

Saliency-Guided Quality Assessment of Screen Content Images

Ke Gu, Shiqi Wang, Huan Yang, Weisi Lin, *Fellow, IEEE*, Guangtao Zhai, *Member, IEEE*, Xiaokang Yang, *Senior Member, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

Abstract—With the widespread adoption of multi-device communication, such as telecommuting, screen content images (SCIs) have become more closely and frequently related to our daily lives. For SCIs, the tasks of accurate visual quality assessment, high-efficiency compression, and suitable contrast enhancement have thus currently attracted increased attention. In particular, the quality evaluation of SCIs is important due to its good ability for instruction and optimization in various processing systems. Hence, in this paper, we develop a new objective metric for research on perceptual quality assessment of distorted SCIs. Compared to the classical MSE, our method, which mainly relies on simple convolution operators, first highlights the degradations in structures caused by different types of distortions and then detects salient areas where the distortions usually attract more attention. A comparison of our algorithm with the most popular and state-of-the-art quality measures is performed on two new screen content image databases (SIQAD and SCD). Extensive results are provided to verify the superiority and efficiency of the proposed IQA technique.

Index Terms—Screen content images (SCIs), image quality assessment (IQA), visual saliency

I. INTRODUCTION

RECENT advances in cloud and mobile computing have drawn wide interests and new challenges to perceptual codings, quality assessment and computer graphics communities. In many scenarios, e.g., cloud-mobile applications [1], [2], [3], remote computing platforms [4], and cloud gaming [5], remote computing is facilitated based on the users' interaction with the local display interface, which typically includes computer-generated screen content images (SCIs). These images are generally created and compressed at the server followed by transmission to the thin client side. This procedure inevitably introduces artifacts, leading to the quality degradation of SCIs and thereby the deterioration of the interactivity performance and the users' experience. Therefore, accurately predicting the screen quality with an objective algorithm plays a variety of roles in cloud and remote computing applications. First, it can be used to dynamically monitor the

This work was supported in part by the Singapore MoE Tier 1 Project M4011379 and RG141/14.

K. Gu is with School of Computer Engineering, Nanyang Technological University, Singapore, 639798, and with Shanghai Jiao Tong University, Shanghai, 200240, China (email: guke.doctor@gmail.com).

H. Yang and W. Lin are with School of Computer Engineering, Nanyang Technological University, Singapore, 639798.

S. Wang is with School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China.

G. Zhai, X. Yang and W. Zhang are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, 200240, China.



Fig. 1: The first row provides screen content images. The second to fourth rows respectively show fixation-inspired saliency maps, saccade-inspired saliency maps, and combined saliency maps.

quality of SCIs and adjust resources to enhance the remote computing experience. Second, it can be used for optimization, e.g., screen content coding [6], towards better rate distortion performance. Third, it can work as a benchmark in the quality evaluation of a remote computing system.

In multimedia signal processing, image quality assessment (IQA) has long enjoyed popularity for decades, especially the last ten years. Wang *et al.* proposed the structural similarity index (SSIM) [7], considering that the human visual system (HVS) is very sensitive to structural information. Since then, many variants have been developed with various tactics for polishing SSIM, e.g., [8], [9], [10], [11], [12]. Other designs were explored using information theory and statistics [13], the brain principle [14] and low-level vision [15], [16], [17], [18], [19]. Interested readers can refer to [20] for a thorough survey. Notwithstanding the prosperity of IQA methods, the abovementioned ones perform well in the quality estimation

of natural scene images (NSIs), but fail in that of SCIs. In this paper, we therefore endeavor to present a new high-accuracy objective IQA technique for SCIs.

In most situations, SCIs are mixed with NSIs and computer-created graphical content, which usually contains both texts and graphics, as shown in Fig. 1. There are many distinctions between SCIs and NSIs. Capturing natural video will induce noise because of the limitations of image sensors. The SCIs are, however, fully noise free and directly created by screen-recording tools. The camera-captured NSIs mostly feature thick lines, rich color variations and complex texture content, whereas the computer-created SCIs are full of thin lines, limited colors and uncomplicated shapes [21], giving rise to a particular challenge in the study of IQA.

One direct strategy to solve the problem is to distinguish various kinds of regions, such as textual and pictorial ones, before systematically integrating the quality score of each part to infer the overall quality score [22], [23]. This approach, however, is considerably time-consuming and sort of ad-hoc. Another strategy that works is to integrate visual saliency since it has been proved to be effective in the quality evaluation of NSIs from theoretical, experimental and practical perspectives [24], [25], [26], [10], [15], [18], [27]. A majority of saliency detection models have been developed during the past 25 years, consisting of top-down task-dependent methods and bottom-up stimulus-driven methods [28]. Most of the existing techniques belonging to the bottom-up methods are devoted to seeking locations with the maximum local saliency based on biologically motivated local features [29], [30], [31], [32], [33]. The features are inspired by neural responses in the lateral geniculate nucleus and V1 cortex, usually including the intensity, contrast, edge, texture, color and orientation [34]. Some other relevant techniques based on global features attempt to find regions from a visual signal that implies unique frequencies in transform domains [35], [36], [37], [38]. This type of model can quickly and precisely detect visual “pop-outs” and locate possible salient objects.

However, based on tests, most saliency detection models do not work well for the IQA of SCIs. Instead, we concentrate on the recognition and clarity of the salient areas, i.e., texts and their surroundings. These areas are detected through a simple detector, motivated by the behaviors of “fixation” and “saccade” [39]. Furthermore, as given in Fig. 1, we illustrate three images from the new screen image quality assessment database (SIQAD) [22], [23]. We can see that SCIs typically contain more thin structures than NSIs. Based on this concern, the gradient operator for highlighting structural variations is leveraged to quantify the visual degradation in the corrupted image with respect to its reference image. Our Saliency-guided Quality Measure of SCIs (SQMS) is finally developed by combining the structural variations with visual saliency.

The remainder of this paper is organized as follows. Section II first presents the SQMS model explicitly. In Section III, a comparison of our SQMS with mainstream and state-of-the-art IQA methods is conducted on the new SIQAD [22], [23] and screen content database (SCD) [40], and some analyses and description of future works are provided. We finally conclude this paper in Section IV.

II. METHODOLOGY

A. Saliency Detection

It was reported in [10], [15], [18], [27] that visual saliency is able to promote the performance of IQA metrics in real applications. However, it was observed using tests that most existing saliency detection technologies do not achieve satisfactory performance in screen content IQA tasks, possibly due to the distinct perceptual mechanisms of NSIs and SCIs. In particular, humans understand the semantic information in a NSI through a comparison of location, orientation, intensity, and color between the foreground and background elements or different objects. By comparison, for a SCI, the semantic information is mainly contained in the textual parts.

A straightforward yet valid method involves simulating the behavior of human eye movements, i.e., “fixation” and “saccade”, by which humans identify information to predict visual saliency. However, this method is not easy because of the existence of many complicated behaviors, e.g., “saccadic suppression” [41], “crowding effect” [42], etc., and their interactions. Our attention is mainly focused on the subsequent two aspects: the first is that we do not want to produce visual saliency as accurate as eye-tracking experiments but search for possible salient points. The second is that an efficient saliency detection technique is one of our primary pursuits. On this basis, we develop a simple saliency detector for quickly detecting salient regions in SCIs.

It has been stated in [39] that “*We move our eyes about three times each second via rapid eye movements (saccades) to reorient the fovea through the scene. Pattern information is only acquired during periods of relative gaze stability (fixations) owing to ‘saccadic suppression’ during the saccades themselves*”. In eye-tracking experiments, some neighboring gaze points are typically clustered to be a fixation point that is remarkably different from its surroundings. Then, all the fixations can yield the final visual saliency map. Motivated by this process, visual saliency can be approximately understood as the points that have larger local contrast compared to their vicinities in a proper window. In reality, this supposition has been broadly employed in numerous classical and recently devised saliency detection models [29], [30], [31], [32], [33], resulting in a fairly good performance. Inspired by this, we attempt to measure visual saliency by comparing the local similarity between a visual signal \mathbf{r} and its smoothed version by Gaussian low-pass filtering:

$$\mathbb{M}_f(\mathbf{r}) = f_s(\mathbf{r}, \mathbf{r}_g) \quad (1)$$

where $f_s(\cdot)$ computes the similarity of two elements in the bracket¹ and \mathbf{r}_g is created by applying a “Gaussian” convolution operator on \mathbf{r} . We apply a $n \times n$ circular-symmetric Gaussian weighting function in this implementation:

$$\mathbf{r}_g = \mathbf{r} \otimes \mathbf{g} = \frac{1}{n^2} \sum_{i=1}^{n^2} g_i r_i \quad (2)$$

¹The gradient magnitude similarity, as defined in (6)-(9) later, is employed for this on account of its efficiency and highlights of important structural information in HVS perception.



Fig. 2: Illustration of the gradient magnitude (GM). From the left to right, the top row includes one clear screen content image and six distorted versions corrupted by Gaussian noise, Gaussian blur, motion blur, contrast change, JPEG compression and JPEG2000 compression. The bottom row shows the associated GM maps.

where “ \otimes ” indicates the convolution operation, i is the pixel index, and $\mathbf{g} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right)$ with the standard deviation of σ being normalized to the unit sum. The values in \mathbf{r}_g lie within the range $[0, 1]$. The parameter values are empirically set to be $n = 11$ and $\sigma = 5.5$. In Fig. 3(a), we illustrate the “Gaussian” convolution kernel. More analyses regarding the sensitivity of the parameter values will be provided in the next section.

Notice that compared to nonsalient pixels, salient pixels of larger local contrast are usually degraded more by Gaussian filtering; thus, they have lower similarity between \mathbf{r} and \mathbf{r}_g and have lower values than their neighbors in \mathbb{M}_f . Taking the abovementioned three sample SCIs in Fig. 1 as examples, we present the associated fixation-inspired saliency maps in the second row. It is not hard to see that the required textual parts and their vicinities are detected successfully. However, it is unlucky that the unexpected frame lines are highlighted, as well.

Another behavior found in human eye movement is “saccade”, which causes one kind of “motion blur” effect that plays an especially significant role in perceiving SCIs [43]. This effect indicates that visual saliency is not only influenced by the pixels in a local window of the proper size but also affected by the comparatively distant pixels along the saccadic direction, e.g., in the textual areas. Specifically, we approximate the saccade-inspired visual saliency using a simple method and define it to be the local similarity of \mathbf{r} and its motion-blurred version:

$$\mathbb{M}_s(\mathbf{r}) = f_s(\mathbf{r}, \mathbf{r}_m) \quad (3)$$

where \mathbf{r}_m is produced by convolving \mathbf{r} with a “motion blur” kernel that is defined as

$$\mathbf{m} = \begin{cases} \frac{1}{t} & \text{if } (x_1, x_2)(\sin \theta, \cos \theta)^T = 0, x_1^2 + x_2^2 \leq \frac{t^2}{4} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where t is the amount of motion in pixels and θ is the motion direction with its angle along the horizontal x axis. The parameter values are empirically set to be $t = 9$ and $\theta = 1$. This “motion blur” convolution kernel is visualized in Fig. 3(b). The parameter sensitivity will be discussed in the next section. Furthermore, we present the associated saccade-inspired saliency maps in the third row in Fig. 1 for easy comparison. We found that in contrast to isolated letters in

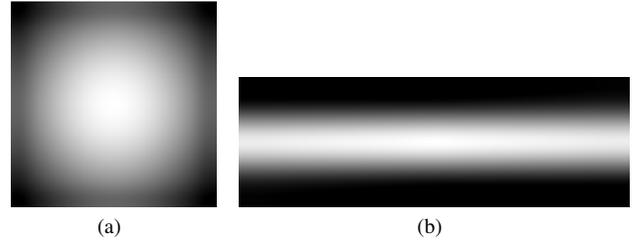


Fig. 3: Convolution kernels: (a) Gaussian blur, (b) motion blur.

fixation-inspired saliency maps, the neighboring letters are considered to simulate the “motion blur” effect of human eye movements. Additionally, we observed that the highlighted undesired frame lines in fixation-inspired saliency maps are decreased, to some extent.

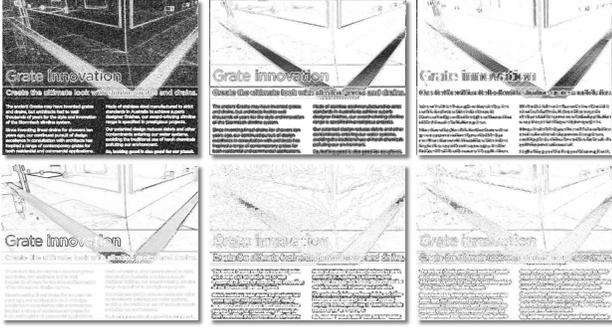
Incorporating the merits of fixation- and saccade-inspired saliency maps for generating a suitable saliency map is still a difficult problem. A simple solution is to use a linear weighted additive model as follows:

$$\mathbb{M}(\mathbf{r}) = \frac{\mathbb{M}_f + \lambda \mathbb{M}_s}{1 + \lambda} \quad (5)$$

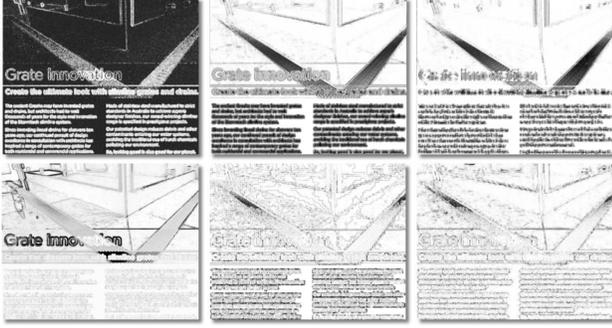
where λ is a fixed positive number used to manipulate the relative importance between two components. As can be seen in the fourth row in Fig. 1, the resulting combined saliency maps with $\lambda = 1$ are able to capture the important texts and the surrounding areas as well as the main contours in natural scene regions. We further investigate the performance based on some classical nonlinear models, e.g., [10], [15], to combine fixation- and saccade-inspired saliency maps. The results tell us that these nonlinear models introduce more computations but do not lead to noticeable performance gain in the overall image database and each type of distortion subset.

B. Quality Evaluation

The HVS is strongly sensitive to structural variations, especially in quality prediction. In this paper, the gradient magnitude (GM) is used to gauge the structural degradation because, on the one hand, it has been widely applied in many image processing and computer vision applications, such as contour detection, optical flow and segmentation, and, on the other hand, it has also been recently found to perform effectively



(a)



(b)

Fig. 4: Comparison of (a) the gradient-based \mathbb{G} maps and (b) the structure-inspired SSIM maps.



Fig. 5: Illustration of the finally generated SQMS_MAP maps.

in the IQA measures [15], [16], [17], [18]. Specifically, the image gradient magnitude is extracted via the Scharr operator [44], which is expressed by applying convolution masks to an input image signal \mathbf{s} :

$$\mathbf{p} = \sqrt{\mathbf{p}_x^2 + \mathbf{p}_y^2} \quad (6)$$

where

$$\mathbf{p}_x = \mathbf{h}_x \otimes \mathbf{s} = \frac{1}{16} \begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix} \otimes \mathbf{s} \quad (7)$$

$$\mathbf{p}_y = \mathbf{h}_y \otimes \mathbf{s} = \frac{1}{16} \begin{bmatrix} +3 & +10 & +3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \otimes \mathbf{s} \quad (8)$$

where \mathbf{p}_x and \mathbf{p}_y represent the Scharr convolution masks along the x- and y-axes, respectively. We chose one representative

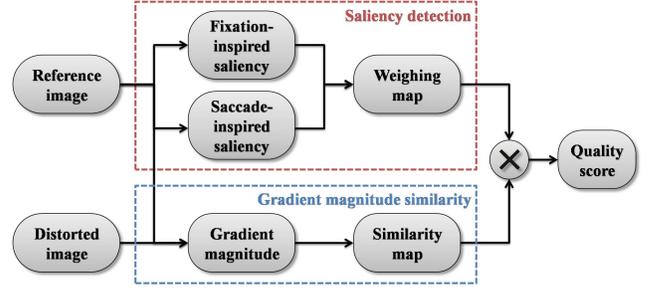


Fig. 6: The flowchart of the proposed image quality metric.

SCI and its six distorted images from the SIQAD database [22], [23] as well as their corresponding GM maps, as given in Fig. 2. For the clear original SCI, the GM highlights texts and the structural information. By comparison, the GM maps of distorted SCIs can reflect the visual degradation caused by different distortion types. For example, the Gaussian blur largely reduces the clarity of texts.

We then make use of the common similarity metric, which provides the three advantages of symmetry, boundedness and unique maximum [7], to measure the difference in GM maps between the reference image \mathbf{r} and the distorted one \mathbf{d} :

$$\mathbb{G}(\mathbf{r}, \mathbf{d}) = \frac{2\mathbf{p}_r \cdot \mathbf{p}_d + c}{\mathbf{p}_r^2 + \mathbf{p}_d^2 + c} \quad (9)$$

where \mathbf{p}_r and \mathbf{p}_d are GM maps of the original and distorted images, respectively, and c is a positive constant for avoiding instability. In Fig. 4(a), we present the six \mathbb{G} maps corresponding to the distorted images in Fig. 2, where a brighter gray level denotes higher similarity and thus a lower distortion level. As expected, the \mathbb{G} maps succeed in capturing the remarkable difference between the original and distorted images.

In addition, we stress that the structure-inspired SSIM [7] is also a good candidate choice. We further provide the SSIM maps in Fig. 4(b) for comparison. As seen, in contrast to \mathbb{G} maps, SSIM maps, due to the use of large-size windows for low-pass filtering, exert a high degree of denoising effects on the noisy image and do not accurately characterize the degradations in the texts of thin lines. Moreover, from the viewpoint of computational cost, the GM operator with four convolution operations in a 3×3 window is also superior to the SSIM metric with five convolution operations in the 11×11 window. Under the above considerations, the GM operator is preferred for the quality evaluation.

Our SQMS metric is developed using the saliency map \mathbb{M} to weight the GM similarity map \mathbb{G} . It should be noted that higher-value elements in \mathbb{M} are more similar to the neighbors and thereby have less visual saliency. Hence, we reverse the visual saliency to derive the weighting map to be:

$$\mathbb{W}(\mathbf{r}) = 1 - \mathbb{M}(\mathbf{r}) \quad (10)$$

with $\lambda = 1$. More sensitivity tests on λ will be provided in the next section. Notice that the whole local distortion map (i.e., gradient-based \mathbb{G} map) is used to predict the quality score, and the salient pixels are highlighted. The SQMS_MAP map

can be obtained by

$$\text{SQMS_MAP}(r_i, d_i) = \frac{\mathbb{G}(r_i, d_i) \cdot \mathbb{W}(r_i)}{\frac{1}{k} \sum_{i=1}^k \mathbb{W}(r_i)} \quad (11)$$

where k denotes the total number of pixels in the image. In Fig. 5, we display the six SQMS_MAP maps for the distorted SCIs in Fig. 2. The SQMS_MAP effectively emphasizes the regions that not only include substantial textual and structural degradations but also attract a lot of visual attention. We finally obtain the SQMS score using the global mean:

$$\text{SQMS}(\mathbf{r}, \mathbf{d}) = \frac{1}{k} \sum_{i=1}^k \text{SQMS_MAP}(r_i, d_i). \quad (12)$$

The flowchart of our proposed IQA model is shown in Fig. 6 for the reader's convenience.

III. EXPERIMENTS AND ANALYSES

A. Screen Content Database

To the best of our knowledge, there are two new image databases dedicated to the screen content IQA, i.e., SIQAD [22], [23] and SCD [40]. The first SIQAD database includes up to 980 distorted SCIs generated by corrupting 20 sources with seven distortion types at seven distortion levels. These 20 reference SCIs are gathered from webpages, slides, PDF files and digital magazines by screen snapshots, as shown in Fig. 7, and are then cropped to the proper sizes so they can be natively displayed on computer screens during the subjective experiment. The seven distortions types, consisting of Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG compression, JPEG2000 compression (JP2K) and Layer Segmentation-based Coding (LSC) [45], are applied to each of the original SCIs. In accordance with the suggestion in ITU-R BT.500-13 [46], this test utilizes one of the most frequently used single stimulus (SS) methods with a 11-point discrete scale (from the worst "0" to the best "10" with an interval of "1") for scoring. All the subjects, who are university undergraduate or graduate students without any experience and knowledge about quality assessment, are asked to provide their visual opinion scores for each testing image at a viewing distance that is about 2-2.5 times the image height. Because the participants are not experts, they will be trained by adequate typical samples with various distortion types and intensities before proceeding to formal subjective testing.

Upon obtaining the opinion scores of each subject for each testing image, the consistency should be examined to guarantee the availability of the subjective quality ratings. In particular, for each image, the number and standard deviation of the scores are used to compute the confidence interval and thus determine the consistency. Under the assumption of a 95% confidence level, the difference between the computed differential mean opinion score (DMOS) and the "real" quality score should be less than the 95% confidence interval limits [46]. As illustrated in Fig. 8, the distribution of confidence intervals with respect to all the DMOS values is centralized on small values, from 0 to 7. Finally, by normalization, the



Fig. 7: The 20 original screen content images in the SIQAD database.

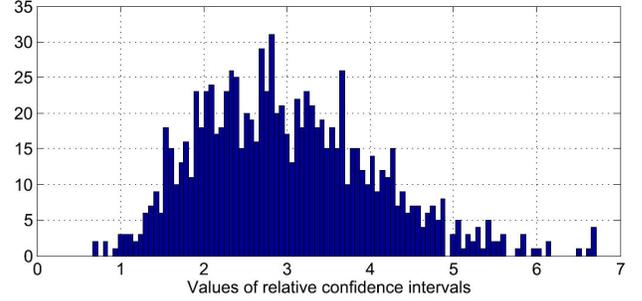


Fig. 8: Distribution of the confidence intervals related to all the DMOS values on SIQAD. Smaller values indicate higher reliability.



Fig. 9: The 24 reference screen content images in the SCD database.

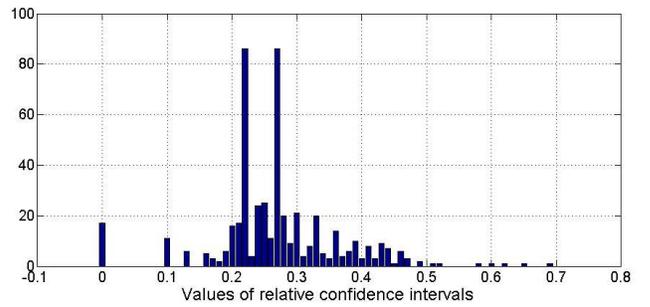


Fig. 10: Distribution of the confidence intervals related to all the DMOS values on SCD. Smaller values denote higher reliability.

DMOS value of each image in the SIQAD database ranges from 24.2 to 90.1.

The second SCD database is composed of a total of 492 compressed SCIs produced using two coding technologies to compress 24 reference SCIs, which have three typical image resolutions of 2560×1440 , 1920×1080 and 1280×720 , as presented in Fig. 9. The two coding technologies are high-efficiency video coding (HEVC), the performance of which is based on a 4:2:0 color format, adaptive quad-tree

TABLE I: Performance comparison of 13 IQA models on SIQAD and SCD databases. We denoted the best-performing model in bold.

<i>SIQAD</i>	NQM [47]	SSIM [7]	MS-SSIM [8]	VIFP [13]	VSNR [48]	IW-SSIM [9]	FSIM _c [15]	GSI [16]	SW-SSIM [11]	SIQM [49]	LTG [17]	VSI [18]	SQMS (Pro.)
\mathcal{X}_P	0.5200	0.7615	0.6195	0.8489	0.5966	0.6536	0.5920	0.5686	0.6117	0.8625	0.7461	0.5568	0.8872
\mathcal{X}_S	0.5321	0.7583	0.6112	0.8451	0.5693	0.6546	0.5817	0.5483	0.6064	0.8582	0.7341	0.5381	0.8803
\mathcal{X}_K	0.3708	0.5682	0.4547	0.6516	0.4374	0.4977	0.4253	0.4054	0.4497	0.6679	0.5580	0.3874	0.6936
\mathcal{X}_A	9.8139	7.1854	8.6781	5.9342	8.8283	8.2759	8.9912	9.1663	8.8459	5.6511	7.1934	9.2875	5.2926
\mathcal{X}_R	12.227	9.2784	11.236	7.5650	11.487	10.833	11.537	11.775	11.324	7.2433	9.5303	11.890	6.6039

<i>SCD</i>	NQM [47]	SSIM [7]	MS-SSIM [8]	VIFP [13]	VSNR [48]	IW-SSIM [9]	FSIM _c [15]	GSI [16]	SW-SSIM [11]	SIQM [49]	LTG [17]	VSI [18]	SQMS (Pro.)
\mathcal{X}_P	0.6837	0.8696	0.8867	0.9028	0.7050	0.8930	0.9019	0.8921	0.8893	0.8920	0.8789	0.8715	0.9059
\mathcal{X}_S	0.6603	0.8683	0.8922	0.9043	0.7172	0.8990	0.9039	0.8947	0.8976	0.8953	0.8804	0.8719	0.9096
\mathcal{X}_K	0.4749	0.6910	0.7185	0.7393	0.5383	0.7260	0.7331	0.7215	0.7257	0.7241	0.7009	0.6941	0.7470
\mathcal{X}_A	1.3350	0.8531	0.7978	0.7082	1.1869	0.7837	0.7354	0.7752	0.7941	0.7862	0.8355	0.8337	0.6949
\mathcal{X}_R	1.6189	1.0953	1.0257	0.9542	1.5733	0.9984	0.9585	1.0025	1.0145	1.0030	1.0583	1.0879	0.9396

TABLE II: Comparison of the statistical significance of our designed SQMS and 12 IQA algorithms on the SIQAD and SCD databases.

<i>SIQAD</i>	NQM	SSIM	MS-SSIM	VIFP	VSNR	IW-SSIM	FSIM _c	GSI	SW-SSIM	SIQM	LTG	VSI
Index	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

<i>SCD</i>	NQM	SSIM	MS-SSIM	VIFP	VSNR	IW-SSIM	FSIM _c	GSI	SW-SSIM	SIQM	LTG	VSI
Index	+1	0	0	0	+1	0	0	0	0	0	0	+1

coding structure, and three new concepts (namely, coding unit, prediction unit and transform unit), and screen content compression (SCC), which is extended from HEVC with a 4:4:4 color format to enhance the coding efficacy on screen content through a series of advanced strategies, e.g., intra block copy and transform skipping. In each coding technology, we consider 11 quality levels from 30 to 50 at an interval of 2.

During the testing process, similar to the SIQAD database, the practical SS subjective testing methodology [46] is also used. For each testing image, 20 inexperienced observers were invited at a viewing distance of 2-2.5 times the screen height to offer their personal quality ratings by choosing one option on a 10-level discrete scale, in which the ten quality scales from the smallest to largest scores were successively labeled to be “Bad (1, 2)”, “Poor (3, 4)”, “Fair (5, 6)”, “Good (7, 8)” and “Excellent (9, 10)”. Before the subjective test began, each inexpert subject was to have undergone a particular training session with several examples compressed by representative quality indices. As shown in Fig. 10, under the assumption of a 95% confidence level, the distribution of confidence intervals with respect to all the MOS values mainly ranges from 0.2 to 0.3. The final MOS value of each image in the SCD database ranges from 1 to 9.9.

B. Testing Metrics and Evaluation Protocols

A comparison is conducted between our SQMS model and twelve IQA metrics. The first group of testing algorithms is composed of five classical quality metrics: 1) Noise quality measure (NQM) [47], built upon Peli’s contrast pyramid; 2) SSIM [7], comparing the original and distorted images in terms of luminance, contrast and structural information; 3) Multi-scale SSIM (MS-SSIM) [8], performing SSIM at each scale followed by integration with distinct weights acquired from a psychophysical test; 4) Pixel-based visual information fidelity (VIFP) [13], defined to be the ratio of the mutual information between the original and distorted images to the

self-information; 5) Visual signal-to-noise ratio (VSNR) [48], estimating the visual fidelity of natural images based on near-threshold and suprathreshold properties of human vision.

The second group contains seven state-of-the-art measures, which have achieved high prediction performance for a broad field of distortion categories. They are the following: 1) Information content weighted SSIM (IW-SSIM) [9], structural similarity weighted SSIM (SW-SSIM) [11] and structure-induced quality metric (SIQM) [49], using IQA frameworks to weight visual quality evaluations; 2) Color-based feature similarity index (FSIM_c) [15] and gradient similarity index (GSI) [16], based on the fact that the HVS perceives an image mainly with low-level features, e.g., gradient and phase; 3) Local-tuned-global model (LTG) [17] and visual saliency-induced index (VSI) [18], systematically pooling global and local degradations to predict the overall quality score.

The calculation of the performance measures for IQA methods requires a regression procedure to reduce the nonlinearity of the predicted scores. We apply logistic regression for the nonlinear regression [50]:

$$f(u) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(u - \beta_3)}} \right) + \beta_4 u + \beta_5 \quad (13)$$

where u and $f(u)$ are, respectively, the original IQA scores and the mapped scores after the regression according to the subjective ratings; β_1 to β_5 are regression model parameters that are determined during the curve fitting process. After the regression, five representative correlation performance indices are computed according to [50]. The first index is the Pearson linear correlation coefficient (\mathcal{X}_P) between the mapped quality scores and the subjective ratings for the prediction accuracy. The second and third indices, for evaluating the prediction monotonicity, are the Spearman rank order correlation coefficient (\mathcal{X}_S) and Kendall’s rank-order correlation coefficient (\mathcal{X}_K) between the objective and subjective quality scores, which are immune to the logistic regression. The fourth and

TABLE III: Performance comparison of the SSIM/GSIM metric and the weighted metrics on the SCD and SIQAD databases.

SIQAD database	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_K	\mathcal{X}_A	\mathcal{X}_R	SCD database	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_K	\mathcal{X}_A	\mathcal{X}_R
GSIM	0.7761	0.7865	0.5844	7.0559	9.0264	GSIM	0.8776	0.8789	0.6979	0.8447	1.0634
Weighted GSIM	0.8872	0.8803	0.6936	5.2926	6.6039	Weighted GSIM	0.9059	0.9096	0.7470	0.6949	0.9396
SSIM	0.7615	0.7583	0.5682	7.1854	9.2784	SSIM	0.8696	0.8683	0.6910	0.8531	1.0953
Weighted SSIM	0.8774	0.8669	0.6799	5.4115	6.8680	Weighted SSIM	0.8931	0.8962	0.7249	0.7832	0.9981

TABLE IV: Performance evaluations of each dataset in the SIQAD database. We highlight the top IQA technique in boldface.

SIQAD database	IW-SSIM [9]		FSIM _c [15]		GSI [16]		SW-SSIM [11]		SIQM [49]		VSI [18]		SQMS (Pro.)	
	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_P	\mathcal{X}_S
GN dataset	0.888	0.874	0.888	0.875	0.851	0.843	0.892	0.877	0.891	0.871	0.884	0.866	0.900	0.886
GB dataset	0.908	0.906	0.823	0.822	0.883	0.880	0.843	0.837	0.924	0.923	0.850	0.850	0.913	0.915
MB dataset	0.842	0.842	0.737	0.731	0.775	0.776	0.734	0.734	0.859	0.858	0.766	0.766	0.867	0.869
CC dataset	0.841	0.756	0.820	0.706	0.817	0.734	0.827	0.754	0.786	0.693	0.774	0.646	0.803	0.695
JPEG dataset	0.800	0.798	0.676	0.678	0.677	0.680	0.715	0.715	0.808	0.811	0.715	0.720	0.786	0.789
JP2K dataset	0.804	0.799	0.713	0.697	0.724	0.712	0.719	0.705	0.821	0.811	0.750	0.730	0.826	0.819
LSC dataset	0.816	0.821	0.704	0.707	0.723	0.717	0.706	0.698	0.795	0.806	0.746	0.742	0.813	0.829

fifth indices are the average absolute error (\mathcal{X}_A) and root mean square error (\mathcal{X}_R) used to measure the prediction consistency. Of the five indices stated above, a value close to 1 for \mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K and near to 0 for \mathcal{X}_A and \mathcal{X}_R denotes superior correlation with human perceptions.

C. Performance Comparison and Statistical Significance

As listed in Table I, we tabulate the performance measures of 13 IQA models on the SIQAD and SCD databases. The best-performing algorithm is highlighted in bold font. The proposed SQMS metric has acquired the highest indices for \mathcal{X}_P , \mathcal{X}_S , and \mathcal{X}_K , and the lowest indices for \mathcal{X}_A and \mathcal{X}_R and thus achieves superior performance compared to the classical and state-of-the-art testing methods. On the SIQAD database, only our IQA method achieves scores greater than 0.88 for \mathcal{X}_P and \mathcal{X}_S and greater than 0.69 for \mathcal{X}_K , but lower than 5.5 for \mathcal{X}_A and lower than 7.0 for \mathcal{X}_R . Relative to the benchmark SSIM metric, the performance of our SQMS method exceeds 16% in terms of performance accuracy and monotonicity. Further, in contrast to the second-ranking SIQM in terms of performance, the performance gain of our SQMS method is greater than 2.8% in terms of performance accuracy and 2.5% in terms of performance monotonicity. For the other SCD database, our metric yields scores as high as 0.9 for \mathcal{X}_P and \mathcal{X}_S , about 0.75 for \mathcal{X}_K , lower than 0.7 for \mathcal{X}_A , and lower than 0.95 for \mathcal{X}_R . Compared with the benchmark SSIM metric, our SQMS yields a higher than 4% gain for performance accuracy and 4.5% gain for performance monotonicity.

Apart from the five frequently used evaluations above, the outlier ratio is also applied to obtain confidence intervals from subjective experiments for the performance comparison of the objective IQA metrics. Specifically, after the regression, the differences between the converted objective scores and subjective ratings are first computed, and then, the outlier ratio is defined as the number of differences outside of the confidence intervals to the number of images in total. A smaller outlier ratio means a better IQA model. We have calculated the outlier ratios for the three best-performing metrics, i.e., SQMS, SIQM and VIFP, and the results demonstrate the superiority of our SQMS metric.

TABLE V: Performance measures on the HEVC and SCC datasets in the SCD database.

SCD database	HEVC dataset			SCC dataset		
	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_K	\mathcal{X}_P	\mathcal{X}_S	\mathcal{X}_K
IW-SSIM [9]	0.899	0.899	0.719	0.885	0.895	0.732
FSIM _c [15]	0.901	0.901	0.722	0.897	0.906	0.744
GSI [16]	0.902	0.903	0.723	0.878	0.884	0.720
SW-SSIM [11]	0.883	0.889	0.708	0.890	0.904	0.747
SIQM [49]	0.901	0.904	0.728	0.877	0.883	0.719
VSI [18]	0.856	0.854	0.660	0.895	0.903	0.744
SQMS (Pro.)	0.915	0.917	0.756	0.892	0.900	0.740

The statistical significance obtained from comparing the prediction residuals of each testing IQA approach after the regression is also determined. We suppose that the prediction residuals of IQA measures follow the Gaussian distribution and thus use the F-test to compute the residuals of our SQMS metric and each IQA model. Assuming a significance level of 0.05, a value of $H = +1$ indicates that our technique is statistically better than the tested model, while a value of $H = -1$ indicates that our technique is statistically worse than the tested technique. A value of $H = 0$ concludes that our technique is statistically comparable with the IQA model tested. The results for the SIQAD and SCD databases of the statistical significance comparison are listed in Table II. For the first SIQAD database, the results are all “+1”, which means that the proposed model is statistically better than the overall IQA metrics considered in this work. For the second SCD database, our SQMS metric is also superior or equivalent to the IQA models tested in terms of statistical significance.

The contribution and generality of the proposed saliency-based weighting model should be verified. First, we measured the contribution of the saliency map on the SIQAD and SCD databases, as tabulated in Table III. The GSIM metric is defined as the global mean of the GM similarity map \mathcal{G} . By comparison, the weighted version based on the saliency map \mathcal{W} always enhances the GSIM metric to a large extent. For the SIQAD database, the performance gain is as high as 14% for \mathcal{X}_P , 11% for \mathcal{X}_S and 18% for \mathcal{X}_K . For the SCD database, the performance gain is also greater than 3.2% for \mathcal{X}_P , 3.4% for \mathcal{X}_S and 7% for \mathcal{X}_K . Second, we replace the GSIM metric with the traditional SSIM metric to verify the

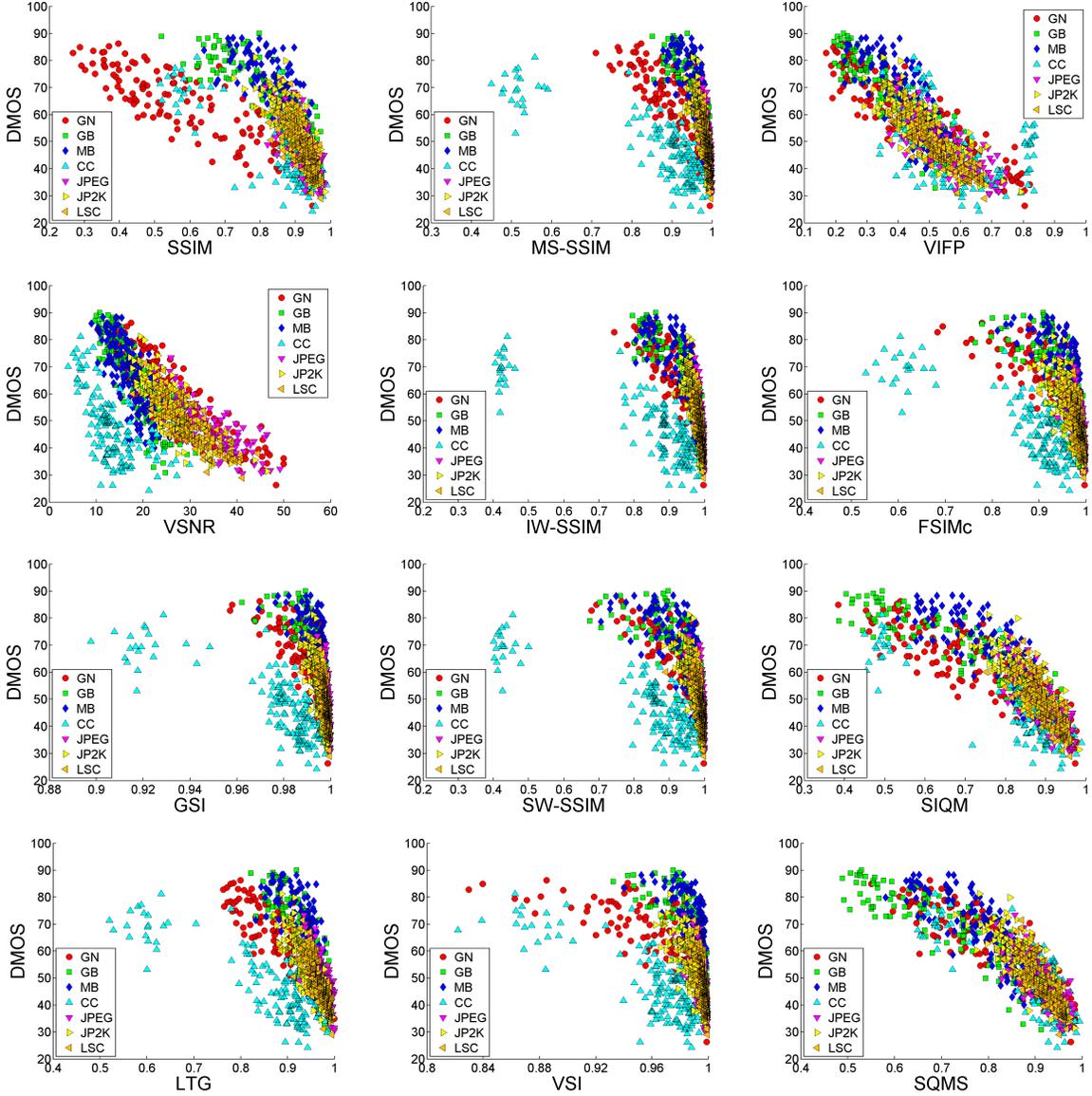


Fig. 11: Scatter plots of DMOS versus SSIM, MS-SSIM, VIFP, VSNR, IW-SSIM, FSIM_c, GSI, SW-SSIM, SIQM, LTG, VSI and our SQMS metrics on the SIQAD database. GN: Gaussian noise (red); GB: Gaussian blur (green); MB: motion blur (blue); CC: contrast change (cyan); JPEG: JPEG compression (magenta); JP2K: JPEG2000 compression (yellow); LSC: Layer Segmentation-based Coding (orange).

generality of the saliency detection model. As can be seen from Table III, the weighted SSIM metric² achieves a high degree of performance improvement over the original SSIM metric, about 15%, 14% and 19% for \mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K on the SIQAD database as well as 2.7%, 3.2% and 4.9% for \mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K on the SCD database, respectively. Furthermore, the statistical significance is also compared between the weighted SSIM/GSIM metric and the associated SSIM/GSIM metric. The results show that the weighted SSIM/GSIM is statistically better than the original version on SIQAD and is statistically superior/comparable to the original version on SCD.

In addition to the overall performance on the entire screen content database, we further conducted a comparison of all the

testing IQA measures on each individual distortion type. For the SIQAD database, two of the most important indices for performance accuracy and monotonicity (\mathcal{X}_P and \mathcal{X}_S) were used. As mentioned above, there are seven distortion types in the SIQAD database. We separately present the results for the seven recent IQA methods in Table IV, and indicate in bold the best-performing method of each type for a straightforward comparison. It can be seen that our SQMS model delivers a fairly good performance for several categories. Our approach clearly outperforms other computing metrics on the GN, MB and JP2K datasets. Among the seven distortion types, our SQMS performs the best 3 times for \mathcal{X}_P and 4 times for \mathcal{X}_S , which is noticeably higher than the second-ranking SIQM that performs the best 2 times both for \mathcal{X}_P and \mathcal{X}_S and the third-ranking IW-SSIM that performs the best 2 times for \mathcal{X}_P and once for \mathcal{X}_S . The results can be made more satisfactory

²The convolution window used in the weighted SSIM metric is assigned to be 5×5 in size for better capturing the variations in the texts of thin lines.

TABLE VI: Computational cost (in seconds/image) of our SQMS and twelve IQA techniques on the overall SIQAD database.

Metrics	NQM	SSIM	MS-SSIM	VIFP	VSNR	IW-SSIM	FSIM _c	GSI	SW-SSIM	SIQM	LTG	VSI	SQMS
Time (s)	0.556	0.084	0.135	0.210	0.370	0.738	0.615	0.032	13.67	0.163	0.053	0.232	0.096

TABLE VII: Performance measures of our SQMS and eight saliency-based metrics on the SIQAD database.

<i>SIQAD</i>	AIM-GM [30]	QTF-GM [31]	NRF-GM [32]	FES-GM [33]	SR-GM [35]	FT-GM [36]	IS-GM [37]	HFT-GM [38]	SQMS (Pro.)
\mathcal{X}_P	0.8073	0.8163	0.7877	0.7784	0.7472	0.7707	0.7647	0.7830	0.8872
\mathcal{X}_S	0.8167	0.8235	0.7962	0.7820	0.7497	0.7768	0.7722	0.7895	0.8803
\mathcal{X}_K	0.6171	0.6235	0.5956	0.5853	0.5502	0.5763	0.5750	0.5912	0.6936

TABLE VIII: The sensitivity test of the four parameters in our SQMS model on the SIQAD database.

n	7	9	11	13	15	σ	1.5	3.5	5.5	7.5	9.5
\mathcal{X}_P	0.8863	0.8877	0.8872	0.8859	0.8841	\mathcal{X}_P	0.8795	0.8872	0.8872	0.8872	0.8871
\mathcal{X}_S	0.8776	0.8802	0.8803	0.8787	0.8763	\mathcal{X}_S	0.8691	0.8800	0.8803	0.8804	0.8804
\mathcal{X}_K	0.6905	0.6934	0.6936	0.6917	0.6890	\mathcal{X}_K	0.6811	0.6935	0.6936	0.6936	0.6936

t	5	7	9	11	13	θ	-5	-2	1	4	7
\mathcal{X}_P	0.8808	0.8857	0.8872	0.8867	0.8861	\mathcal{X}_P	0.8850	0.8864	0.8872	0.8855	0.8845
\mathcal{X}_S	0.8720	0.8778	0.8803	0.8807	0.8809	\mathcal{X}_S	0.8792	0.8800	0.8803	0.8796	0.8791
\mathcal{X}_K	0.6841	0.6908	0.6936	0.6933	0.6933	\mathcal{X}_K	0.6913	0.6930	0.6936	0.6919	0.6910

by modifying the devised weighting scheme in terms of the distortion intensities and categories. We also use three indices for performance accuracy and monotonicity (\mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K) to compare our SQMS and recently developed IQA measures on the HEVC and SCC datasets in the SCD database. The results in Table V illustrate that the proposed SQMS model delivers a substantially high correlation performance.

D. Computational Cost and Visualized Comparison

The computational cost is another important performance index because high-volume visual data that are compressed and transmitted are assessed and monitored at every instance. Table VI lists the average running time for all the 980 images in the SIQAD database. MATLAB R2010a (7.10.0) on a computer equipped with a 3.40 GHz CPU processor and 4.00 GB of RAM was used in this test. It can be seen that our SQMS approach requires less than 0.1 seconds, which is comparable to the benchmark SSIM metric.

The scatter plots of the human ratings versus the objective quality scores of 12 testing IQA techniques on the SIQAD database are shown in Fig. 11 for a visual comparison. The methods used are SSIM, MS-SSIM, VIFP, VSNR, IW-SSIM, FSIM_c, GSI, SW-SSIM, SIQM, LTG, VSI and our devised SQMS metrics. In each scatter plot, we use distinct colors to label the sample points associated with various distortion categories: red for GN, green for GB, blue for MB, cyan for CC, magenta for JPEG, yellow for JP2K, and orange for LSC. A good IQA model has the ability to predict the visual quality consistently across distinct types of distortions. As displayed in Fig. 11, the scatter plot shows that our SQMS is more robust across distortion types and thereby has better performance consistency. Especially for the contrast change data set, the sample points of the proposed SQMS are near the other six distortion types, while the sample points for most testing IQA algorithms are relatively far from the other six types. Therefore, the proposed IQA model can yield a high performance index.

E. Impact of the Different Saliency Detection Models

From another viewpoint, our proposed weighting strategy has a function similar to visual saliency. Eight representative visual saliency detection models, including AIM [30], QTF [31], NRF [32], FES [33], SR [35], FT [36], IS [37] and HFT [38], are thus used to tune the gradient magnitude similarity and propose AIM-GM, QTF-GM, NRF-GM, FES-GM, SR-GM, FT-GM, IS-GM and HFT-GM techniques. Specifically, we first use each of above saliency detection models to predict the saliency map to be used for weighting instead of $\mathbb{W}(\mathbf{r})$ in (11)-(12), resulting in a visual quality estimation. The performance measures of the eight saliency-based IQA models and our SQMS metric are given in Table VII. The first three typical performance indices, i.e., \mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K , are applied here. As illustrated, the proposed method has a remarkably higher performance than the other eight models. We further examine the statistical significance of our method and that of each of the above eight models, and the results demonstrate the superiority of our weighting scheme in predicting the visual quality of the screen content images. It is interesting to note that the QTF-GM metric yields a very promising result that is comparable to that yielded by the VIFP algorithm and merely inferior to that yielded by our SQMS model.

F. Testing of the Parameter Sensitivity

The impact of the parameter values used in an IQA model on the correlation performance is also significant. The database is always composed of images with limited visual contents, distortion types and levels, far less than the contents of real-world images. Hence, a poor IQA algorithm sensitive to distinct parameter values is very likely to work well only on the testing databases but fail in assessing other images; conversely, a good IQA method must be tolerant to varying parameter values. In our SQMS metric, the most important λ is assigned the value one. We thereby evaluate the influence of λ on the prediction performance of our approach. The test

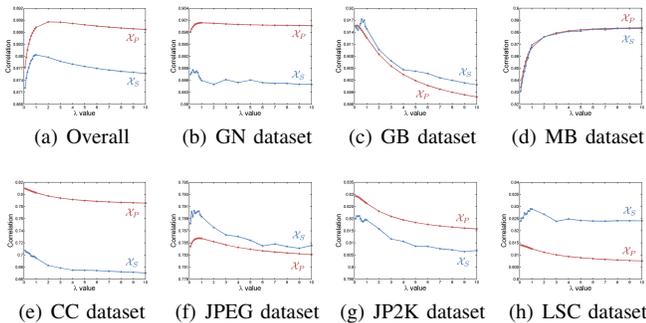


Fig. 12: Influence of varying λ values on the correlation performance (\mathcal{X}_P and \mathcal{X}_S) for the whole SIQAD database and each dataset.

includes 19 numbers, from 0.1 to 1 at an interval of 0.1 and from 2 to 10 at an interval of 1. We plot the results for \mathcal{X}_P and \mathcal{X}_S on the entire SIQAD database in Fig. 12(a). It can be easily seen that, on the one hand, our IQA metric is affected by the varying values of λ to a very small extent and, on the other hand, the worst performance result ($\mathcal{X}_P = 0.8742$, $\mathcal{X}_S = 0.8692$) is still larger than that of the second-ranking SIQM metric ($\mathcal{X}_P = 0.8625$, $\mathcal{X}_S = 0.8582$) and the third-ranking VIFP metric ($\mathcal{X}_P = 0.8489$, $\mathcal{X}_S = 0.8451$). We further report the performance results associated with different λ values on each of the seven datasets, as shown in Figs. 12(b)-(h). It can be noticed that except for the MB dataset, our SQMS model is fairly robust to varying λ values. We also find that the effects of the λ values are totally converse for the GB and MB datasets, which highlights the design of adaptive assignment for λ in various distortion types.

We have also tested the sensitivity of the other four parameters used in the two convolution operators of our SQMS metric: n and σ in the ‘‘Gaussian’’ kernel and t and θ in the ‘‘motion blur’’ kernel. For each of four parameters, we enumerate four numbers in a proper interval around the assigned value while fixing the other parameters. \mathcal{X}_P , \mathcal{X}_S and \mathcal{X}_K are used here for comprehensive validation. The sensitivity of testing parameters on the SIQAD database is presented in Table VIII. The values used in our SQMS technique are highlighted. It is clear that when the parameters change, the proposed IQA metric has a considerably stable performance that is particularly better than the second- and third-ranking SIQM and VIFP methods, and thus, it is quite robust and tolerant to the varying values of the parameters.

G. Discussion and Future Work

Based on the gradient magnitude similarity weighted by a fast saliency detection model, our SQMS metric demonstrates superior performance, computational efficiency and parameter insensitiveness. Moreover, we emphasize that our IQA metric is very applicable based on the following two aspects. The first is that it can be used for the perceptual coding of screen content images/videos. Following the predictive video coding framework, the perceptual weighting map derived from the original/input frame can be directly used to optimize the macroblock/coding unit level bit allocation by adjusting the

local quantization parameter (QP), such that the perceptually important areas are allocated more coding bits to improve the overall quality. The second is that it can facilitate the study of blind/no-reference (NR) IQA models of screen content images. Obtaining raw human opinion scores via subjective testing is usually time- and labor-consuming. Despite the emergence of the two recent SIQAD and SCD databases, the data available to train blind/NR IQA metrics are not sufficient. From the above tests and analyses, our IQA measure yields a high correlation with human judgement of quality and thus serves as an approximation of human opinion ratings for training blind/NR quality metrics [49], [51].

Some further efforts can be made to enhance the proposed method. First, because of the intrinsic edge-preserving ability, some complicated operators, such as the bilateral filter [52], may be good candidates for replacing the Gaussian weighting function for a better performance. Second, the proposed linearly weighted integration of fixation- and saccade-inspired saliency maps is simple and effective but not the best choice because for a given area, humans are inclined to one manner over another, and thus, an adaptive or nonlinearly weighted combination would be better. Third, our weighting strategy can be appropriately fused with saliency detection models, such as QTF [31] that performs fairly well in screen content IQA, to derive a more effective weighting technology. Fourth, the framework used in [10] (or [15]) can be employed to form a weighting map by systematically using saliency features from the original and distorted images (or measuring the saliency similarity between the original and distorted images) instead of our pooling technology towards a better performance. Fifth, although the IQA technique in [22], [23] is inferior to the proposed quality measure, a better performance is likely to be obtained by properly combining the above two strategies, for example, dividing the input screen content images into textual and pictorial components followed by applying our SQMS algorithm.

IV. CONCLUSION

In this paper, we have investigated a hot research topic – screen content IQA. Screen content images are remarkably different from natural scene images because of a great amount of computer-created graphical content that usually consists of texts and graphics. We have developed a new IQA technique using saliency-guided gradient magnitude similarity. The proposed metric is simple and has low time complexity. Thorough experiments on the two novel SIQAD and SCD databases dedicated to screen content images confirm the high performance of our SQMS model compared to relevant classical and state-of-the-art IQA algorithms. We stress that our approach is suitable for the development of perceptual coding and blind/NR IQA metrics for screen content images. In the future, improved weighting and pooling strategies will be considered towards obtaining superior performance. In the future work, improved weighting and pooling models will be considered for obtaining superior performance. Our source code will be released at <http://www.ntu.edu.sg/home/wslin/> and <https://sites.google.com/site/guke198701/publications>.

REFERENCES

- [1] 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.
- [2] C. F. Lai, H. Wang, H.-C. Chao, and G. Nan, "A network and device aware QoS approach for cloud-based mobile streaming," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 747-757, Jun. 2013.
- [3] S. Wang and S. Dey, "Adaptive mobile cloud computing to enable rich mobile multimedia applications," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 870-883, Jun. 2013.
- [4] H. Shen, Y. Lu, F. Wu, and S. Li, "A high-performance remote computing platform," in *Proc. IEEE Int. Conf. Pervasive Computing and Communications*, pp. 1-6, Mar. 2009.
- [5] C.-Y. Huang, K.-T. Chen, D.-Y. Chen, H.-J. Hsu, and C.-H. Hsu, "GamingAnywhere: The first open source cloud gaming system," *ACM Trans. Multimedia Computing*, vol. 10, no. 1, pp. 173-185, Jan. 2014.
- [6] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Screen content coding based on HEVC framework," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1316-1326, Aug. 2014.
- [7] 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.
- [8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1398-1402, Nov. 2003.
- [9] 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.
- [10] K. Gu, G. Zhai, X. Yang, L. Chen, and W. Zhang, "Nonlinear additive model based saliency map weighting strategy for image quality assessment," in *Proc. IEEE Workshops on Multimedia Sig. Process.*, pp. 313-318, Sept. 2012.
- [11] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Structural similarity weighting for image quality assessment," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, pp. 1-6, Jul. 2013.
- [12] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, C.-W. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. T. Circuits Syst.*, vol. 4, no. 1, pp. 95-105, Mar. 2014.
- [13] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, Feb. 2006.
- [14] 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.
- [15] 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.
- [16] A. Liu, W. Lin, and M. Narvaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512, Apr. 2012.
- [17] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process.*, pp. 506-510, Oct. 2014.
- [18] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270-4281, Oct. 2014.
- [19] 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*, 2016, to appear.
- [20] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297-312, May 2011.
- [21] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 173-185, Jan. 2013.
- [22] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," in *Proc. IEEE Int. Workshops on Quality of Multimedia Experience*, pp. 257-262, Sept. 2014.
- [23] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408-4421, Nov. 2015.
- [24] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Sig. Process. Mag.*, vol. 28, no. 6, pp. 50-59, Nov. 2011.
- [25] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971-982, Apr. 2011.
- [26] X. Min, G. Zhai, Z. Gao, and K. Gu, "Visual attention data for image quality assessment databases," in *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 894-897, Jun. 2014.
- [27] S. Wang, K. Gu, S. Ma, W. Lin, X. Liu, and W. Gao, "Guided image contrast enhancement based on retrieved images in cloud," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 219-232, Feb. 2016.
- [28] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185-207, 2013.
- [29] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [30] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *J. Vision*, vol. 9, no. 3, pp. 1-24, Mar. 2009.
- [31] 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.
- [32] C. Kim and P. Milanfar, "Visual saliency in noisy images," *J. Vision*, vol. 13, no. 4, pp. 1-14, Mar. 2013.
- [33] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Visual saliency detection with free energy theory," *IEEE Sig. Process. Lett.*, vol. 22, no. 10, pp. 1552-1555, Oct. 2015.
- [34] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Reviews Neuroscience*, vol. 18, pp. 193-222, 1995.
- [35] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. CVPR*, pp. 1-8, Jun. 2007.
- [36] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1597-1604, Jun. 2009.
- [37] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194-201, Jan. 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. M. Henderson, "Human gaze control during real-world scene perception," *Trends in Cog. Sci.*, vol. 7, no. 11, pp. 498-504, Nov. 2003.
- [40] S. Shi, X. Zhang, S. Wang, R. Xiong, and S. Ma, "Study on subjective quality assessment of screen content images," in *Proc. IEEE Picture Coding Symp.*, pp. 75-79, May 2015.
- [41] A. Thiele, P. Henning, M. Kubischik, and K.-P. Hoffmann, "Neural mechanisms of saccadic suppression," *Science*, vol. 295, pp. 2460-2462, Nov. 2002.
- [42] D. G. Pelli and K. A. Tillman, "The uncrowded window of object recognition," *Nat. Neurosci.*, vol. 11, no. 10, pp. 1129-1135, Oct. 2008.
- [43] N. M. Putnam, H. J. Hofer, N. Doble, L. Chen, J. Carroll, and D. R. Williams, "The locus of fixation and the foveal cone mosaic," *J. Vision*, vol. 5, no. 7, pp. 632-639, May 2005.
- [44] B. Jähne, H. Haubecker, and P. Geibler, *Handbook of Computer Vision and Applications*. New York: Academic, 1999.
- [45] Z. Pan, H. Shen, S. Li, and N. Yu, "A low-complexity screen compression scheme for interactive screen sharing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 949-960, 2013.
- [46] Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union Recommendation ITU-R BT.500-13, 2012.
- [47] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636-650, Apr. 2000.
- [48] 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.
- [49] K. Gu, S. Wang, G. Zhai, S. Ma, and W. Lin, "Screen image quality assessment incorporating structural degradation measurement," in *Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 125-128, May 2015.
- [50] 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/>.
- [51] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, 2016, to appear.
- [52] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE ICCV*, pp. 836-846, Jan. 1998.

PLACE
PHOTO
HERE

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 the reviewer for IEEE T-IP, T-CSVT, T-MM, T-CYB, T-BC, SPL, Neurocomputing, SPIC, JVCI, SIViP, IET-IP, etc. He has reviewed more than 30 journal papers each year. His research interests include quality assessment, contrast enhancement and visual saliency detection.

PLACE
PHOTO
HERE

Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Germany.

PLACE
PHOTO
HERE

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from the Peking University, in 2014. He is currently a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From Apr. 2011 to Aug. 2011, he was with Microsoft Research Asia, Beijing, as an Intern. He has proposed more than 20 technical proposals to ISO/MPEG, ITU-T and AVS video coding standards.

His current research interests include video compression and image/video quality assessment.

PLACE
PHOTO
HERE

Xiaokang Yang (SM'04) received the B. S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2000. He is currently a Full Professor and Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From September 2000 to March 2002, he was a Research Fellow in Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has published over 80 refereed papers, and has filed six patents. His research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition.

PLACE
PHOTO
HERE

Huan Yang received the B.S. degree in computer science from the Heilongjiang Institute of Technology, Harbin, China, in 2007, the M.S. degree in computer science from Shandong University, Jinan, China, in 2010, and the Ph.D. degree in the School of Computer Engineering, Nanyang Technological University, Singapore, in 2015. Her current research interests include object detection/recognition, image processing, and visual quality assessment.

PLACE
PHOTO
HERE

Weisi Lin (F'16) received the Ph.D. degree from Kings College London. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with more than 340 refereed papers published in international journals and conferences. He has been on the Editorial Board of the IEEE Transactions on Image Processing (T-IP), the IEEE Transactions on Multimedia (T-MM, 2011-2013),

the IEEE Signal Processing Letters (SPL), and the Journal of Visual Communication and Image Representation (JVCI). He has been elected as an APSIPA Distinguished Lecturer (2012/13). He served as a Technical-Program Chair for 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. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

PLACE
PHOTO
HERE

Wenjun Zhang (F'11) received the B.S., M.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987 and 1989, respectively. From 1990 to 1993, He worked as a post-doctoral fellow at Philips Kommunikation Industrie AG in Nuremberg, Germany, where he was actively involved in developing HDMAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a full professor in the Department of Electronic Engineering in 1995. As the national HDTV TEEG project leader, he successfully developed the first Chinese HDTV prototype system in 1998. He was one of the main contributors to the Chinese Digital Television Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He holds more than 40 patents and published more than 90 papers in international journals and conferences.