there exists an approximated linear relationship between their structural degradation measures and free energy entropy, as shown using the black dash line in Fig. 2(g). With the increment of distortions, the point of structural degradation measures and free energy entropy values will be deviated from that black line. This NSS model associated to five DIBR-synthesized

images is shown in Fig. 2(g) with five different colored stars and one can see that these stars are located quite near. This experiment implies that this NSS model cannot distinguish the DIBR-synthesized images of different quality scores. That is to say, this NSS model is still incapable of reliably assessing the quality of DIBR-synthesized views.

To this aim, we consider introducing the concept of local image similarity, which indicates that, in a natural image, the correlation between one pixel and its neighborhood is quite akin to those of other adjacent pixels in a local region. In this work we use the classical AR operator to describe the local image similarity of a given DIBR-synthesized image, due to its good attribute of invariance to object transformations such as translation, rotation, scaling, etc. The other local-based operators such as bilateral and non-local means filter may be also employed to describe the local image similarity.

In previous researches of psychophysics and computational neuroscience, a widely acknowledged viewpoint of biological perceptual systems is that the natural environment modulates the visual apparatus to be highly adapted to it and to have evolved to the most efficient manner towards extracting visual information from it [32], [33]. A recently revealed human brain theory, free energy principle, can be perceived as one part of the evolved visual apparatus stated above. Simply speaking, the free energy principle synthesizes many prevailing brain theories about biological and physical sciences, and unveils that the internal generative mechanism in brain governs the human cognition process [34]. This mechanism points out that the brain deploys a constructive manner to separate a given image into orderly parts and disorderly parts, in order for the follow-up image perception and analysis. Through a series of analyses provided in [35], the process of minimizing free energy approximates to predictive coding and it can be well simulated based on the AR model towards quality assessment and saliency detection [8], [36]. The aforementioned analyses also promote the choice of AR operator for local image description.

In this work a local AR model is used for image analysis. Autocorrelation, also known as serial correlation, is the correlation of a signal with itself. For one specific pixel in an image  $I$ , we denote its location index as  $i$  and its value as  $x_i$ , and accordingly construct a relationship between this pixel and its neighborhood:

$$x_i = \Omega_\theta(x_i)\mathbf{s} + d_i \quad (1)$$

where  $\Omega_\theta(x_i)$  defines a neighborhood vector which includes its surrounding  $\theta$  pixels in the local  $\sqrt{\theta+1} \times \sqrt{\theta+1}$  patch. While  $\mathbf{s} = (s_1, s_2, \dots, s_\theta)^T$  constitutes a vector of AR parameters to be determined and  $d_i$  represents the error difference between the current pixel value and its corresponding AR prediction. In the proposed algorithm, we merely take account of the most adjacent 8 pixels next to the present pixel. Of course, more adjacent pixels can be used to build a more complicated relationship. From experiments, it was found that enlarging the local patch to introduce more pixels must cause implementation cost largely increased, but cannot lead to definite gain in the performance. With this view, we assign the parameter  $\theta = 8$ . Apparently, the parameter  $\theta$  affects the

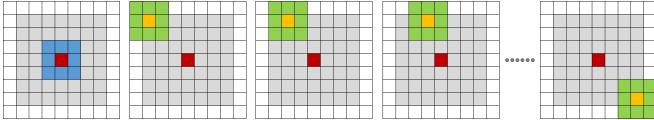


Fig. 3. Illustration of how to implement the local AR model.

autocorrelation since the autocorrelation of natural images is local and it will be not local anymore with  $\theta$  increased. In the next section, we particularly analyze the influence of  $\theta$  on the performance of the APT metric.

Our next concern is how to determine the AR parameters. In order to estimate the reliable vector of AR parameters, we first represent the linear system in a matrix way:

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{x} - \mathbf{X}\mathbf{s}\|_2 \quad (2)$$

where  $\mathbf{x} = (x_{i,1}, x_{i,2}, \dots, x_{i,\phi})^T$ , and  $\mathbf{X}(j,:) = \Omega_\theta(x_{i,j})$ , where  $j = \{1, 2, \dots, \phi\}$ . Next we solve the linear system in equation (2) via the least square method and infer the best estimation of the vector of AR parameters to be

$$\hat{\mathbf{s}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}. \quad (3)$$

Note that we set  $\phi = 48$  based on the assumption of wide sense stationarity in a local patch and validity of principle of “geometric duality”. Specifically, we hypothesize that the relationship that is built upon the current pixel using equation (1) also exists for adjacent 48 pixels in the local  $7 \times 7$  patch. Except the present pixel, we make use of the information of surrounding 48 pixels to ascertain the relationship, that is the best estimation of the AR coefficient vector  $\hat{\mathbf{s}}$  yielded with equation (3). In addition, advanced optimization technologies, such as simulated annealing method and genetic algorithm, might be considered in the future work for better prediction of pixels.

For the readers’ conveniences, we use a schematic diagram to help to illustrate how to implement the local AR model, as shown in Fig. 3. Let us see this figure from the left to right. In the first subfigure, each small block means a pixel, and the middle 49 grey small blocks constitute the local  $7 \times 7$  patch. The center red small block refers to the current pixel to be processed. It and its adjacent 8 blue small blocks constitute the local  $\sqrt{\theta+1} \times \sqrt{\theta+1}$  patch, namely the local  $3 \times 3$  patch because of  $\theta = 8$ . In the second subfigure, the golden small block is close to the center red small block, within the range of the local  $7 \times 7$  patch, and it and its surrounding 8 green small blocks constitute a local patch with the same size of the middle  $3 \times 3$  patch in the first subfigure. The third to last subfigures provide other 47 conditions.

Basically speaking, the AR predicted image can be thought of as the orderly part, while the absolute value of the error difference map between the input image and its associated AR-based predicted image, i.e.  $|d_i|$ , is perceived as the disorderly part. For natural images or natural image patches, the AR-model can efficiently clarify the each type of contents (i.e., smooth, edge and texture) and consequently, values in the error map are always very small. In Fig. 4(a), we have shown three typical natural image patches. Smooth, textural

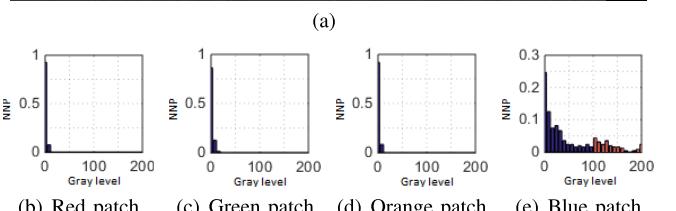


Fig. 4. Comparison of natural image patches and geometry distorted patch: (a) a typical DIBR-synthesized distorted image containing labelled smooth, edge, textural and geometry distorted patches; (b)-(e) histograms of error difference of the labelled four patches above, where NNP means the normalized number of pixels.

and edge patches are labelled using the red, green and orange boxes, respectively. We have also shown their corresponding error maps’ histograms in Figs. 4(b)-(d). As can be viewed, these three histograms have very similar distributions, in which the majority of values are zero and the few non-zero values are still less than 20. According to the observation above, we are able to derive the subsequent two conclusions.

- 1) The disorderly part always exists in the natural images and the magnitude of the disorder part depends on the characteristics of patches, e.g. smooth, edge or texture.
- 2) The values in the error map are zero or small and typically these values are not greater than 20 in most conditions for DIBR-synthesized images.

It is also interesting to observe that the second conclusion indirectly reflects the good ability of the AR predictor used for local image description, since the values in the error map are usually small.

On the other hand, very different results were found when the AR modeling is applied to a patch which has geometric distortions. An example is shown in Fig. 4 to give more insight about these different results. In Fig. 4(a), we label a blue patch that includes a typical geometry distortion region. Likewise, we calculate the error difference of this patch and moreover plot its histogram, as illustrated in Fig. 4(e). One can see from the histogram that the AR modeling cannot effectively predict the regions with geometrical distortion and consequently, it includes many big values of error difference, few of which are even greater than 100 (as marked with red color). This observation reveals that the used AR operator is capable of highlighting the geometry distortions. Actually, in theoretical,

geometry distortions (e.g. the blue patch in Fig. 4(a)) destroy the assumption of local image similarity; that is, the above-described relationship established via equation (1) does not exist anymore, and thus the error difference in the geometry distorted area suddenly increases compared with typical smooth, edge and textural patches in natural images mentioned above. This phenomenon can be validly used for judging the quality of a DIBR-synthesized image.

Based on the observations and analyses above, we propose to use the AR model to predict the input DIBR-synthesized image and thus attain the error difference map. Some simple modifications should be made to refine the computed error difference map. To be more specific, we first “clean” the error map with a Gaussian filter for the purpose of removing some isolated small “noise”. Then the  $\gamma\%$  most salient regions are detected to be used for excluding the associated parts in the error map, since these non-geometry distorted areas that are very likely to have large-valued error differences should be left aside. We have considered several recently proposed saliency detection methods [36]–[39], the FES model in [36] is selected due to its better performance. Influences of using different saliency detection techniques on correlateion performance will be compared in the experimental results. Lastly, we binarize the filtered error map with a threshold to generate a “0-1” map:

$$M_d = \begin{cases} 1, & \text{if } M_e < \Upsilon_t \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where  $M_e$  means the error map processed by the Gaussian filter;  $\Upsilon_t$  is a constant threshold. We further apply the median filter to process  $M_d$  for burrs removal.

Our next consideration focuses on pooling the binary map  $M_d$  generated based on equation (4) to get the final quality score. The majority of existing IQA metrics were developed based on the NSS model. Some FR-IQA methods compare the difference of structural variations of a distorted image and its associated reference (natural) image. RR-IQA metrics extract RR feature vectors from the reference (natural) and contaminated images followed by measuring the distance of two vectors containing RR features. In NR-IQA models, one certain statistic regulation derived from natural images, and the quality of the distorted image is estimated based on its deviation from the aforesaid natural statistic regulation. So, it is a reliable way to assess image quality by comparing the distance of a corrupted image and its associated natural image in structural, RR vector or statistic domains. Note that, using equation (4), the small values in the error difference map  $M_e$  are assigned as the unit, which are associated to the natural image parts. In comparison, those large values corresponding to geometry distorted regions are set as zero. This means that the values in the binary map  $M_d$  of a distortion-free natural image should be all one. We can compare  $M_d$  with the binary map of natural images  $M_r$  to predict the quality of a DIBR-synthesized image. The procedures defined in equations (1)–(4) can be perceived as a new NSS regulation in an “AR-plus-thresholding” domain.

Our goal is to find out how much the binary map of image with geometric distortions  $M_d$  differs with the binary map

of the natural image  $M_r$ . With this view, we use a reference map  $M_r$  which has the same size with  $M_d$ . The binary map of reference  $M_r$  can be produced by applying equations (1)–(4) to the natural images. Then, similar to the frequently used similarity measure, which has three merits such as symmetry, boundedness and unique maximum [3], we define the quality measure of the distorted DIBR-synthesized image by

$$Q_s = \frac{1}{L} \sum_{l=1}^L \left( \frac{2M_d(l) \cdot M_r(l) + \varepsilon}{M_d(l)^2 + M_r(l)^2 + \varepsilon} \right)^\alpha \quad (5)$$

where  $l$  is the pixel index;  $L$  indicates the number of the pixels in image;  $\varepsilon$  is a constant number which is very small for avoiding the problem of division-by-zero;  $\alpha$  is a positive exponent that is associated to the Minkowski summation. More emphasis is generally shifted to the low-quality regions with  $\alpha$  increased. This sounds reasonable since more human attentions are absorbed by one kind of small region, in which most distortions are included [4]. Typical value of  $\alpha$  is from 1 to 4. In fact, state-of-the-art FR methods deploy different pooling schemes to highlight high distortion regions [9]–[11].

We assume that all the values in the binary reference map  $M_r$  are the unit, as geometric distortions in natural images do not occur and the value of residual errors is lower than the threshold, as shown in Figs. 4(b)–(d). So, equation (5) can be rewritten as

$$Q_s = \frac{1}{L} \sum_{l=1}^L \left( \frac{2M_d(l) + \varepsilon}{M_d(l)^2 + 1 + \varepsilon} \right)^\alpha. \quad (6)$$

From this equation, it can be found that in the denominator, the term  $M_d(l)^2 \geq 0$  and thereby  $M_d(l)^2 + 1 \geq 1$ , which indicates that the denominator cannot be zero. Thus, we can remove the small variable  $\varepsilon$  in equation (6) and rewrite it to be

$$\begin{aligned} Q_s &= \frac{1}{L} \sum_{l=1}^L \left( \frac{2M_d(l)}{M_d(l)^2 + 1} \right)^\alpha \\ &= \underbrace{\frac{1}{L} \sum_{l \in L_0} \left( \frac{2M_d(l)}{M_d(l)^2 + 1} \right)^\alpha}_{\text{(I) Geometric distorted regions}} + \underbrace{\frac{1}{L} \sum_{l \in L_1} \left( \frac{2M_d(l)}{M_d(l)^2 + 1} \right)^\alpha}_{\text{(II) Non-geometric distorted regions}} \end{aligned} \quad (7)$$

where  $L_0$  and  $L_1$  are respectively associated to the regions in which all the values are zero and one, and  $L = L_0 + L_1$ . In other words, it can be said that, in a DIBR-synthesized image,  $L_0$  and  $L_1$  respectively stand for the region which has geometric distortion and the region which do not have geometric distortion. Therefore, we can simplify equation (7) when  $\alpha$  is positive:

$$\begin{aligned} Q_s &= \frac{1}{L} \sum_{l \in L_0} \left( \frac{2 \cdot \mathbf{0}}{\mathbf{0} + 1} \right)^\alpha + \frac{1}{L} \sum_{l \in L_1} \left( \frac{2 \cdot \mathbf{1}}{\mathbf{1} + 1} \right)^\alpha \\ &= \frac{L_1}{L}. \end{aligned} \quad (8)$$

The Minkowski pooling has no influence on the final quality estimation, and therefor we can remove the exponent  $\alpha$  from equation (8).

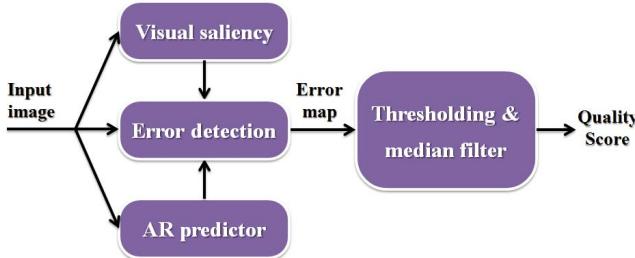


Fig. 5. Block diagram of the proposed blind APT metric.

The quality score, i.e. the similarity between the binary maps of a distorted DIBR-synthesized image and natural images, can be readily solved as the ratio of the number of one-value pixels to the overall map size (i.e. the ratio of number of pixels in non-geometry distorted regions and the total number of pixels in a DIBR-synthesized image). From equation (8), we can see that a DIBR-synthesized image has a higher quality when the score  $Q_s$  is close to one. For such images, most of the values in the binary synthesized map  $M_d$  are one and consequently, the corresponding image has less amount of geometric distortions (pixels in  $L_0$  regions). From equation (8), it can be also perceived that the score  $Q_s$  does not depend on any information about the reference image, so our proposed algorithm is completely blind/referenceless. We further show a block diagram of the proposed blind APT metric in Fig. 5 for improving the readability of this paper.

### III. EXPERIMENTAL RESULTS

In this research, we devise a valid local image description based blind quality assessment metric, specifically to match the characteristics of the DIBR-synthesized images. In order to check the efficacy of the proposed algorithm, we apply it on the IRCCyN/IVC database [2] dedicated to the DIBR-synthesized images. This database is composed of the total 96 images, in which 12 are original images and the remaining 84 images are associated to synthesized views. The 84 DIBR-synthesized views are predominantly corrupted due to the presence of geometric distortions and used for performance validation.

#### A. Competing IQA Metrics and Evaluation Methodology

We have compared the proposed algorithm with recently developed quality assessment methods based on the various phenomena. We have divided existing IQA models used for the comparison in two categories. In the first category, we have included IQA metrics which were designed for natural images. The second category is composed of the IQA models which were designed to match the characteristics of the DIBR-synthesized images. The IQA methods in the first category include full-reference algorithms (such as, PSNR, SSIM [3], Visual Signal-to-Noise Ratio [40], Most Apparent Distortion (MAD) [41], IW-SSIM [4], Feature SIMilarity (FSIM) [9], ADD-SSIM [5], Perceptual SIMilarity (PSIM) [11]), reduced-reference algorithms (such as, Reduced-Reference Entropic Differencing (RRED) [42], Free Energy based Distortion

Metric (FEDM) [7], Fourier Transform based Quality Measure (FTQM) [43], Orientation Selectivity based Visual Pattern (OSVP) [44]), and referenceless algorithms (such as, NIQE [16], Quality-Aware Clustering (QAC) [45], IL-NIQE [17], Six-Step BLInd Metric (SISBLIM) [18]).

On the other hand, we have also compared the proposed referenceless quality assessment algorithm with existing full reference algorithms designed for DIBR-synthesized images (in the second category), such as, VSQA [26], 3D-SWIM [27], MW-PSNR [28], MP-PSNR [29], and MP-PSNR-reduc [31]. To the best of our knowledge, no blind quality assessment algorithm of DIBR-synthesized views has been proposed in the literature. In this paper, we do not compare the proposed algorithm with those learning-based IQA algorithms, since the IRCCyN/IVC database contains only 96 images and insufficient training data are very likely to make those learning-based model to be susceptible to the bias.

For the purpose of performance evaluation, we have used four widely employed criteria, namely Pearson Linear Correlation Coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC), Kendall's Rank-order Correlation Coefficient (KRCC), and Root Mean Square Error (RMSE). PLCC and RMSE are deployed to measure the prediction accuracy, while SRCC and KRCC are used to find the monotonicity of the prediction. A better quality assessment algorithm should attain a higher value of PLCC, SRCC, and KRCC, while achieve a lower value of RMSE. It needs to reduce the nonlinearity of objective prediction scores before conducting these four correlation measures. So the objective prediction scores are mapped to subjective human ratings using the following five parameter nonlinear logistic function beforehand:

$$f(Q_s) = \kappa_1 \left( 0.5 - \frac{1}{1 + e^{\kappa_2(Q_s - \kappa_3)}} \right) + \kappa_4 Q_s + \kappa_5 \quad (9)$$

where  $Q_s$  and  $f(Q_s)$  represent the predicted scores using the IQA methods and its corresponding mapped scores.  $\kappa_i$  ( $i \in 1, 2, 3, 4, 5$ ) are the parameters which are needed to be fitted during the nonlinear regression. Then we compute those four correlation criteria using the converted objective predictions  $f(Q_s)$  and subjective quality scores.

#### B. Performance Comparison

In Table I, the comparison results of the proposed APT metric with existing IQA algorithms (both the first and second categories) are presented. The proposed algorithm achieves 0.7307, 0.7157, 0.5766, and 0.4546 of PLCC, SRCC, KRCC, and RMSE, respectively, which is much better than those competing IQA methods. From Table I, we are able to derive two important conclusions:

- 1) Those existing IQA algorithms that were designed for natural images (in the first category) are unable to perform effectively. The MAD algorithm [41] performs the best among FR IQA methods tested and it obtains 0.6667, 0.5603, 0.4077, and 0.4963 of PLCC, SRCC, KRCC, and RMSE, respectively. Across eight RR and NR IQA metrics, the FTQM and SISBLIM methods

TABLE I

PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM WITH RECENTLY DEVELOPED METRICS (QUALITY ASSESSMENT METRICS FOR BOTH IMAGES AND DIBR). THE BEST PERFORMANCE IN EACH TYPE IS HIGHLIGHTED WITH THE BOLD-FACES

Metric	PLCC	SRCC	KRCC	RMSE	Designed for	Category
PSNR	0.3976	0.3095	0.2063	0.6109	Monoscopic Views	Full-Reference
SSIM [3]	0.4850	0.4368	0.2877	0.5823	Monoscopic Views	Full-Reference
VSNR [40]	0.4370	0.3851	0.2756	0.5989	Monoscopic Views	Full-Reference
MAD [41]	<b>0.6077</b>	<b>0.5994</b>	<b>0.4372</b>	<b>0.5288</b>	Monoscopic Views	Full-Reference
IW-SSIM [4]	0.5831	0.4053	0.2733	0.5409	Monoscopic Views	Full-Reference
FSIM [9]	0.5828	0.4148	0.2733	0.5411	Monoscopic Views	Full-Reference
ADD-SSIM [5]	0.5512	0.4672	0.3252	0.5556	Monoscopic Views	Full-Reference
PSIM [11]	0.5315	0.4576	0.3033	0.5640	Monoscopic Views	Full-Reference
RRED [42]	0.4072	0.3090	0.2104	0.6081	Monoscopic Views	Reduced-Reference
FEDM [7]	0.2252	0.1817	0.1307	0.6487	Monoscopic Views	Reduced-Reference
FTQM [43]	<b>0.5628</b>	<b>0.5543</b>	<b>0.3864</b>	<b>0.5504</b>	Monoscopic Views	Reduced-Reference
OSVP [44]	0.4767	0.3829	0.2508	0.5853	Monoscopic Views	Reduced-Reference
NIQE [16]	0.4374	0.3739	0.2421	0.5987	Monoscopic Views	No-Reference
QAC [45]	0.3519	0.3108	0.1936	0.6232	Monoscopic Views	No-Reference
IL-NIQE [17]	0.4998	<b>0.5348</b>	<b>0.3668</b>	0.5767	Monoscopic Views	No-Reference
SISBLIM [18]	<b>0.5225</b>	0.3832	0.2721	<b>0.5677</b>	Monoscopic Views	No-Reference
VSQA [26]	0.5742	0.5233	0.3673	0.5451	DIBR-Synthesized Views	Full-Reference
3D-SWIM [27]	0.6584	0.6156	0.4496	0.5011	DIBR-Synthesized Views	Full-Reference
MW-PSNR [28]	0.5622	0.5757	0.4378	0.5506	DIBR-Synthesized Views	Full-Reference
MP-PSNR [29]	0.6174	0.6227	0.4833	0.5238	DIBR-Synthesized Views	Full-Reference
MP-PSNR-reduc [31]	<b>0.6772</b>	<b>0.6634</b>	<b>0.5382</b>	<b>0.4899</b>	DIBR-Synthesized Views	Reduced-Reference
Proposed APT	<b>0.7307</b>	<b>0.7157</b>	<b>0.5766</b>	<b>0.4546</b>	DIBR-Synthesized Views	No-Reference

separately lead to the optimal results in their personal types, more than 0.5 of PLCC. These simulation results confirm our claim that those existing algorithms in the first category cannot catch the geometric distortions, which are the predominant artifacts contained in the DIBR-synthesized images.

- Despite the fact that the IQA methods designed for the DIBR-synthesized images (in the second category) implement better than those designed for natural images while the performance indices of those algorithms are not sufficient yet. For example, the MP-PSNR-reduc metric performs the best among the IQA models designed for DIBR-synthesized views and it attains 0.6772 of PLCC, which is still smaller than the PLCC value obtained by our proposed blind APT algorithm. And furthermore, in those IQA approaches, it requires the complete information about the reference synthesized views, which are generally not accessible in most real application scenarios.

Overall, the proposed APT metric is completely referenceless and it can achieve much higher value of PLCC, SRCC, and KRCC and lower value of RMSE as compared with the IQA algorithms considered.

Moreover, we used 6 parts of the IRCCyN/IVC database to examine the influence of certain DIBR algorithms, i.e. A2-A7, on the performance of our APT metric. Table II illustrates the results and points out that APT has achieved high correlation with subjective ratings for the frames synthesized by the algorithms A3, A4, A5 and A7. But simultaneously, it was also

TABLE II  
RANK AND PLCC OF DIBR ALGORITHMS A2-A7

Algorithm	A2	A3	A4	A5	A6	A7
Rank	6	1	3	1	5	4
SRCC	0.26	0.90	0.88	0.90	0.55	0.83

found that the performance for A2 and A6 is not good. This phenomenon is very possibly due to the reason that few highly obvious geometry distorted regions are included in the frames synthesized by the algorithms A2 and A6, and thus our APT metric mistakes those frames for high-quality natural images and results in the poor correlation performance, as shown in Table II. In reality, our APT metric is good at predicting poor-quality 3D-synthesized views which are corrupted by obvious geometry distortions. As for assessing high-quality 3D-synthesized views which are quite close to natural images, the MP-PSNR model with the ability to discern small differences between the reference and synthesized images has better performance than our proposed blind APT metric. So, our further work will be devoted to developing new NSS models specific to this IQA task for promoting the performance of our APT method. It deserves attention that, despite the poor results for A2 and A6, the proposed APT metric works without reference, while, the MP-PSNR is reference-based, and thus the blind APT algorithm can be thought of as a good DIBR-synthesized IQA model. Furthermore, as provided in Section III-D later, the APT metric can be also considered as

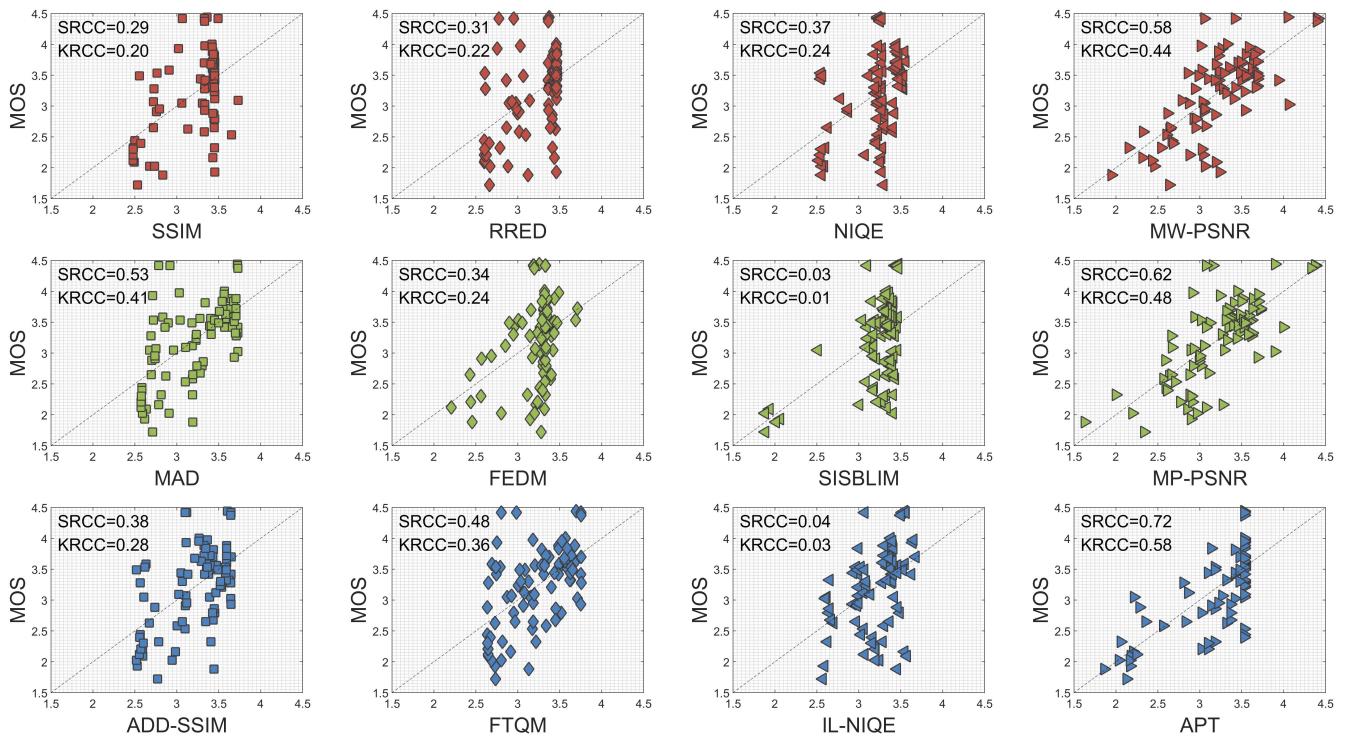


Fig. 6. Scatter plots of MOS versus FR models (SSIM, MAD and ADD-SSIM), RR models (RRED, FEDM and FTQM), NR models (NIQE, SISBLIM and IL-NIQE), and models designed for DIBR-synthesized views (MW-PSNR, MP-PSNR and our proposed APT) on the IRCCyN/IVC database. The black diagonal dash line means the perfect prediction.

a feature to be incorporated into state-of-the-art FR methods, such as MP-PSNR, towards higher performance.

In Fig. 6, using the IRCCyN/IVC database, we provide the scatter plots of subjective MOS values versus objectives quality estimations derived by different IQA metrics, which include: 1) FR SSIM, MAD and ADD-SSIM; 2) RR RRED, FEDM and FTQM; 3) NR NIQE, SISBLIM and IL-NIQE; 4) DIBR-relevant MW-PSNR, MP-PSNR and our proposed APT. It is apparent that, as compared with competitors, the sample points of the proposed algorithm present higher convergence and linearity, which means that our APT model can deliver more consistent objective scores in line with the subjective ratings.

### C. Performance Dependency of Used Parameters

A robust quality assessment model is more desired; that is, a good IQA metric refers to a metric whose performance should not vary significantly with the slight change of the parameters. Let us first discuss the influence of the threshold  $\Upsilon_t$  used in equation (4) on the performance. Seeing Fig. 4(e), the error difference pixels belonging to the geometric distorted regions have very large values. Any higher value of threshold can be used to distinguish between the pixels in natural image patches and those corrupted by geometric distortions, and this can be used for highlighting those geometric distorted pixels. As thus, we have examined the dependency of the proposed algorithm on the threshold  $\Upsilon_t$  and provided the result in Fig. 7. From this figure, we can perceive two observations below.

- 1) Even with a large range of variation of the threshold  $\Upsilon_t$ , the performance of the proposed APT metric does not change significantly. Its worst performance (0.6301 and

0.5345) is still better than the majority of state-of-the-art IQA metrics tested. The worst performance of the proposed algorithm has achieved the 0.6396 of PLCC.

- 2) It was found that, when the threshold  $\Upsilon_t$  varies in a comparative large interval (from 30 to 110), the values of PLCC and SRCC are greater than 0.6841 and 0.6236, respectively. By comparing these results with those reported in Table I, one can see that even the lowest performance (0.6841 and 0.6236) is still superior to the overall testing IQA models, regardless of using reference information or not and designed for monoscopic views or DIBR-synthesized views.

In the proposed algorithm, we have fixed  $\Upsilon_t$  value to be 100, based upon the experiment above. It is also interesting to notice that some existing algorithms based upon the free energy theory [6], [7] (AR modeling) perform poorly for the quality assessment of DIBR-synthesized images, whereas the proposed APT metric based on the AR modeling followed by the thresholding implement very effectively. These simulation results demonstrate the superiority of the proposed new NSS model, namely “AR-plus-thresholding”, for blindly assessing the quality of DIBR-synthesized views.

Subsequently, we will discuss the influence of the ratio  $\gamma\%$  on the performance. This parameter decides the regions to be excluded in the error map. As shown in Fig. 8, we have presented the result to examine the performance dependency of the proposed APT metric on the ratio  $\gamma\%$ . According to the two curves in the figure, we are able to derive three conclusions below.

- 1) We compare the correlation performance of the proposed APT model in a large interval of the ratio  $\gamma\%$  (from

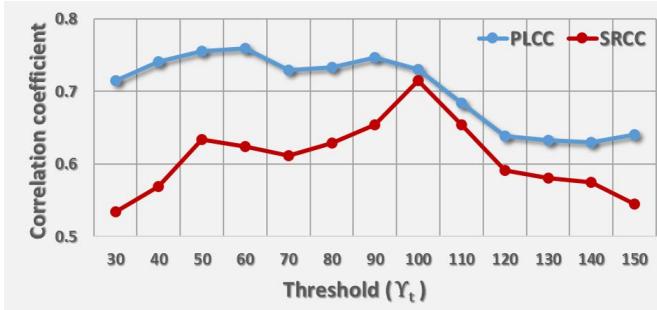


Fig. 7. Performance dependency of the proposed APT metric with the changing thresholds ( $\gamma_t$ ).

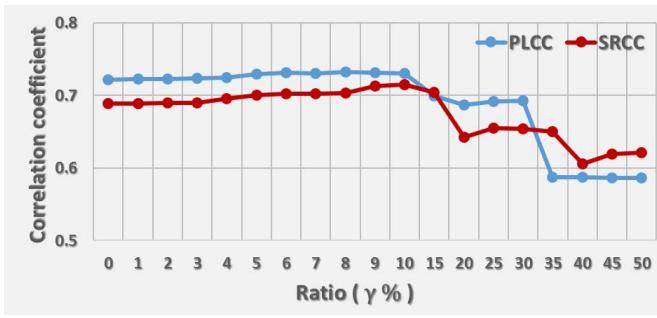


Fig. 8. Performance dependency of the proposed APT metric with the changing ratios ( $\gamma$ %).

0 to 50). The result illustrates the robustness of our APT metric, only with a small fluctuations. The lowest performance scores are 0.5864 and 6055 of PLCC and SRCC, even superior to most popular and recently proposed quality methods.

- 2) If we focus our attention on the performance variations in the range of the ratio  $\gamma\%$  from 0 to 25, it can be found that, in such situation, all the PLCC and SRCC values are greater than 0.6871 and 0.6419, respectively. This means that, in a comparative large range, our proposed APT algorithm outperforms all the competing quality metrics, even though some metrics were designed for DIBR-synthesized views and using the complete reference information.
- 3) It has to be emphasized that the performance scores at  $\gamma = 0$  are associated to the APT metric without using any saliency detection model, which is named APT<sub>ns</sub>. At this time, the values of PLCC and SRCC are 0.7213 and 0.6885. Comparing these two values with state-of-the-art metrics, as shown in Table I, the APT<sub>ns</sub> delivers a considerable high performance, much better than the competitors.

Eventually, we examine the influence of different  $\theta$  values on the correlation performance. In the proposed APT metric,  $\theta$  is set to be 8 since we only consider the local patch which include the most adjacent 8 pixels, i.e. 1-distance pixels. Here we further examine  $n$ -distance pixels, where  $n = \{2, 3, 4, 5\}$ , and accordingly  $\theta = \{24, 48, 80, 120\}$ . The performance and implementation cost are compared, as illustrated in Table III. One can see that, in general, the performance indices decrease

TABLE III  
INFLUENCE OF DIFFERENT  $\theta$  VALUES ON PERFORMANCE

$\theta$	4	8	24	48	80	120
PLCC	0.72	0.73	0.75	0.72	0.68	0.63
SRCC	0.64	0.72	0.61	0.53	0.56	0.46
KRCC	0.48	0.58	0.45	0.37	0.42	0.33
RMSE	0.46	0.46	0.44	0.46	0.49	0.52
Time (minute)	2.64	2.81	4.97	4.93	6.68	8.88

TABLE IV  
PERFORMANCE OF IMPROVED FR IQA MODELS FOR DIBR-SYNTHESIZED VIEWS BY COMBINING OUR BLIND APT METRIC

Metric	PLCC	SRCC	KRCC
MW-PSNR [28]	0.5622	0.5757	0.4378
Proposed APT	0.7307	0.7157	0.5766
MW-PSNR*APT	0.7568	0.7341	0.5555
Gain I	↑ 34.6%	↑ 27.5%	↑ 26.9%
Gain II	↑ 3.57%	↑ 2.57%	↓ 3.66%
MP-PSNR [29]	0.6174	0.6227	0.4833
Proposed APT	0.7307	0.7157	0.5766
MP-PSNR*APT	0.7604	0.7611	0.5843
Gain I	↑ 23.2%	↑ 22.2%	↑ 20.9%
Gain II	↑ 4.06%	↑ 6.34%	↑ 1.34%

with  $\theta$  increased. In the meantime, the implementation time (in terms of minute) promptly grows yet. Moreover, we also check  $\theta = 4$  since 4 connected neighborhood is used frequently, and results can be found in Table III. Similarly, our proposed blind APT metric has achieved better performance in comparison. So, in our blind APT metric, we assign  $\theta$  and  $n$  as 8 and 1.

#### D. Improving Full-Reference DIBR-Synthesized IQA Models

Actually, FR IQA metrics might be improved by including part or all of features (e.g. NSS-based features) used in NR models [22]. So we have also examined whether the proposed APT measure can serve as a feature to improve existing FR IQA algorithms which were designed for DIBR-synthesized views. We define the improved metric  $Q_i$  as

$$Q_i = Q_s * Q_f^\beta \quad (10)$$

where  $Q_s$  and  $Q_f$  are predicted quality scores derived by using our APT model and a FR model for DIBR-synthesized IQA, and  $\beta$  is a weighting parameter to adjust the relative importance between  $Q_s$  and  $Q_f$ . In this paper, we consider two recently proposed FR IQA metrics for DIBR-synthesized views, namely MW-PSNR [28] and MP-PSNR [29]. The parameter  $\beta$  is an empirically selected constant for attaining greater performance on the IRCCyN/IVC database. Table IV summarizes the performance results of the improved metrics and the associated relative gains on the IRCCyN/IVC database. As seen, Gain I indicates the performance gain of improved metric relative to its original version, while Gain II indicates the performance gain of improved metric relative to our

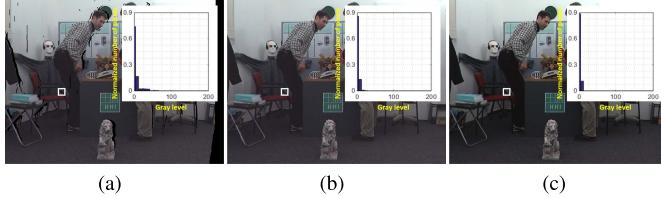


Fig. 9. Influence of inpainting techniques on our APT algorithm.

proposed APT metric. It can be observed that, as compared with the original MW-PSNR (or MP-PSNR), incorporating the novel APT metric constantly introduces a high level of performance improvements with respect to its original version. Even compared with our APT metric, the improved metrics are always leading to a remarkable performance gain. The above phenomenon can be explained by the fact that the design philosophy behind the proposed APT metric is based on a new NSS model, different from existing IQA models for DIBR-synthesized views.

### E. Discussions

In comparison to existing studies, there include some spotlights in our work. First, based on AR-based local image description, we have proposed a new NSS model which can capture the geometric distortion appeared in DIBR-synthesized images. Second, to our best knowledge, this work is the first NR IQA model designed for DIBR-synthesized views, which has achieved superior performance as compared with state-of-the-art IQA models. Third, our proposed APT metric is a quite robust algorithm, not sensitive to the parameters used in it. Fourth, by incorporating our APT metric, we can improve existing FR methods for DIBR-synthesized IQA to a large extent. Fifth, the saliency detection can be regarded as an optional step of our APT metric towards faster implementation since only a small performance decrease will be introduced after removal of saliency detection model, i.e.  $\gamma = 0$ .

However, in terms of computational cost, the APT metric consumes around 2.77 minutes for one  $768 \times 1024$  DIBR-synthesized image, whereas the MW-PSNR and MP-PSNR only require about 1.21 and 0.0823 seconds. Finally, we also analyze the impact of using inpainting techniques on the performance of our APT metric. As displayed in Fig. 9, (a) and (b) represent two DIBR-synthesized images, respectively without and with hole-filling techniques, and (c) is the associated reference natural image. As for the labelled white patch, we provide the histograms of their error maps at the top right corner, akin to Figs. 4(b)-(d). Comparing (a) and (b), it was found that, by means of the inpainting technique, the performance of our APT metric is influenced. Specifically, the histogram's tail reduces and thus it cannot be used to validly catch the distortion. In contrast, we also compare (b) and (c). The labelled white patch in (b) includes an annoying bended edge, while that in (c) includes a natural edge. One can see that, despite not obvious evidence, the annoying-bended-edge patch has longer tail than the natural-edge patch.

### IV. CONCLUSIONS

The 3-D synthesized views are the backbone of free-view point videos (FVV), virtual reality (VR), and augmented reality (AR), and mixed reality (MR). In the literature, several algorithms were proposed to generate the 3-D synthesized view but referenceless quality assessment metrics are missing. The referenceless metrics are essentially required to evaluate and monitor the synthesized views, as reference synthesized views are generally not available. With this view, in this work, we have proposed an effective and first referenceless quality assessment algorithm for DIBR-synthesized views using the proposed new natural scene statistics (NSS) model called as “AR plus threshold”. The residual error between a DIBR-synthesized image and its reconstructed image using the AR-modeling can validly capture the geometry distortions. The visual saliency is then leveraged to modify the proposed blind quality metric to a sizable margin. Experiments validate the superiority of our no-reference quality method as compared with prevailing existing full-, reduced-, and no-reference approaches, which includes both type of algorithms specifically designed for the natural and DIBR-synthesized images. The code will be released at <https://sites.google.com/site/guke198701/publications>.

### 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. Frontiers*, 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, Nov. 2011.
- [3] 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.
- [4] 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.
- [5] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, “Analysis of distortion distribution for pooling in image quality prediction,” *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.
- [6] J. Wu, W. Lin, G. Shi, and A. Liu, “Perceptual quality metric with internal generative mechanism,” *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.
- [7] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, “A psychovisual quality metric in free-energy principle,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.
- [8] 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.
- [9] 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.
- [10] K. Gu *et al.*, “Saliency-guided quality assessment of screen content images,” *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1–13, Jun. 2016.
- [11] 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.
- [12] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, “Learning a blind quality evaluation engine of screen content images,” *Neurocomputing*, vol. 196, pp. 140–149, Jul. 2016.
- [13] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From scene statistics to perceptual quality,” *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [14] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

- [15] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 209–212, Mar. 2013.
- [17] 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.
- [18] 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.
- [19] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
- [20] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.
- [21] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2017.2649101.
- [22] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
- [23] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [24] F. Shao, W. Lin, G. Jiang, and Q. Dai, "Models of monocular and binocular visual perception in quality assessment of stereoscopic images," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 123–135, Jun. 2016.
- [25] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.
- [26] P. H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," *Electron. Imag. Int. Soc. Opt. Photon.*, pp. 8256–8288, Feb. 2012.
- [27] 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.
- [28] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological wavelets," in *Proc. IEEE Int. Workshop Quality Multimedia Experience*, Jan. 2015, pp. 1–6.
- [29] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological multi-scale approach," *True Vision-Capture, Transmiss. Display 3D Video*, vol. 1, pp. 1–4, Oct. 2015.
- [30] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological pyramids," *EURASIP J. Image Video Process.*, vol. 21, p. 3, Mar. 2017.
- [31] D. Sandic-Stankovic, 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.
- [32] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.
- [33] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.
- [34] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010.
- [35] H. Attias, "A variational Bayesian framework for graphical models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 2000, pp. 209–215.
- [36] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Visual saliency detection with free energy theory," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1552–1555, Oct. 2015.
- [37] Y. Fang, W. Lin, B.-S. Lee, C.-T. Lau, Z. Chen, and C.-W. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 187–198, Feb. 2012.
- [38] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.
- [39] J. Zhang and S. Sclaroff, "Exploiting surroundedness for saliency detection: A Boolean map approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 889–902, May 2016.
- [40] 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.
- [41] 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, pp. 1–2, Jan. 2010.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 995–1002.



**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 has reviewed over 50 journal papers each year. His research interests include quality assessment, contrast enhancement, visual saliency detection, and air quality prediction. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and received the excellent Ph.D. thesis award from the Chinese Institute of Electronics in 2016. He is the leading special session organizer in VCIP2016 and ICIP2017. He is currently an Associated Editor of the IEEE ACCESS, and is the Reviewer of the IEEE T-NNLS, T-IP, T-MM, TCYB, T-CSVT, T-IE, T-BC, J-STSP, SPL, Access, Information Sciences, Neurocomputing, SPIC, JVCI, DSP, MTAP, ELL, and so on.



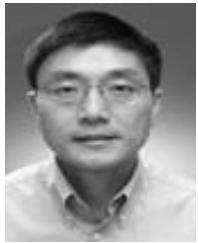
**Vinit Jakhetiya** received the B.Tech. degree in computer and communication engineering from the LNM Institute of Information Technology, India, in 2011, and the Ph.D. degree in electronics and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2016. In 2015, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, where he joined as a Project Officer. He is currently an Assistant Professor with Bennett University, India. His research interests include image/video processing, image quality assessment, and visual perceptual modeling. He was a recipient of the Best Undergraduate Paper Award from the 28th IEEE International Instrumentation and Measurement Technology Conference (2011).



**Jun-Fei 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 member of the IEEE Computational Intelligence Society. 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.



**Xiaoli Li** received the bachelor's and master's degrees in control theory and control engineering from the Dalian University of Technology in 1994 and 1997, respectively, and the Ph.D. degree from Northeastern University, China, in 2000. He is currently a Professor with the Faculty of Information Technology, Beijing University of Technology. His research interests include multiple model control, adaptive control, and visual tracking. In 2006, he was awarded as the Nova of Science and Technology of Beijing Municipality of China. In 2011, he was awarded as the New Century Excellent Talents of Ministry of Education of the People's Republic of China.



**Weisi Lin** (F'16) received the Ph.D. degree from King's College London. He is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has authored or co-authored over 340 refereed papers published in international journals and conferences. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling. He is a fellow of the Institution of Engineering Technology and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been on the Editorial Board of the IEEE TIP, T-MM (2011-2013), SPL, and JVCI. He has been elected as an APSIPA Distinguished Lecturer (2012/13). He served as a Technical-Program Chair at the Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo 2013, and the International Workshop on Quality of Multimedia Experience 2014.



**Daniel Thalmann** received the Ph.D. degree in computer science from the University of Geneva, Geneva, Switzerland, in 1977, and the Ph.D. degree (Hons.) from University Paul Sabatier, Toulouse, France, in 2003. He is currently with the Institute for Media Innovation, Nanyang Technological University, Singapore. He is also a pioneer in research on virtual humans. He has been the Founder of the Virtual Reality Laboratory with École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; a Professor with the University of Montreal, Montreal, QC, Canada; and a Visiting Professor/Researcher with CERN, Geneva, the University of Nebraska-Lincoln, Lincoln, NE, USA, The University of Tokyo, Tokyo, Japan, and The National University of Singapore, Singapore. He has authored over 500 papers in graphics, animation, and virtual reality. He has co-edited 30 books, and co-authored several books, including *Crowd Simulation 2nd Edition* (2012). He received the Eurographics Distinguished Career Award in 2010 and the Canadian Human Computer Communications Society Achievement Award in 2012. He is the Co-Editor-in-Chief of the *Journal of Computer Animation and Virtual Worlds*, and the Editorial Board Member of six other journals. He was the Program Chair and Co-Chair of several conferences, including the IEEE Virtual Reality, the ACM Symposium on Virtual Reality Software and Technology, the ACM International Conference on Virtual Reality Continuum and Its Applications in Industry, the Computer Graphics International Conference, and the International Conference on Computer Animation and Social Agents.