A Prediction backed Model for Quality Assessment of Screen Content and 3D synthesized Images

Vinit Jakhetiya, Ke Gu, Weisi Lin *Fellow, IEEE,* Qiaohong Li, and Sunil Prasad Jaiswal

***Abstract*—In this paper, we address problems associated with free-energy-principle based image quality assessment (IQA) algorithms for objectively assessing the quality of Screen Content (SC) and 3D synthesized images and also propose a very fast and efficient IQA algorithm to address these issues. These algorithms separate an image into predicted and disorder residual parts and assume disorder residual part does not contribute much to the overall perceptual quality. These algorithms fail for quality estimation of SC images as information of textual regions in SC images are largely separated into the disorder residual part and less information in the predicted part and subsequently, given a negligible emphasis. However, this is in contrast with the characteristics of human vision. Since our eyes are well trained to detect text in daily life. So, our human vision has prior information about text regions and can sense small distortions in these regions. In this paper, we proposed a new reduced-reference IQA algorithm for SC images based upon a more perceptually relevant prediction model and distortion categorization, which overcomes problems with existing free-energy-principle-based predictors. From experiments, it is validated that the proposed model has a better capability of efficiently estimating the quality of SC images as compared to the recently developed reduced-reference IQA algorithms. We also applied, the proposed algorithm to judge the quality of 3D synthesized images and observed that it even achieves better performance than the full-reference IQA metrics specifically designed for the 3D synthesized views.**

***Index Terms*—Screen content images, 3D synthesized images, prediction, image quality assessment (IQA), distortion categorization, textual region, human vision.**

I. Introduction

Recently, Screen Content (SC) images [1] are used in several applications, such as virtual screen sharing [2], online gaming, remote education, etc. These SC images contain regions with both text and pictures, thus have different characteristics to natural images [1]. With this in view, researchers have aimed to develop compression methods specifically for SC images and videos [3-5], as the existing algorithms performed poorly on SC images/videos. The importance of SC images is demonstrated by the fact that separate proposals were called to efficiently compress SC images, in the extension of HEVC [6]. With the increasing applications and usage of SC images/videos, methods to accurately and automatically quantify their quality are required.

Similar to SC images, free viewpoint videos (FVV) have aroused much attention because of its broad range of utility values in the fields of entertainment, medical applications, remote education, etc. In free viewpoint videos, the need is to produce fresh viewpoint frames from the adjacent multiple views [7]. The techniques which create those new viewpoints are called the Depth-Image-Based-Rendering (DIBR) techniques [7]. These newly generated views are affected by several distortions and geometric distortion is the most important among them. Contrary to the commonly occurring structure distortions in natural images, the geometric distortion has very different characteristics [7], [8] and is thus difficult to measure. The overall quality of a whole free viewpoint video sequence will be heavily degraded even though a few DIBR-synthesized views contained therein are polluted. As a consequence, an accurate quality assessment method of 3D-synthesized images is fairly important.

During the last few decades, many image quality assessment (IQA) algorithms have been proposed to assess the perceptual quality of natural images. These IQA algorithms are broadly divided into three categories, namely full-reference (FR), reduced-reference (RR) and no-reference (NR). In [9], Wang et al. proposed an FR metric namely, SSIM which is based upon the structural similarity between the reference and distorted images. In [10] and [11], the authors assume that the human visual system (HVS) performs two different ways to perceive the image quality, i.e., near-threshold and suprathreshold. The authors of [12] and [13] proposed FR algorithms using the gradient similarity and deviation of gradient similarity, respectively. For RR algorithms, the authors of [14] and [15] used the structural degradation model and orientation selectivity based visual pattern's, respectively. The authors of [16] proposed an RR algorithm for stereoscopic images using the binocular perceptual information. In [17], Soundrarajan et al. proposed an RR algorithm namely, RRED using the entropy differencing of original and reference image. As for the studies of NR IQA, some state-of-the-art quality metrics were developed based on early human vision modeling [18] or some natural scene statistics (NSS) models [19-20]. However, they all were found to take effects on the IQA of natural scene images only, but fail to faithfully assess the quality of SC images.

Recently, researchers have also paid attention to the quality assessment of SC images and the authors of [1] and [21]

This work was supported in part by National Natural Science Foundation of China under Grants 61533002, 61703009, Nova Programme Interdisciplinary Cooperation Project under Grant Z161100004916041, Singapore MoE Tier 1 Project M4011379 and RG141/14.

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

K. Gu (Corresponding Author) is with Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (email: guke@bjut.edu.cn).

W. Lin, and Q. Li are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798 (email: wslin@ntu.edu.sg, qli013@e.ntu.edu.sg).

S. P. Jaiswal is with Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: spjaiswal@connect.ust.hk).

modified the SSIM [9] algorithm to better match the characteristics of the SC images. In [22], Wang et al. proposed an IQA algorithm for SC images based upon the gradient magnitude and blur functions. While Gu et al. used saliency detection [23] and machine learning approach [24] to quantify the quality of SC images. Similar to the SC images, researchers have paid attention to the quality assessment of 3D synthesized images [8] and authors of [25] and [26] used morphological wavelets and morphological pyramids, respectively for this purpose. To the best of our knowledge, no RR IQA algorithm has been proposed for quality evaluation of both SC and 3D synthesized images. In some applications, such as image/video transmission, complete information about the reference signal is unavailable, which makes FR methods impractical. For NR methods, without any information about the reference, it is difficult to achieve satisfactory quality prediction performance. It is more practical to send limited information about the reference to the decoder side to perform a quality assessment. With this in view, we propose a new RR algorithm, specifically for SC images.

Free-energy-principle has been utilized in several IQA algorithms in recent years. Zhai et al. [27] proposed an RR algorithm based upon the residual uncertainty; Wu et al. proposed FR [28] and RR [29] metrics, assuming that the residual uncertainty and primary visual information cause different level of visual sensations and along the same line, authors of [30], [31] and [32] proposed blind IQA algorithms based on the free-energy principle and learning.

The internal generative mechanism [28] of our brain tries to predict the primary visual information and to discard uncertainty. Generally, these free-energy-principle based IQA algorithms separate an image into predicted and disorder residual parts using auto-regressive (AR) modeling. The disorder residual part refers to residual errors between the interest image and its predicted image. They also assume that human vision can mostly sense distortions in the predicted part and disorder residual part does not contribute much to the perception of image quality. Unfortunately, when applying such prediction methods on SC images, most of the information of textual regions are separated into the disorder residual part and consequently, assume textual regions are not significantly important for judging the quality of the SC images.

Contrary to this assumption, human vision has prior information about the text, and can easily perceive small distortions in textual regions. Existing free-energy based IQA methods underestimate the contribution of the textual region and thus cannot perform effectively for SC images. In order to overcome such problem, we propose an IQA algorithm using a local predictor which separates most of the information of the textual region into the predicted part and build an RR IQA method that gives importance to the distortion in the textual regions during the estimation of the quality score for SC images.

II. Proposed Algorithm

In this paper, we propose an RR IQA algorithm for SC images based upon the prediction and distortion categorization. In the proposed algorithm, we predict the SC image in such a way that most of the information of the textual region are separated in the predicted part and textual region can play an important role when evaluating the quality.

A. Free-energy-principle-based IQA models for SC images

The free-energy-principle based IQA algorithms [28], [27], [29] separate an image into two parts: the predicted part, which is mainly the visual structural information, and the disorder residual part. These algorithms assume that the degradation on the predicted part has a significant impact on perceptual quality, while the disorder residual part has only a small influence on the overall perceptual quality.

These algorithms rely on the fact that the correlation between the central pixels and their neighboring pixels is quite high for pixels which contain primary visual information and they can be efficiently predicted using ordinary least squares (OLS) based autoregressive (AR) modeling. The OLS based AR modeling assumes that center pixel and its neighboring pixels in a local window are stationary and it can efficiently predict perceptually important pixels. In other words, the pixels which can not be efficiently predicted using OLS based AR modeling have more information in the disorder residual part and distortions in such regions do not cause much discomfort to human vision.

These assumptions are valid for natural images, but for SC images, the correlation among neighboring pixels is low [33] as compared to natural images and pixels in the neighboring local window are not stationary. In such cases, the assumptions of geometric duality and stationarity are missing [34] and consequently, AR model using OLS based predictors cannot predict textual regions and separate more information of the textual regions to the disorder residual part. In order to support these arguments the predicted and disorder residual parts of an SC image using the predictor adopted by IGM [28] (or RRVIF [29], as the same predictor is used in both of the algorithms) and FEDM [27], are shown in Fig. 1. From the second and third rows of Fig. 1, it can be observed that the AR modeling based predictors are not able to efficiently predict the text regions and these algorithms separate most of the information of the textual regions to the disorder residual part.

Contrary to the assumption made by the existing free-energy-principle based IQA algorithms, our eyes are well trained to see text in daily life and human vision has prior information about the text and is therefore quite sensitive to text and textual regions. Above arguments and importance of textual regions in assessing the quality of SC, images can be supported by the fact that

1) The human vision selects the text region as the salient part [35], [36].
2) From subjective evaluation [1], it has been validated that correlation between quality of text and overall quality is higher than the correlation between quality of picture and overall quality, which suggests that the textual part of an SC image contributes more to the overall perceptual quality than the pictorial part.

First, we give an example to show that human vision is more sensitive to distortions in textual regions than the

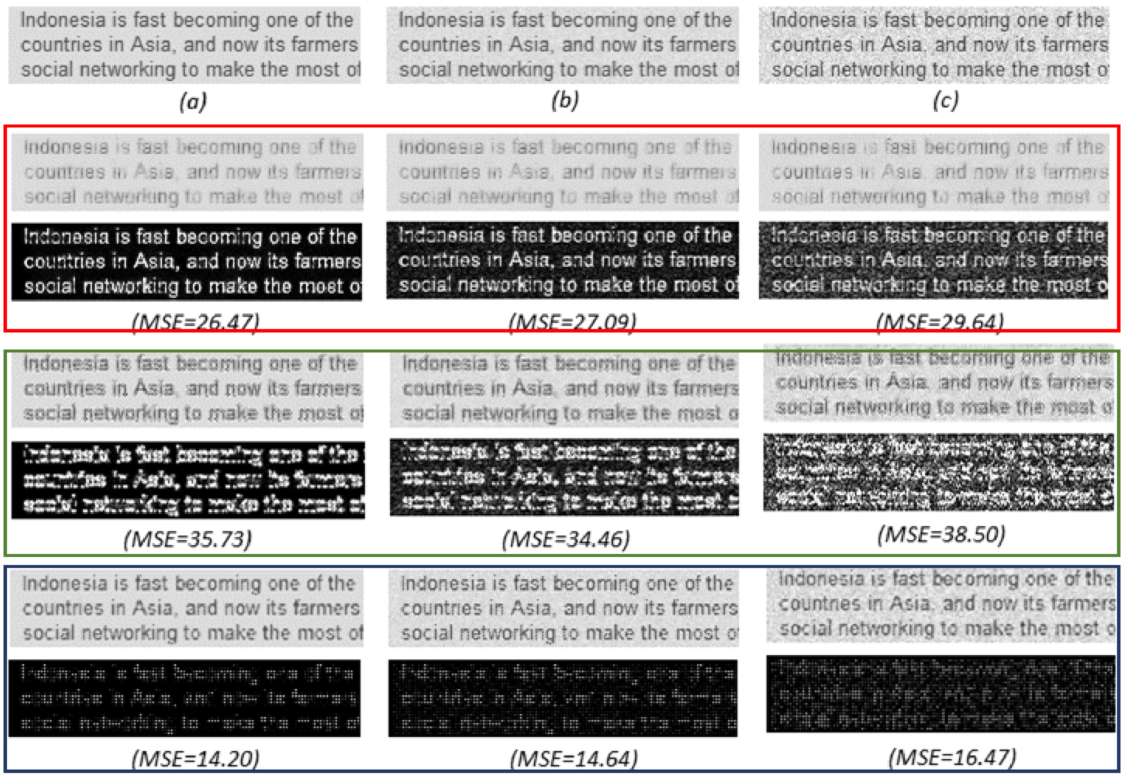

Fig. 1: The inability of AR modeling based predictors to predict the textual regions in existing free-energy principal based IQA algorithms. Here, Fig. 1. (a), (b) and (c) are the original image and distorted images due to Gaussian noise with σ = 20 and 40. The first and second row in red, green and blue blocks represent the predicted part and disorder residual part of Fig. 1. (a)-(c) using the predictors used in IGM [28], FEDM [27] and the proposed algorithm, respectively.

texture regions. Fig. 2 shows a cropped SC image from the SIQAD [1] database and a cropped natural image from TID database [37] and these distorted versions due to Gaussian noise contamination and JPEG compression. From Fig. 2, it can be seen that even a small distortion in the textual region can be easily perceived by human vision. At the same time, distortions in the textural regions in a natural image are not easily perceivable by our human vision, which validates the assumption of free-energy-principle-based IQA algorithms for natural images. This example suggests that our human vision is trained to sense such small distortions in text and textual regions play an important role in assessing the quality of images.

B. Prediction based IQA for SC images:

In the proposed algorithm, our goal is to separate most of the information of the textual regions into the predicted part than in the disorder residual part so that the textual region can significantly influence the overall perceptual quality of SC images. Generally, SC images contain sharp edges and thin lines [33], [32], [38] and local predictors are able to predict efficiently, as compared to block based [33] predictors using least squares or global predictors. With this view, we propose to use an observation-model based bilateral filter (OBF) [34] to separate an SC image into predicted and disorder residual parts. The i^{th} pixel of predicted part ($\hat{X}_d$), which is associated to a distorted image (X_d), is obtained as

$$\hat{X_{d_i}} = \frac{X_{d_i}\lambda + \sum_{k\in N_i}\omega_{k_i}A_{k_i}}{\lambda + \sum_{k\in N_i}\omega_{k_i}}, \tag{1}$$

where A_{k_i} and w_{k_i} are the pixels in the neighboring 3×3 window N_i of the i^{th} pixel and corresponding weights, respectively, while λ is the parameter which controls the prediction accuracy. In this work, we choose value of λ to be 0.1. These weights (w_k) are calculated based upon the pixel gradient and radiometric distance, as suggested by Jakhetiya et al. [34], [38]. The separated disorder residual part of distorted image is obtained as follows

$$R_{d_i} = |\hat{X_{d_i}} - X_{d_i}|. \tag{2}$$

The separated predicted and disorder residual parts of textual regions using the OBF are shown in the last row of Fig. 1. From this figure, one can observe that the OBF predicts textual regions efficiently and it separates most of the information of the textual region to the predicted part and, in turn, significantly consider it when assessing the quality of the SC images.

In the literature, a strength of the edge structures (sharpness) have shown great success in extracting the primary visual information [1], [28], [29], [39] of an image. For this reason, we also use the sharpness similarity between the distorted and reference images to estimate the level of degradation to

Fig. 2: SC and natural images showing the sensitivity and insensitivity of human vision to perceive distortions in the text and textural regions. (a) and (d) original SC and natural images, (b) and (e) distorted images due to Gaussian noise (with MSE = 17.3) and (c) and (f) JPEG compressed images (with MSE = 26.5).

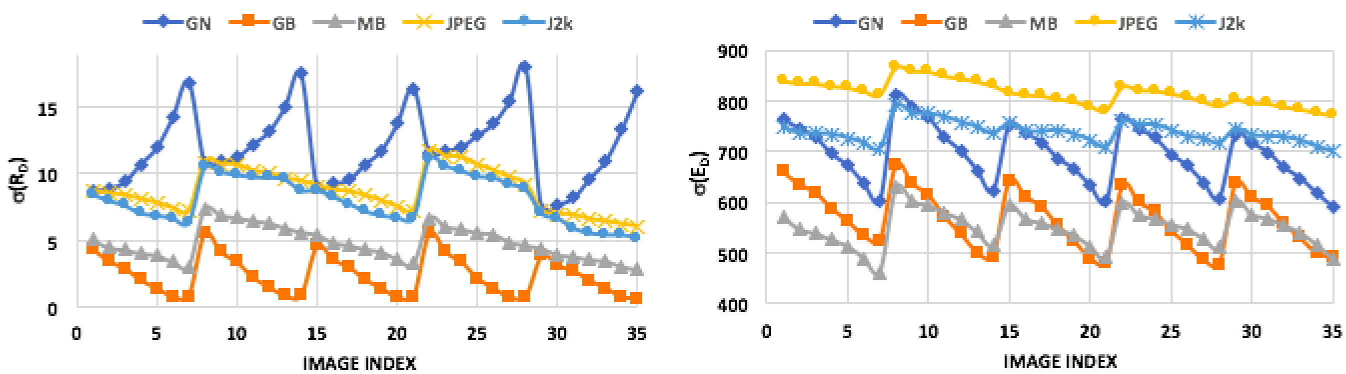

Fig. 3: The dependency of the standard deviation of the residual image ($\sigma(R_d)$) and standard deviation of the sharpness ($\sigma(E_d)$) with an increasing level of distortion for five images from SIQAD database with seven levels of distortions for Gaussian noise, Gaussian blur, motion blur, JPEG, and J2K compression, respectively.

the reference image. The sharpness of the distorted image is estimated as follows:

$$E_d = \max_{k=1....4} \quad Mag_k(X_d) \tag{3}$$

$$Mag_k(X_d) = |X_d \star F_k|, \tag{4}$$

Here, $\star$ represents the convolution operator and F_k represents the high-frequency filters in four directions (horizontal, vertical, diagonal and anti-diagonal) to calculate the sharpness value. More details about these four filters F_k can be found in [29]. The edge structure of an image deteriorates with the addition of distortions, and in turn, image sharpness decreases. With this view, the sharpness of an image can be used to determine the quality of the image. In the proposed algorithm, standard deviation is used to pool sharpness of the reference ($\sigma(E_r)$) and distorted ($\sigma(E_d)$) images. The dependency of $\sigma(E_d)$ with the increasing distortion for 35 distorted images (five images from the SIQAD database [1] with seven increasing levels of distortions) is shown in Fig. 3 (b). From this figure, one can observe that the $\sigma(E_d)$ decreases with the increment of all types of distortions. So as the difference between the sharpness of the distorted and reference images increases, the perceptual quality of the distorted image reduces.

On the other hand, with the increment of low-frequency distortions (such as Gaussian blur, motion blur, JPEG and J2K compression), the standard deviation of the disorder residual image ($R_d = |X_d - \hat{X}_d|$) decreases, as such kinds of distortions make an image smooth. Contrary to this, with a Gaussian noise distortion increment, the high-frequency parts in the distorted image increased, which behave as outliers [34]. These outliers in the distorted image are difficult to predict [34] and consequently, the standard deviation of the residual image ($\sigma(R_d)$) increases, as shown in Fig. 3 (a).

The simplest way is to utilize function used in SSIM [9] to obtain final quality score S but function used in SSIM treats a different kind of distortions equally. The researchers in just noticeable difference [39] area have shown that high-frequency distortions cause more sensation to human vision as compared to low-frequency distortions and high-frequency distortions more degrade the overall perceptual quality of an image than the low-frequency distortions. So, the impact of high-frequency distortions (i.e., Gaussian noise) should be distinguished from the low-frequency distortions (Gaussian and motion blur, JPEG and J2K compression) for quantifying the quality of an SC image. With this view, the final quality score using the proposed algorithm is calculated as

$$S = \begin{cases} \left[\frac{\sigma(R_r)}{\sigma(R_d)}\right]^{\alpha} \times \left[\frac{\sigma(E_d)}{\sigma(E_r)}\right]^{\beta} & \text{if } \sigma(R_d) \geq \sigma(R_r) \\ \\ \left[\frac{\sigma(R_d)}{\sigma(R_r)}\right]^{\gamma} \times \left[\frac{\sigma(E_d)}{\sigma(E_r)}\right]^{\beta} & \text{else.} \end{cases} \quad (5)$$

In equation (5), α should attain a higher value that γ, as Gaussian noise distortions impact the overall perceptual quality more than the low-frequency distortions. The proposed algorithm requires only two reference data to assess the quality of an image, namely the standard deviation of the sharpness ($\sigma(R_r)$) and standard deviation of the disorder residual error ($\sigma(E_r)$). We choose to quantize $\sigma(R_r)$ and $\sigma(E_r)$ using a 16-bit quantizer, so the proposed algorithm requires to send only 32-bit information about the reference image to the decoder side during transmission.

In (5), α, β and γ and in (1) λ are the fixed parameters, which need to be determined reliably. Big data is a very important concept, which has been broadly used in various directions [40-44]. Particularly, similar to a recent work [24], we have first collected thousands of "webpage" and "screen snap" images from the "Google Images" website, and then maintained 1,000 high-quality images with human eyes' observations. We apply the aforesaid 1,000 high-quality SC images as original references to generate above 100,000 images as training samples with six typical distortion types, which include Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), and JPEG and JPEG2000 (JP2K) compressions, and 15 distortion levels for each type. Next, we apply the state-of-the-art SQMS metric [23] to label those distorted SC images. The parameters used in our proposed quality metric, α, β, γ, and λ, are determined to make our metric have the highest correlation with the SQMS metric in terms of prediction monotonicity. Note that the above training images are exactly content-independent of those in the SIQAD database, and this means those parameters are quite reliable and generalized. Based upon the above described independent experiments, the values of α, β, γ, and λ values are chosen as 0.5, 0.2, 0.1 and 1.9, respectively.

III. Experimental Results

In order to verify our proposed algorithm, we applied it to the recently developed screen image quality assessment database (SIQAD) [1] to judge the quality of SC images.

A. IQA Metrics used for Comparison and Evaluation Methodology

The SIQAD database contains 20 reference SC images and corresponding 980 distorted images which are distorted due to seven prevailing distortions: Gaussian noise, Gaussian blur, motion blur, contrast change, JPEG compression, J2K compression, and layer segmentation based coding. In SIQAD database each image has 7 different level of distortions. The proposed reduced-reference algorithm is compared with five well-known full-reference IQA algorithms, SSIM [9], IGM [28], GSIM [12], SQMS [23] and VIF [45]; and five recently developed reduced-reference algorithms: OSVP [15], FEDM [27], RRED [17], RRVIF [29] and RQMSH [22].

The proposed algorithm is evaluated in terms of three criteria, namely the Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), and root means squared error (RMSE), as suggested by the Video Quality Expert Group (VQEG) [46]. A better IQA algorithm should attain a lower value of the RMSE and higher value of the PLCC and SRCC. In order to remove the non-linearity of the objectively predicted scores, a five parameterized non-linear logistic function is used which is defined as:

$$f(Q_s) = \Delta_1 \times \left(\frac{1}{2} - \frac{1}{1 + e^{\Delta_2(Q_s - \Delta_3)}}\right) + \Delta_4 Q_s + \Delta_5. \quad (6)$$

In equation (6), Q_s represents the predicted scores obtained from the IQA algorithms and $f(Q_s)$ represents the respective mapped scores. While Δ_i (i $\in$ 1, 2, 3, 4, 5) are the model parameters which are estimated using a non-linear regression.

B. Performance Comparison

Simulation results (in terms of PLCC, SRCC, RMSE and running time) for the SIQAD database using the proposed algorithm, five FR metrics, and five RR metrics are presented in Table I. From Table I, it can be observed that the proposed algorithm performs significantly better than the other RR metrics (OSVP, FEDM, RRED, and RRVIF). At the same time, the proposed algorithm achieves a better performance than several FR IQA metrics, except the SSIM [9] and GMSD [13] algorithms but these algorithms require complete information about the reference SC image.

It is also interesting to note from Table I that the proposed algorithm performs much better than the existing IQA algorithms based upon the free-energy-principle (IGM [28], FEDM [27] and RRVIF [29]), which are designed for assessing the perceptual quality of natural images. From Table I, one can observe two important aspects:

1) The proposed algorithm achieves PLCC value of 0.7264, while IGM, FEDM, and RRVIF algorithms are only able to achieve 0.6422, 0.5388, 0.5758, respectively. These simulation results validate the assertion that the existing free-energy-principle based IQA algorithms separate most of the information of the textual regions to the disorder residual part and in turn assume that the textual regions do not significantly affect the overall perceptual quality. While the proposed algorithm separates most of the information of the textual region in the predicted part and consequently, considers distortions in these regions during the estimation of the quality score.
2) The running time of the existing free energy principle based IQA algorithms is quite high as compared to the proposed algorithm, as these algorithms use least-square

TABLE I: Comparison results of the proposed algorithm with five FR (SSIM [9], IGM [28], GSIM [12], SQMS [23] and VSI [45]) and five RR algorithms (OSVP [15], FEDM [27], RRED [17], RRVIF [29], and RQMSH [22]) in terms of PLCC, SRCC and RMSE for the SIQAD database. The last row represents the running time (in seconds) of corresponding algorithm to predict the quality of an SC image. Here N is the number of pixels in the SC image.

		Full-Reference (FR)					Reduced-Reference (RR)					
	Distortions	SSIM	IGM	GSIM	SQMS	VSI	OSVP	FEDM	RRED	RRVIF	RQMSH	Prop.
	No. of scalers	N	N	N	N	N	9	2	9	2	1	2
PLCC	GN	0.8806	0.9017	0.8448	0.9004	0.8836	0.7571	0.7793	0.8963	0.8657	0.8876	0.8831
	GB	0.9014	0.8858	0.8831	0.9126	0.8504	0.8794	0.7840	0.8939	0.8830	0.8698	0.8960
	MB	0.8060	0.7655	0.7711	0.8673	0.7658	0.8186	0.5020	0.8105	0.7350	0.8135	0.6328
	CC	0.7435	0.8203	0.8077	0.8027	0.7734	0.8116	0.7170	0.7347	0.7567	0.6873	0.7559
	JPEG	0.7487	0.7980	0.6778	0.7857	0.7149	0.4231	0.4365	0.7857	0.6912	0.5965	0.7594
	J2K	0.7749	0.8324	0.7242	0.8263	0.7498	0.0939	0.1317	0.7673	0.7647	0.5948	0.7952
	LSC	0.7307	0.8287	0.7218	0.8126	0.7457	0.4269	0.0794	0.8215	0.7321	0.5701	0.6898
	Overall	**0.7561**	0.6422	0.5663	**0.8872**	0.5568	0.6341	0.5388	0.5557	0.5758	0.7555	**0.7264**
SRCC	GN	0.8694	0.8819	0.8404	0.8860	0.8655	0.7607	0.7675	0.8810	0.8479	0.8727	0.8750
	GB	0.8921	0.8766	0.8796	0.9149	0.8495	0.8730	0.7660	0.8808	0.8715	0.8595	0.8865
	MB	0.8041	0.7620	0.7753	0.8695	0.7658	0.8140	0.4795	0.8058	0.7214	0.8107	0.6295
	CC	0.6405	0.6777	0.7148	0.6948	0.6459	0.7093	0.5120	0.5601	0.6493	0.5031	0.6678
	JPEG	0.7576	0.7936	0.6796	0.7893	0.7196	0.3998	0.4010	0.7700	0.6803	0.5913	0.7464
	J2K	0.7603	0.8211	0.7125	0.8194	0.7299	0.1802	0.1286	0.7581	0.7588	0.5758	0.7994
	LSC	0.7371	0.8372	0.7145	0.8290	0.7419	0.4062	0.0911	0.8290	0.7347	0.5658	0.7063
	Overall	**0.7566**	0.6400	0.5551	**0.8803**	0.5381	0.5855	0.4348	0.5358	0.6082	0.7534	**0.7168**
RMSE	GN	7.0679	6.4484	7.9811	6.4906	6.9846	9.7455	9.3476	6.6144	7.4666	6.8695	6.9993
	GB	6.5701	7.0423	7.1210	6.2041	7.9849	7.2244	9.4213	6.8042	7.1231	7.4888	6.7398
	MB	7.6967	8.3654	8.2788	6.4722	8.3620	7.4680	11.245	7.6158	8.8157	7.5610	10.0678
	CC	8.4116	7.1933	7.4160	7.5008	7.9743	7.3480	8.7683	8.5337	8.2239	9.1368	8.2348
	JPEG	6.2295	5.6634	6.9085	5.8124	6.5705	8.5141	8.4540	5.8123	6.7908	7.5417	6.1139
	J2K	6.5691	5.7604	7.1675	5.8539	6.8765	10.347	10.303	6.6655	6.6970	8.3552	6.3013
	LSC	5.8283	4.7749	5.9046	4.9731	5.6846	7.7156	8.5051	4.8647	5.8123	7.0095	6.1773
	Overall	**9.3676**	10.972	11.798	**9.6039**	11.890	11.069	12.059	11.901	11.703	9.3784	**9.8375**
Run Time		0.038	8.395	1.369	0.063	0.233	0.321	205.596	0.644	10.535	0.141	0.849

TABLE II: Comparison of the statistical significance of our algorithm and 10 IQA models on the SIQAD database.

SIQAD	SSIM [9]	IGM [28]	GSIM [12]	GMSD [13]	VSI [45]	OSVP [15]	FEDM [27]	RRED [17]	RRVIF [29]	RQMSH [22]
S	0	+1	+1	0	+1	+1	+1	+1	+1	0

based autoregressive modeling for separating the SC image into predicted and disorder residual part.

We have also examined the proposed algorithm with the recently proposed RQMSH algorithm [22] and observed that the proposed algorithm performs slightly inferior to RQMSH. The proposed algorithm achieves 0.7264 value of PLCC, while RQMSH algorithm achieves 0.7555 value of PLCC. At the same time, both the algorithms are statistically similar, as shown in Table II. From Table I, we can observe that the proposed algorithm performs better than the RQMSH, except for the Gaussian noise and motion blur. As authors of [22] have specifically incorporated the Gaussian and motion blur kernels to judge the quality of Gaussian noise and motion blurred images. By this motivation, we might incorporate the similar kernels with the proposed algorithm towards a large performance gain.

C. Statistical Significance

We assume that the prediction errors (between predicted scores and subjective scores) of IQA algorithms follow a Gaussian distribution and with this view, we use F-Test to check the statistical significance between the proposed algorithm and other IQA algorithms. Assuming a significance level of 0.05, a value S = +1 or S = -1 suggests that the proposed algorithm is statistically better or worse than the other IQA algorithm, respectively. While S = 0, shows that the proposed algorithm is statistically comparable to other IQA algorithm. The results of the statistical significance comparison are reported in Table II. From Table II, one can observe that the proposed algorithm is statistically better than the most of the other IQA algorithms and comparable to the SSIM and GMSD metrics. Although, proposed reduced-reference IQA algorithm achieves slightly worse PLCC as compared to the two full-reference IQA algorithms, namely, SSIM and GMSD but proposed the algorithm and these algorithms are statistically comparable.

At the same time, an efficient RR IQA metric should use a lesser amount of data of reference image to accurately estimate the quality of the distorted image. The proposed algorithm, OSVP, FEDM, RRED, and RRVIF require 2, 9, 2, 9 and 2 reference data, respectively. So the proposed algorithm requires a similar amount of reference data as compared to existing RR algorithms but attains much better performance.

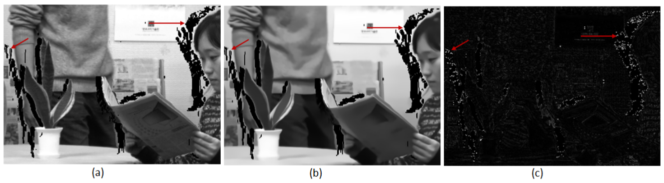

Fig. 4: The 3D synthesized image (a) with strong geometric distortions from IRCCyN/IVC database [7] and corresponding separated predicted (b) and disorder residual part (c) using the free-energy-principle-based RRVIF [29] IQA algorithm.

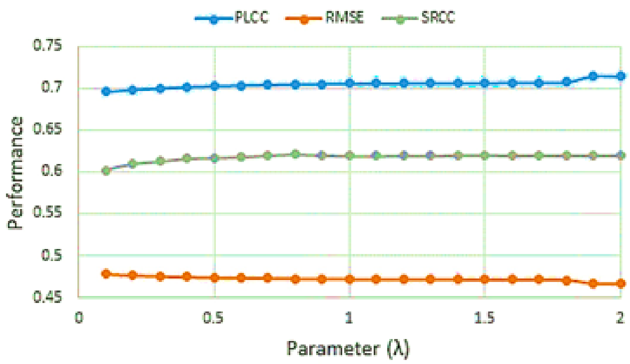

Fig. 5: The relationship between the proposed algorithm and the parameter λ to judge the quality of 3D synthesized images.

IV. Application to Quality Assessment of 3D-Synthesized Images

Similar, to the SC images, free-energy-principle-based IQA algorithms, separate most of the information of the regions with geometric distortions of 3D synthesized images into disorder residual part, as illustrated in Fig. 4. Consequently, assume geometric distortions do not contribute significantly to the overall perceptual quality. Contrary to this, geometric distortions are the prevailing distortions in the 3D synthesized images [7] and it significantly degrades the overall perceptual quality.

Our proposed quality metric has the ability to capture the geometric distortions, and thus we further applied it to the recently constructed IRCCyN/IVC database [7] for examining the quality of 3D-synthesized images. The IRCCyN/IVC [7] database has 96 number of images. Among these 96 images, 12 are reference images and other 84 images are the synthesized views. The geometric distortions are predominantly present in these 84 DIBR-synthesized views.

Similar to the SC images, the proposed algorithm is compared with the 6 FR and 4 RR IQA algorithms specifically designed for natural images and 4 FR algorithms, specifically designed for 3D synthesized images (FR_{DIBR}); these four FR_{DIBR} algorithms are the VSQA [47], 3D-SWIM [48], MW-PSNR [25] and MP-PSNR [26]. The comparison study of the proposed algorithm with 14 IQA algorithms are shown in Table III and the proposed algorithm achieves 0.7145 and 0.4659 value of PLCC and RMSE, respectively. In Table III, proposed1 and proposed2 represent the performance of the proposed algorithm for 3D synthesized images with similar parameters (λ, α, β, and γ) as used for judging the SC images and optimized parameters, respectively. From Table III, one can observe that

TABLE III: Simulation results comparison of the proposed algorithm with existing IQA algorithms for IRCCyN/IVC [7] database. Here $FR_{natural}$ and FR_{DIBR} represents the FR algorithms designed for the natural and the 3D synthesized images, respectively. While $RR_{Natural}$ suggests the RR algorithms designed for the natural images. In each category best performances are denoted with the bold-faces.

	Metric	PLCC	RMSE
$FR_{Natural}$	PSNR	0.3976	0.6109
	SSIM [9]	0.4850	0.5823
	IGM [28]	0.4325	0.6003
	GSIM [12]	0.5246	0.5668
	GMSD [13]	0.4077	0.6080
	VSI [45]	**0.6667**	**0.4963**
$RR_{Natural}$	OSVP [15]	0.4767	0.5853
	FEDM [27]	0.2252	0.6487
	RRED [17]	0.4072	0.6081
	RRVIF [29]	**0.5953**	**0.5351**
FR_{DIBR}	VSQA [47]	0.5742	0.5451
	3D-SWIM [48]	**0.6584**	**0.5011**
	MW-PSNR [25]	0.5622	0.5506
	MP-PNSR [26]	0.6164	0.5238
	Proposed1	**0.7145**	**0.4659**
	Proposed2	**0.7523**	**0.4386**

1) the existing free-energy-principle-based IQA algorithms (IGM [28], FEDM [27], and RRVIF [29]) ignores the impact of regions with geometric distortions and can not catch geometric distortions and subsequently, achieve a much lower value of PLCC as compared to the proposed algorithm.

2) the existing FR and RR algorithms designed for natural images perform poorly for 3D synthesized images due to the different characteristics of these natural images as compared to the 3D synthesized images. In natural image category, VSI [45] algorithm achieves the highest value of PLCC, which is much lower than the proposed algorithm.

3) two state-of-the-art and widely-used IQA algorithms; namely SSIM [9] and GMSD [13] perform poorly on 3D synthesized images, with PLCC value less than 0.5.

4) We have also applied recently developed RQMSH [22] algorithm to judge the quality of 3D synthesized images and it can only able to achieve 0.6216 value of PLCC, which is much lower than the PLCC value achieved by the proposed algorithm.

5) although, the proposed algorithm is an RR metric in nature but it is able to achieve considerably higher performance than the FR algorithms specifically designed for 3D synthesized images.

A good quality metric should not be sensitive to any parameters and its performance should not vary significantly with a slight change of the parameters. In Fig. 5, the dependency of the proposed algorithm for judging the quality of 3D synthesized images on threshold λ used in equation (1) is shown. From this figure, we can see that the proposed algorithm does not significantly depend on the parameter λ.

These results further demonstrated that the proposed quality metric is not only capable of capturing the structural distortions but also can capture geometric distortions. Of this ability, our quality metric also can be used to measure other geometry-type of distortions. The proposed algorithm can be seen as a further step towards forming a universal IQA algorithm, as it can be applied to judge the quality of images with different characteristics.

V. Conclusion

In this paper, we have addressed the issues associated with the free-energy-principle based IQA algorithms for objectively assessing the quality of SC images. These IQA algorithms treat textual regions in such a way that these regions do not play an important role in evaluating the quality of an SC image. This hypothesis is found to be untrue, which makes the performance of these algorithms poor for evaluating the quality of an SC image. With this in view, we have proposed a new reduced-reference IQA algorithm for SC images based upon more perceptual relevant prediction model, which considers the importance of textual regions during quality evaluation. From experiments, it is validated that the proposed algorithm has a better ability to efficiently quantify the quality of SC images as compared to other recently developed reduced-reference IQA algorithms. Furthermore, our proposed quality metric is also very effective in assessing the quality of 3D synthesized images, because of its ability to capture geometry-type of distortions, and this implies the underlying utility of our proposed metric.

Furthermore, it is worthy to mention that, in reality, during the transmission, these screen content images might be contaminated with several distortions (for example, JPEG compression noise) and before transmission (in encoder side). That is, we cannot clear what kind of distortion will be associated with the screen content images and what will be the level of distortion. So in encoder side, we have the information of reference image and no information about the distorted image. In this scenario, we can send very little information (32 bits) about the reference image, and in decoder side, we can decide the quality with the help of this information and distorted image.

References

[1] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Processing,* vol. 24, no. 11, pp. 4408-4421, Nov. 2015.

[2] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud mobile convergence," *IEEE Multimedia,* vol. 18, no. 2, pp. 4-11, Feb. 2011.

[3] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," *IEEE Trans. Image Processing,* vol. 14, no. 8, pp. 993-1005, Aug. 2005.

[4] C. Lan, G. Shi, and F. Wu, "Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation," *IEEE Trans. Image Processing,* vol. 19, no. 4, pp. 946-957, Apr. 2010.

[5] Z. Pan, H. Shen, Y. Lu, S. Li, and N. Yu, "A low-complexity screen compression scheme for interactive screen sharing," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 23, no. 6, pp. 949-960, Jun. 2013.

[6] "Requirements for an extension of HEVC for coding of screen content", document ISO/IEC JTC1/SC29/WG11 MPEG2013/N14174, 2014.

[7] E. Bosc, R. Pépion, P. L. Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J. Select. Top. Signal Process.*, vol. 5, no. 7, pp. 1332-1343, Sep. 2011.

[8] K. Gu et al., "Model-based referenceless quality metric of 3D synthesized images using local image descriptors," *IEEE Trans. Image Processing*, vol. PP, no. 99, pp. 1-1.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing,* vol. 13, no. 4, pp. 600-612, April 2004.

[10] 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, Sept. 2007.

[11] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, Mar. 2010.

[12] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Processing,* vol. 21, no. 4, pp. 1500-1512, April 2012.

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

[14] K. Gu, G. Zhai, X. Yang and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS), 2013*, pp. 1095-1098.

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

[16] Q. Feng, D. Zhao and W. Gao "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Trans. Multimedia,*vol. 17, no. 12, pp. 2338-2344, Dec. 2015.

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

[18] 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.

[19] 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.

[20] 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.

[21] S. Wang, K. Gu, K. Zeng, Z. Wang and W. Lin, "Perceptual screen content image quality assessment and compression," in *Proc. IEEE International Conference on Image Processing (ICIP), 2015* pp. 1434-1438.

[22] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, 'Reduced-reference quality assessment of screen content images," *IEEE Trans. Circuits and Systems for Video Technology,* vol. PP, no. 99, pp. 1-1.

[23] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1-13, Jun. 2016.

[24] 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.

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

[26] D. S. Stankovic, D. Kukolj, and P. L. 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.

[27] G. Zhai, X. Wu, X. Yang, W. Lin and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Processing,* vol. 21, no. 1, pp. 41-52, Jan. 2012.

[28] J. Wu, W. Lin, G. Shi and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Processing,* vol. 22, no. 1, pp. 43-54, Jan. 2013.

[29] J. Wu, W. Lin, G. Shi and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia,* vol. 15, no. 7, pp. 1700-1705, Nov. 2013.

[30] K. Gu, G. Zhai, X. Yang and W. Zhang, “No-reference image sharpness assessment in autoregressive parameter space,” *IEEE Trans. Image Processing,* vol. 24, no. 10, pp. 3218-3231, Oct. 2015.

[31] 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.

[32] J. Qian et al., “Towards Efficient Blind Quality Evaluation of Screen Content Images Based on Edge-Preserving Filter,” *IET Electronics Letters*, vol. 53, no. 9, pp. 592-594, April. 2017.

[33] H. Chen, A. Saxena and F. Fernandes, “Nearest-neighbor intra prediction for screen content video coding,” in *Proc. IEEE International Conference on Image Processing (ICIP), 2014* pp. 3151-3155.

[34] V. Jakhetiya, O. C. Au, S. Jaiswal, L. Jia, and H. Zhang, “Fast and efficient intra-frame deinterlacing using observation model based bilateral filter,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014* pp. 5819-5823.

[35] A. Shahab, F. Shafait, A. Dengel and S. Uchida, “How salient is scene text?,” in *Proc. 10th IAPR International Workshop on Document Analysis Systems (DAS), 2012* pp. 317-321.

[36] Y. Li, W. Jia, C. Shen and A. van den Hengel, “Characterness: an indicator of text in the wild,” *IEEE Trans. Image Processing,* vol. 23, no. 4, pp. 1666-1677, April 2014.

[37] N. Ponomarenko and K. Egiazarian, Tampere Image Database 2008 TID2008, 2008. [Online]. Available: http://www.ponomarenko.info/tid2008.htm

[38] V. Jakhetiya, W. Lin, S. P. Jaiswal, S. C. Guntuku and O. C. Au, “Maximum a posterior and perceptually notivated reconstruction algorithm: A generic framework,” *IEEE Trans. Multimedia,* vol. 19, no. 1, pp. 93-106, Jan. 2017.

[39] V. Jakhetiya, W. Lin, S. Jaiswal, K. Gu, and S. C. Guntuku, “Just Noticeable Difference for Natural Images using RMS Contrast and Feedback Mechanism,” *Neurocomputing,* to appear.

[40] S. Xiao, S. H. Ma, G. Li and S. K. Mukhopadhyay, ”European option pricing with a fast fourier transform algorithm for big data analysis,” *IEEE Trans. Industrial Informatics,* vol. 12, no. 3, pp. 1219-1231, June 2016.

[41] X. Ding, Y. Tian and Y. Yu, ”A real-time big data gathering algorithm based on indoor wireless sensor networks for risk analysis of industrial operations,” *IEEE Trans. Industrial Informatics,* vol. 12, no. 3, pp. 1232-1242, June 2016.

[42] G. A. Susto, A. Schirru, S. Pampuri and S. McLoone, ”Supervised aggregative feature extraction for big data time series regression,” *IEEE Trans. Industrial Informatics,* vol. 12, no. 3, pp. 1243-1252, June 2016.

[43] W. Bi, M. Cai, M. Liu and G. Li, ”A big data clustering algorithm for mitigating the risk of customer churn,” *IEEE Trans. Industrial Informatics,* vol. 12, no. 3, pp. 1270-1281, June 2016.

[44] K. Lin; J. Luo; L. Hu; M. S. Hossain; A. Ghoneim, ”Localization based on social big data analysis in the vehicular networks,” *IEEE Trans. Industrial Informatics,* vol.PP, no.99, pp.1-1

[45] L. Zhang, Y. Shen, and H. Li, “VSI: A visual saliency-induced index for perceptual image quality assessment,” *IEEE Trans. on Image Processing,* vol. 23, no. 10, pp. 4270-4281, Oct. 2014.

[46] “Final report from the video quality experts group on the validation of objective models of video quality assessment ii”, 2003, video Quality Expert Group (VQEG). [Online]. Available: http://www.vqeg.org/,”

[47] 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.

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

Vinit Jakhetiya received the B.Tech. degree in computer and communication engineering from LNM Institute of Information Technology, India, in 2011, and the Ph.D. degree in electronics and computer engineering from Hong Kong University of Science and Technology, Hong Kong, in 2016. From January 2015 to December 2015, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, and later he joined as a Project Officer at the same university. Currently, he is working as an assistant professor, Bennett University, India. His research interests include image/video processing, image quality assessment, and visual perceptual modeling.

Ke Gu received the B.S. and PhD degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is currently the associated editor for IEEE Access, and is the reviewer for 20 SCI journals. His research interests include quality perception and image processing. Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and received the excellent Ph.D. thesis award from the Chinese Institute of Electronics (CIE) in 2016. He is the leading special session organizer in VCIP2016 and ICIP2017.

Weisi Lin (M’92-SM’98-F’16) received his Ph.D. from King’s College, London University, U.K. He served as the Lab Head of Visual Processing, Institute for Infocomm Research, Singapore. Currently, he is an Associate Professor in the School of Computer Science and Engineering. His areas of expertise include image processing, video compression, and multimedia communication, in which he has published 160 journal papers and 230+ conference papers, filed 7 patents, and authored 2 books. He is an AE for IEEE Trans. on Image Processing, and IEEE Trans. Circuits and Systems for Video Tech. He is an IEEE fellow.

Qiaohong Li received the B.E. degree and M.E. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include image quality assessment, speech quality assessment, computer vision, and visual perceptual modeling.

Sunil P. Jaiswal received the B. Tech degree from the LNM Institute of Information Technology (LNMIIT), Jaipur, India, in 2012. He is currently pursuing the Ph.D. degree in electronic and computer engineering in the Hong Kong University of Science and Technology (HKUST), Hong Kong since 2012. He also used to be a visiting scholar in the visual analytics and imaging lab in Stony Brook University, NY USA, and SUNY Korea in 2015/2016. His research interest includes multispectral imaging, 3-D medical imaging and Image quality assessment.