

# USING MULTISCALE ANALYSIS FOR BLIND QUALITY ASSESSMENT OF DIBR-SYNTHESIZED IMAGES

Ke Gu<sup>†</sup>, Jun-Fei Qiao<sup>†</sup>, Patrick Le Callet<sup>‡</sup>, Zhifang Xia<sup>§</sup>, and Weisi Lin<sup>§</sup>

<sup>†</sup>Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

<sup>‡</sup>Luman Université, Université de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes, France

<sup>§</sup>The State Information Center of P.R.China, Beijing, China

<sup>§</sup>School of Computer Science and Engineering, Nanyang Technological University, Singapore

Email: guke.docotr@gmail.com; guke@bjut.edu.cn

## ABSTRACT

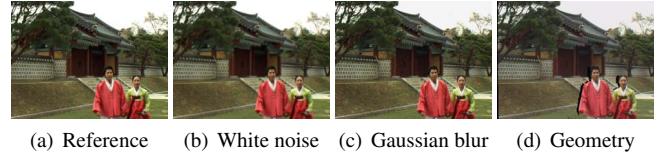
In this paper we propose to blindly evaluate the quality of images synthesized based on a depth image-based rendering (DIBR) procedure. As an important branch in virtual reality (VR), superior DIBR techniques provide free viewpoints in many real applications such as remote surveillance and education, but few efforts have been made to measure the performance of DIBR methods (i.e. the quality of DIBR-synthesized images), especially in the condition of reference unavailable. To this aim, we put forward a new no-reference (NR) image quality assessment (IQA) model via multiscale analysis, dubbed as MSA. The design philosophy of our proposed MSA model is that the DIBR-introduced geometry distortions damage the self-similarity characteristic of natural images and the damage degrees present regular variations at distinct scales. Through systematically incorporating the measurements of the variations provided above, our MSA model can faithfully predict the quality of images generated using different DIBR technologies. Results of experiments demonstrate that the proposed blind MSA model has delivered noticeably better performance than state-of-the-art full- and no-reference IQA methods.

**Index Terms**— Image quality assessment (IQA), blind/no-reference (NR), depth image-based rendering (DIBR), virtual reality (VR), multiscale analysis

## 1. INTRODUCTION

In numerous practical systems such as remote education and surveillance, medical and entertainment applications, etc, free viewpoint videos (FVV) and its 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, which provides the users with more flexible selection of direction and viewpoint [1]. When

This work was supported by the National Natural Science Foundation of China under Grants 61533002.

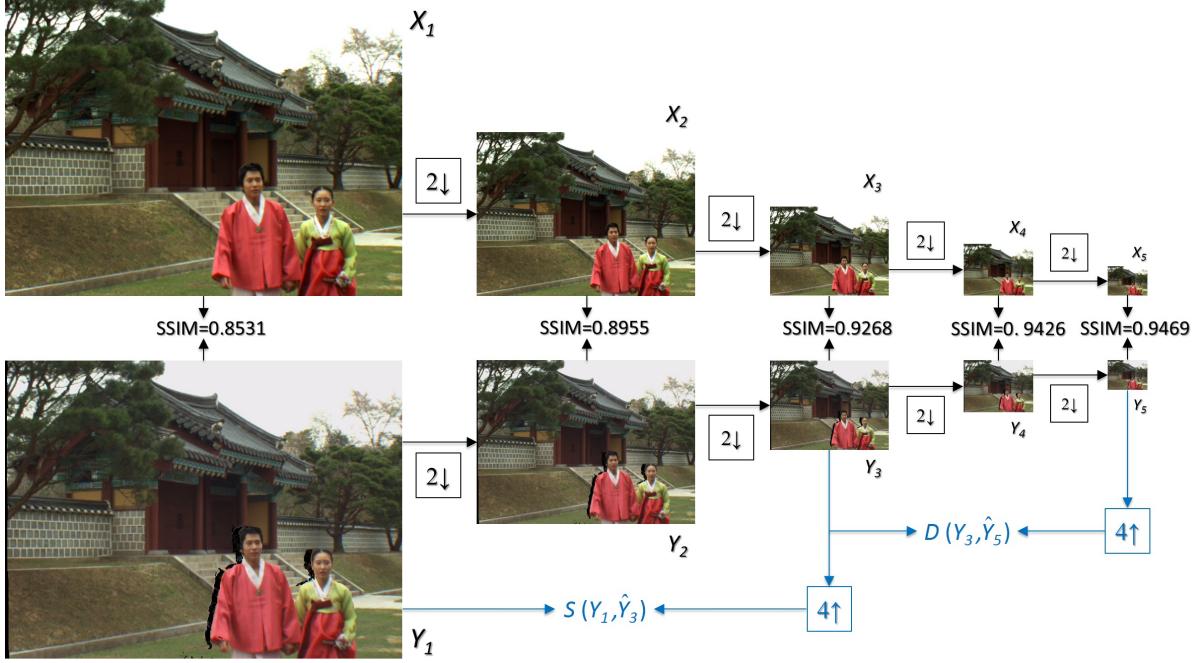


**Fig. 1.** Illustration of the difference between typical distortions (white noise and Gaussian blur) and the geometry distortion.

generating new frames that were once totally non-exist, distortions (actually geometry distortions) are inevitably introduced during this process. So a reliable quality assessment and monitoring technique is highly desirable.

As compared with typical artifacts, e.g. blur, noise, compression, contrast change [2, 3, 4, 5], which have been deeply researched in the last few decades, the geometry distortion presents a quite different visual degradation, as illustrated in Fig. 1. Many IQA algorithms, no matter for full reference (FR) [6, 7, 8, 9] or for no reference (NR) [10, 11, 12, 13, 14], cannot yield reliable quality predictions in assessing DIBR-synthesized images. However, there exist very limited efforts devoted to the IQA problem of DIBR-synthesized views. In [15], the VSQA metric was developed by considering the influences of orientation, texture and contrast to modify the SSIM model. In [16], the 3D-SWIM metric was proposed to conduct the statistical comparisons of wavelet subbands between a reference DIBR-synthesized image and its corrupted version. In [17, 18], MW-PSNR and MP-PSNR algorithms were respectively explored based on morphological wavelet decomposition and morphological pyramid decomposition for evaluating the quality of DIBR-synthesized views. The aforementioned IQA metrics were designed specifically for DIBR-synthesized quality evaluation based on the benchmark reference images. This might highly restrict their application scopes since reference images are usually not accessible in FVV systems.

To cope with the above problem, we attempt to develop a new IQA model used for blindly predicting the quality of DIBR-synthesized views. The proposed NR-IQA model is



**Fig. 2.** Comparison of reference and DIBR-synthesized images at multiscale.  $2\downarrow$  : downsampling by 2.  $4\uparrow$  : upsampling by 4.

based on multiscale analysis and we call it MSA for simplicity. The design theory behind our MSA model lies in two observations: one is that the DIBR-introduced geometry distortion damages the self-similarity characteristic of natural images; the other is that the damage degree tends to decrease as the scale reduces. Through a valid combination of damage degrees at distinct scales, the proposed MSA model is able to blindly generate a reliable quality estimation of an input DIBR-synthesized image. Experiments validate the efficacy of our MSA quality model.

The rest of the paper is organized as follows. Section 2 illustrates the design principle and implementation details of our proposed blind MSA model. Section 3 compares the MSA model with state-of-the-art FR- and RR-IQA metrics on the IRCCyN/IVC database [1]. Section 4 derives some concluding remarks.

## 2. QUALITY ASSESSMENT MODEL

As for natural images, the self-similarity characteristic is an essential attribute and it has been broadly applied to many applications such as image/video compression. One can see from Fig. 1 that, in comparison to white noise or Gaussian blur which affects the global image self-similarity attribute, the geometry distortion merely changes the self-similarity characteristic in some typical local regions while 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 Figs. 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 refers to the DIBR-synthesized image. The SSIM values [2] accurately reflect the variation tendency stated above<sup>1</sup>. If the reference image  $X_i$ , where  $i = \{1, 2, \dots, 5\}$ , 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.

However, in real applications, reference free viewpoint images are usually unavailable, so the DIBR is required to generate virtual free viewpoint images. In other words, blind quality assessment of the DIBR-synthesized view is more close to the practical application scenarios. A straightforward way to solve the above issue 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  $D(Y_3, X_3) \approx D(Y_3, \hat{X}_5) \approx D(Y_3, \hat{Y}_5)$ , where ‘ $\hat{\cdot}$ ’ is a perfect upsampling operator used to keep the two inputs having the matchable size. As thus, we can compute  $D(Y_i, \hat{Y}_5)$  to approximate  $D(Y_i, X_i)$ , where  $i = \{1, 2, \dots, 5\}$ . But unfortunately, current upsampling technologies are far from ideal due to the introduction of a large amount of blur distortion, and thus one more approximation is needed.

We treat the above problem from another point of view.

<sup>1</sup>The greater SSIM value indicates the higher similarity and the lower distance between the two compared image signals.

$X_1, X_2, \dots, X_5$  are reference images having 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  $i = \{1, 2, 3, 4, 5\}$ . 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,  $S(Y_1, \hat{Y}_i)$  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, such as SRCC [19, 20], rather than precise estimations of subjective scores. Therefore, this paper uses  $S(Y_1, \hat{Y}_i)$  to be an alternative of  $D(Y_i, \hat{Y}_5)$  (or  $D(Y_i, X_i)$ ), where  $i = \{1, 2, 3, 4, 5\}$ . Despite the use of upsampling during the computation of  $S(Y_1, \hat{Y}_i)$ , there is a big difference. As for  $D(Y_i, \hat{Y}_5)$ , we upsample  $Y_5$  as the benchmark image for comparison, while  $Y_1$  is itself the benchmark when computing  $S(Y_1, \hat{Y}_i)$ . This alternative can well avoid the problem of finding perfect upsampling operator. Results of trials also verify this statement, leading to an about 15% performance gain (in terms of SRCC). In our work, the very efficient bilateral interpolation technology is used and more advanced upsampling methods will be considered in future researches.

Multiscale analysis is a significant attribute of the human visual system (HVS), which has been widely used in many image processing fields, for example, quality evaluation [21] and saliency detection [22]. Following the multiscale analysis in [21], we fuse each similarity map and derive

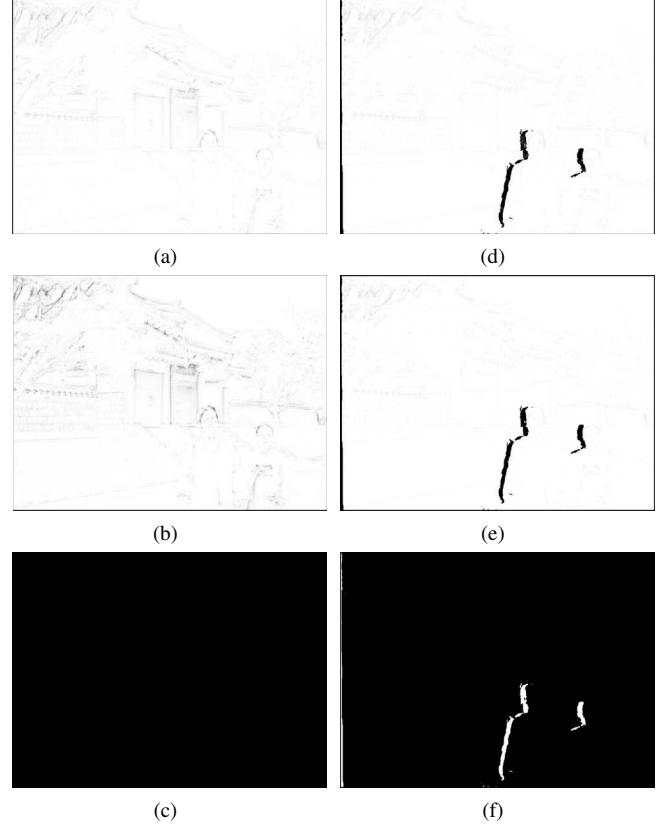
$$\bar{S}_j = \prod_{i=1}^N [S_j(Y_1, \hat{Y}_i)]^{\gamma_i}. \quad (1)$$

where  $N = 5$ ;  $j$  indicates the pixel index;  $\{\gamma_1, \gamma_2, \dots, \gamma_5\}$  are assigned according to a psychophysical experiment as  $\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ . We leverage the commonly used similarity measure that has three merits of symmetry, boundedness, and unique maximum:

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

where  $\varepsilon$  is a constant number for avoiding division-by-zero. 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  $\bar{S}$  and generate  $\dot{S}$  in order to remove isolated noisy pixels. Furthermore, even for natural images, distortions which do not originally occur will be included into the fused similarity map  $\ddot{S}$  since the bilateral interpolation method is not perfect upsampling technology and must introduce blur distortions. Therefore, we exert a thresholding on the filtered similarity map  $\dot{S}$  for extracting geometry distorted regions:

$$\ddot{S}_j = \begin{cases} 1, & \text{if } \dot{S}_j < T \\ 0, & \text{otherwise} \end{cases} \quad (3)$$



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

where  $T$  is a small threshold set as 0.1. The newly generated blur distortions are beyond this threshold and thus removed. By contrast, the geometry distorted regions are beneath this threshold and thus preserved. In Fig. 3, we show  $\hat{S}$ ,  $\dot{S}$  and  $\ddot{S}$  of the reference and DIBR-synthesized images (i.e. Fig. 1(a) and (d)). Comparing Figs. 3(c) and (f), one can easily see that  $\ddot{S}$  well extracts the geometry distorted areas.

Eventually, we blindly estimate the overall quality score of an input DIBR-synthesized image as

$$Q_{MSA} = \frac{1}{L} \sum_{j=1}^L \ddot{S}_j \quad (4)$$

where  $L$  stands for the number of all the pixels in  $\ddot{S}$ . It is worthy to stress that the smaller  $Q_{MSA}$  value refers to less geometry distortion and larger subjective mean opinion score (MOS), and vice versa.

### 3. PERFORMANCE VALIDATION

In this section, we apply the proposed blind MSA model on the IRCCyN/IVC database [1] to validate its performance. The IRCCyN/IVC database was constructed in the year of 2011, dedicated to the DIBR-synthesized views. Numerous

**Table 1.** Performance comparison on the IRCCyN/IVC database [1].

Metrics	Type	SRCC	KRCC	PLCC	MAE	RMSE
SSIM [2]	FR	0.4671	0.3194	0.5636	0.4530	0.5500
FSIM [6]	FR	0.4148	0.2738	0.5828	0.4217	0.5411
IGM [7]	FR	0.4893	0.3385	0.5476	0.4717	0.5571
ADD-SSIM [8]	FR	0.5451	0.3939	0.6130	0.4273	0.5261
PSIM [9]	FR	0.4576	0.3033	0.5315	0.4269	0.5640
MW-PSNR [17]	FR	0.6634	0.4758	0.7166	0.3434	0.4644
MP-PSNR [18]	FR	<b>0.6995</b>	<b>0.5064</b>	<b>0.7278</b>	<b>0.3312</b>	<b>0.4566</b>
NIQE [10]	NR	0.3739	0.2421	0.4374	0.4641	0.5987
SISBLIM [11]	NR	0.3832	0.2721	0.5225	0.4523	0.5677
IL-NIQE [12]	NR	0.5348	0.3668	0.4998	0.4415	0.5767
ASIQE [13]	NR	0.4948	0.3443	0.5854	0.4163	0.5398
MSA (Proposed)	NR	<b>0.6521</b>	<b>0.5258</b>	<b>0.6393</b>	<b>0.3938</b>	<b>0.5120</b>

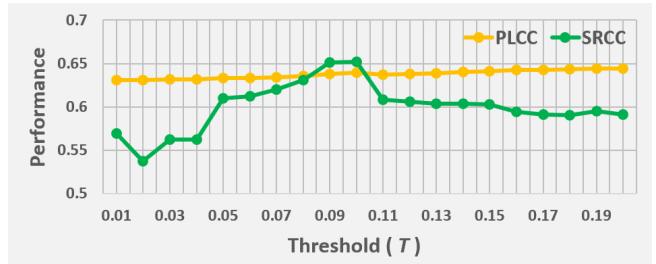
mainstream and state-of-the-art IQA metrics are included for comparison. They are divided into two types: one includes seven FR-IQA methods, namely SSIM [2], FSIM [6], IGM [7], ADD-SSIM [8], PSIM [9], MW-PSNR [17], and MP-PSNR [18]; the other one includes four opinion-unaware (OU) NR-IQA methods, namely NIQE [10], SISBLIM [11], IL-NIQE [12], and ASIQE [13].

To compute the correlation performance, five frequently applied evaluation criteria are taken advantage of. They are Spearman rank order correlation coefficient (SRCC), Kendall's rank-order correlation coefficient (KRCC), Pearson linear correlation coefficient (PLCC), mean absolute error (MAE), and root mean square error (RMSE) respectively. SRCC and KRCC measure the prediction monotonicity between objective quality predictions and subjective opinion scores, while the other three focus on the measurement of prediction consistency. Before computing PLCC, MAE and RMSE, the nonlinearity of objective quality predictions is required to be eliminated based on the subsequent five parameter nonlinear logistic function [23]:

$$f(Q_{\text{MSA}}) = \beta_1 \left( 0.5 - \frac{1}{1 + e^{\beta_2(Q_{\text{MSA}} - \beta_3)}} \right) + \beta_4 Q_{\text{MSA}} + \beta_5 \quad (5)$$

where  $Q_{\text{MSA}}$  and  $f(Q_{\text{MSA}})$  are respectively the raw quality predictions and the associated mapped scores.  $\{\beta_1, \dots, \beta_5\}$  are the parameters to be fitted during the nonlinear regression. A superior quality assessment model should obtain a greater value of SRCC, KRCC, PLCC, and meanwhile, obtain a smaller value of MAE and RMSE.

We list the performance results in Table 1. For making the readers' conveniences, we highlight the best performing model among all the FR-IQA methods and all the NR-IQA algorithms. Specifically, in comparison to testing blind IQA metrics, the proposed MSA model has achieved much better performance. As compared with FR-IQA metrics, our blind MSA model is also superior to most of them, except inferior to recent MW-PSNR and MP-PSNR metrics which were



**Fig. 4.** Performance dependency with the changing thresholds ( $T$ ).

dedicated to DIBR-synthesized IQA. Note that FR metrics require the entire reference image and this might seriously constrain their application scenarios. We further examine the influence of parameters on prediction performance. There is only one variable parameter used in the MSA model, namely the threshold  $T$ . We enumerate 20 numbers from 0.01 to 0.2 with an interval of 0.01. As shown in Fig. 4, the performance measures are affected very little. The least values of PLCC and SRCC are 0.63 and 0.54, still far beyond all the NR-IQA methods tested. This verifies the robustness of the proposed blind MSA model.

#### 4. CONCLUSIONS

This paper has proposed a novel training-free NR quality assessment model of DIBR-synthesized views. Our quality model is based on the following principle; that is, the self-similarity characteristic of natural images is damaged by the DIBR-introduced geometry distortion, and the damage degree tends to reduce as the scale becomes smaller. Exploiting an effective integration of damage degrees at distinct scales, the proposed blind MSA model can yield a reliable quality prediction of a given DIBR-synthesized image. Results of experiments demonstrate that our MSA model has acquired very high performance in comparison to state-of-the-art FR- and NR-IQA methods.

## 5. REFERENCES

- [1] E. Bosc, R. P  pion, P. Le Callet, M. K  ppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J-STSP*, vol. 5, no. 7, pp. 1332-1343, Sep. 2011.
- [2] 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. Online at: <http://live.ece.utexas.edu/research/quality>
- [3] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *JEI*, vol. 19, no. 1, Mar. 2010. Online at: <http://vision.okstate.edu/csiq>
- [4] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcasting*, vol. 61, no. 3, pp. 520-531, Sep. 2015.
- [5] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284-297, Jan. 2016.
- [6] 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.
- [7] 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.
- [8] 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. Broadcasting*, vol. 62, no. 2, pp. 446-456, Jun. 2016.
- [9] 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.
- [10] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE SPL*, vol. 22, no. 3, pp. 209-212, Mar. 2013.
- [11] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcasting*, vol. 60, no. 3, pp. 555-567, Sep. 2014.
- [12] 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.
- [13] 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.*, 2017.
- [14] 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. Learning Syst.*, 2017.
- [15] P. H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," *Electron. Imag. Int. Society for Optics and Photonics*, pp. 8288-8256, Feb. 2012.
- [16] F. Battisti, E. Bosc, M. Carli, and P. Le Callet, "Objective image quality assessment of 3D synthesized views," *SPIC*, vol. 30, pp. 78-88, Jan. 2015.
- [17] D. S. Stankovic, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological wavelets," in *QoMEX*, pp. 1-6, Jan. 2015.
- [18] D. S. Stankovic, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological pyramids," *The True Vision-Capture, Transmission and Display of 3D Video*, pp. 1-4, Oct. 2015.
- [19] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J-STSP*, vol. 3, no. 2, pp. 193-201, Apr. 2009.
- [20] 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.
- [21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *IEEE ACSSC*, pp. 1398-1402, Nov. 2003.
- [22] C. Kim and P. Milanfar, "Visual saliency in noisy images," *J. Vision*, vol. 13, no. 4, pp. 1-14, Mar. 2013.
- [23] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000, <http://www.vqeg.org/>.