BLIND QUALITY EVALUATOR FOR SCREEN CONTENT IMAGES VIA ANALYSIS OF STRUCTURE

Guanghui Yue[1,*], *Chunping Hou*[1], *and Weisi Lin*[2]

[1]School of Electrical and Information Engineering, Tianjin University, Tianjin, China 300072
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

ABSTRACT

Existing blind evaluators for screen content images (SCIs) are mainly learning-based and require a number of training images with co-registered human opinion scores. However, the size of existing databases is small, and it is labor-, time-consuming and expensive to largely generate human opinion scores. In this study, we propose a novel blind quality evaluator without training. Specifically, the proposed method first calculates the gradient similarity between a distorted image and its translated versions in four directions to estimate the structural distortion, the most obvious distortion in SCIs. Given that the edge region is easier to be distorted, the inter-scale gradient similarity is then calculated as the weighting map. Finally, the proposed method is derived by incorporating the gradient similarity map with the weighting map. Experimental results demonstrate its effectiveness and efficiency on a public available SCI database.

***Index Terms*—** Screen content image, quality assessment, blind, no reference, structure

1. INTRODUCTION

To date, many image quality assessment (IQA) methods have been designed and obtained considerable achievement for natural scene images (NSIs) [1, 2, 3]. Compared with NSIs, screen content images (SCIs) show different characteristics, consisting of camera-captured NSIs, computer-rendered graphics, text, etc. Traditional IQA methods designed for NSIs are incapable of solving the quality assessment problem of SCIs [4]. New techniques are highly desired for SCIs. Unfortunately, only limited efforts have been made in the past years. Generally, existing methods lie in full-reference (FR), reduced-reference (RR) and no-reference (NR) measures. FR and RR measures are reference-based and only suit for situations where complete or partial reference information is available [5, 6, 7]. However, we cannot obtain the reference information in many situations. The NR measure (also denoted as the blind measure) that does not require reference information is desired [8, 9, 10, 11].

This work was supported by the National Natural Science Foundation of China under Grants 61520106002 and 61731003. *Corresponding author.

In recent years, only very few NR attempts have been conducted for evaluating SCIs. In the exploration phase, these methods are mainly learning-based, i.e., designing a quality assessment model that connects the quality-sensitive features and human opinion scores. For instance, Fang *et al*. [10] first extracted both local and global features to characterize luminance and texture information. Then, the quality prediction model was built via the support vector regression (SVR) to bridge the relationship between features with human opinion scores. Inspired by orientation selectivity mechanism, Lu *et al*. [12] captured statistical orientation and structure features to estimate the visual distortion of the distorted SCIs. Similar to reference [10], a quality estimation model was built by employing the SVR. In view of the fact that learning-based methods may suffer from over-fitting problems, it would obtain more promising performance by utilizing big-data training samples. Gu *et al*. [13] learnt a quality prediction model by collecting 100, 000 distorted images. To avoid the labor of human ratings, quality scores of these images were labeled by an FR method designed for SCIs. Shao *et al*. [8] learnt both local and global dictionaries with quality constrained. Subsequently, the test image's quality was estimated by combining the local and global scores obtained from two dictionaries. During dictionary construction stage, the quality scores were coarsely estimated by a number of FR IQA methods.

Although preliminary success has been achieved by these learning-based NR methods, we still need to move forward due to the following reasons. First of all, these learning-based methods require a large number of training samples to obtain a good prediction model. However, existing database are small. It is labor-consuming and expensive for training sample generation via subjective evaluation. Second, the distortion type should be known in advance for these methods that utilize the scores calculated by FR methods as training labels [8]. Therefore, these methods are not completely blind and have limited application scope. Last but not the least, by utilizing the scores of FR methods as training labels, the performance of the generated NR model is directly determined by the effectiveness of the selected FR methods. To obtain good results, it requires designers' experience to choose suitable FR methods. Obviously, the future of such an NR method is directly restricted by the development of FR methods.

In this study, we attempt to design a completely blind quality assessment method for SCIs. Contrary to existing learning-based NR methods, it is free of training operation, thereby avoiding the potential over-fitting problem caused by insufficient training samples and performance bias induced by the selected FR methods for label generation. Compared with non-learning FR methods, it provides a novel way for quantifying structural distortion without a reference image. The key strategy of the proposed method rests with the use of the distorted image's self-similarity to extract structural information, which is sensitive to distortions. Specifically, the structural information is extracted by calculating the gradient similarity between a distorted image and its translated versions along four directions. Since the edge region is easily distorted, a weighting map is calculated by obtaining the similarity map between the distorted image and its blurred version. By combining the structural information with the weighting map, the proposed method obtains good performance.

2. METHODOLOGY

2.1. Structure Extraction

Structure information has been found as the main sensitive element during visual perception [14]. SCIs contain distinctive structural characteristics, such as text, graphic line, and so on. How to effectively quantify the structure distortion is the critical issue for designing the IQA method of SCIs. In the literature, gradient magnitude has been broadly validated and used to characterize structural information [15]. However, it is incompetent for estimating the distortion by merely calculating the gradient map of the test image. Actually, many IQA methods calculate the gradient similarity for structure variation estimation between the reference and distorted images [5, 16, 17]. Much to our regret, the reference image is unavailable in designing NR IQA method. In this study, we propose a new way to extract the gradient similarity and estimate the structure distortion according to the intrinsical characteristics of SCIs.

Fig. 1 illustrates an example of SCIs suffered from Gaussian blur, Motion blur and JPEG compression. For reader's convenience, we enlarge the text region marked by the red rectangle and picture region marked by the green rectangle. It is clear that different distortions induce diverse appearance changes over the reference image (i.e., Fig. 1(a)). The text region and picture region exhibits different responses to the distortion. In spite of this, one can still observe that all these distortions corrupt the image main edges (where the gradient magnitude is large) on both text and picture regions. Specifically, these edges are either blurred or shifted horizontally/vertically. As the graphic line and text in SCIs contain obvious edges, we propose to estimate the structure variation by the following operations. First, for an image I, we first obtain its translated versions $I^n(n = \{1, 2, 3, 4\})$ by shifting its

Fig. 1. Illustration of SCIs. (a) is the reference image and (b)$-$(d) are its corrupted versions processed by Gaussian blur, Motion blur and JPEG compression, respectively.

pixel position with distance d along four directions, i.e., horizontal, vertical, main-diagonal and secondary-diagonal. d is arranged as 2 in this study. Obviously, the shifting operation has large influence on the main edges but is with little influence on the smooth region. Then, the gradient similarity G_S^n between I and I^n is calculated by:

$$G_S^n(x, y) = \frac{2G_0(x, y) \cdot G_n(x, y) + T_1}{G_0^2(x, y) + G_n^2(x, y) + T_1}, \tag{1}$$

where (x, y) stands for the pixel position. T_1 is a constant. Here, it is set as 600. By Eq. (1), it can effectively capture the influence of distortion on these regions (e.g., graphic line, text and main edges in the picture). G_0 is the gradient magnitude of I, and G_n is the gradient magnitude of translated version in the n-th direction. By definition, G_0 is computed by:

$$G_0 = \sqrt{(I \otimes g_x)^2 + (I \otimes g_y)^2}, \tag{2}$$

where $\otimes$ is the convolutional operator. g_x and g_y are filter kernels in horizontal and vertical directions, respectively. In this study, Scharr operators are adopted as filter kernels [18], i.e.,

$$g_x = \frac{1}{16}\begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix}, \quad g_y = g_x^T, \tag{3}$$

where symbol ^{T} is the transpose operator. Likewise, G_n can also be obtained using the same definition in Eq. (2). By Eq. (1), we can totally obtain four gradient similarity maps along diverse directions. Obviously, the proposed gradient similarity generation method, to some extent, considers the gradient direction, which is sensitive to structure distortion. As discussed in reference [5], maximum operation is a simple but effective measure to simplify and generate the structure variation map. Inspired by this, we generate the structure variation map by selecting the maximum value among the gradient similarity responses over all directions:

$$G(x,y) = max(G_S^n(x,y)), \;\; n = \{1,2,3,4\}. \tag{4}$$

Figs. 2(a)−2(d) present the structure variation maps of Figs. 1(a)−1(d). A higher gray value denotes to higher gradient similarity. One can intuitively observe that the structure variation map of the reference image has lower values along the contour of main edges. This is attributed to that the reference image contains clear edge boundary. Its pixel values in smooth region do not have significant influence, while those in edge regions change greatly when it is shifted to the translated version. Therefore, once processed by Eq. (1), the edge contour is with lower similarity, whereas the smooth region is with higher similarity. By contrast, the edge boundaries of SCIs processed by Gaussian blur and Motion blur are dispersive (as shown in Figs. 1(b)−1(c)), while that of the SCI processed by JPEG compression is full of blocks and ringing effects (as shown in Fig. 1(d)). During gradient similarity computation, the edge surrounding region of the distorted SCIs has completely different effects as compared to that of the reference image, thereby leading to diverse image structure variation maps as shown in Figs. 2(a)−2(d). Specifically, the structure variation map exhibits the intrinsical characteristics of distortions, e.g., edge dispersion, ringing, etc. Readers are encouraged to enlarge Fig. 2 and see the edges for more details. It is worth stressing that similar observations can also be found in other distortion types, like Gaussian noise, JPEG2000 compression, and contrast change. Due to space limitation, we do not present them here.

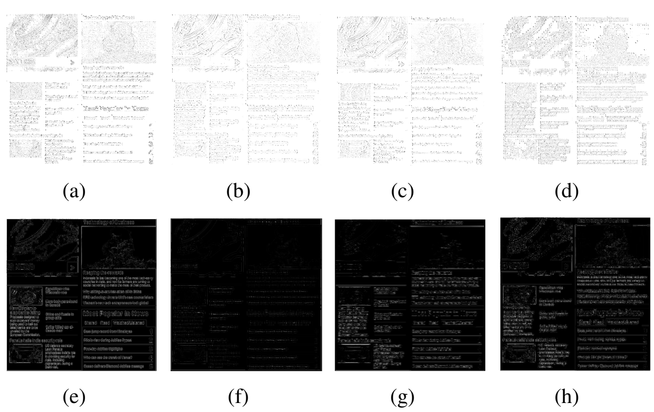

Fig. 2. Illustration of the structure variation maps and weighting maps. (a)−(d) and (e)−(h) are the structure variation maps and weighting maps of Figs. 1(a)−1(d), respectively.

2.2. SCI Quality Assessment

After obtaining the structure variation map, our next task is how to estimate the quality score from it. In the literature, mean pooling [19] and standard deviation pooling [20] have been validated effectively in quality assessment of NSIs. Since SCIs contain complex contents, directly utilizing mean pooling or standard deviation pooling may ignore the content's influence. In this study, we propose to utilize the weighting pooling by considering the distortion characteristics of SCIs. Revisiting Fig. 1, it is obvious that the surrounding region of main edges are more sensitive to distortions than other regions. Therefore, such regions should receive more attention during the pooling stage. Here, a simple but effective measure is used to extract the weighting map. To be specific, we first process the distorted image I with a Gaussian blur operation. Then, the gradient similarity G_f between I and the Gaussian-blurred version I_2 is computed by:

$$G_f(x,y) = \frac{2G_0(x,y)\cdot G_b(x,y) + T_2}{G_0^2(x,y) + G_b^2(x,y) + T_2}, \tag{5}$$

where $T_2 = 1$ is utilized for stability. G_b is the gradient magnitude (calculated by Eq. (2)) of I_2. To generate I_2, a two-dimensional symmetric Gaussian kernel is used with window size $[2d+1, 2d+1]$, variance $\sigma = 1.5$. By such arrangement, the window size exactly covers the edge surrounding region highlighted by the shifting operation. Through observation, we find that the edge surrounding region has smaller similarity value, while the smooth region has larger similarity value. To emphasize the importance of the edge surrounding region, the weighting map G_w is finally calculated as $G_w = 1 - G_f$. Figs. 2(e)-2(h) show the weighting maps of Figs. 1(a)-1(d). As seen, the edge surrounding region is arranged with high importance.

In the end, the quality of a test SCI can be estimated by pooling the structure variation map with the weighting map:

$$S = \frac{\sum_{(x,y)\in\Lambda} G(x,y)\cdot G_w(x,y)}{\sum_{(x,y)\in\Lambda} G_w(x,y)}, \tag{6}$$

where Λ is the pixel coordinate set of the test SCI.

3. EXPERIMENTAL RESULTS

3.1. Experimental Protocol

The public available SCI database (SIQAD [4]) is selected as the test platform. It consists of 20 high-quality reference images. For each reference image, it is processed by seven distortion types at seven levels, including Gaussian blur, Motion blur, Gaussian noise, JPEG compression, JPEG2000 compression, contrast change and layer segmentation based coding. As a result, 980 distorted images are totally involved in this database, and the quality score of each image is reported through subjective experiments.

Four evaluation criteria are used to measure the prediction monotonicity and accuracy of objective IQA methods. To be specific, Spearman Rank order Correlation Coefficient (SRCC) and Kendall's Rank-order Correlation Coefficient (KRCC) are utilized to evaluate the prediction monotonicity. Whereas, Pearson Linear Correlation Coefficient (PLCC) and Root

Mean Square Error (RMSE) are taken as the criteria for measuring prediction accuracy. As suggested by [21], a nonlinear regression procedure is required for reducing the nonlinearity of estimated scores by objective IQA methods before calculating PLCC and RMSE. Here, the five-parameter logistic regression used in [17, 22] is adopted for nonlinear regression.

3.2. Performance Comparison

Ten IQA methods are selected for comparisons. They can be divided into two groups. The methods in the first group are reference-based, including FSIM [19], MAD [23], GSIM [15], GSS [5], SFUW [17], SIRR [6] and Wang' method [22]. FSIM, MAD, GSIM are specifically designed for NSIs, while the remainder is proposed for SCIs. The second group contains three opinion-unaware NR IQA methods, such as NIQE [24], IL-NIQE [25] and BQMS [13]. These three methods aim to build a quality assessment model without the help of human opinion scores (actually, BQMS utilizes the estimated scores by an FR method as the label for model construction). Table 1 lists the comparison results, where the best performance in each type is highlighted in boldface. Since we fail to obtain the code of SFUW, SIRR and Wang's method, the associated results are directly copied from the original papers.

Table 1. Performance Comparisons.

Method	Type	PLCC	SRCC	KRCC	RMSE	Time (s)
FSIM [19]	FR	0.591	0.582	0.425	11.551	0.762
MAD [23]	FR	0.619	0.607	0.461	11.241	0.718
GSIM [15]	FR	0.569	0.548	0.405	11.775	0.117
GSS [5]	FR	0.846	0.836	0.639	7.631	1.168
SFUW [17]	FR	**0.891**	**0.880**	-	**6.499**	-
SIRR [6]	RR	0.754	0.729	-	9.403	-
Wang [22]	RR	**0.801**	**0.766**	0.576	**6.802**	-
NIQE [24]	NR	0.341	0.370	0.255	13.467	0.070
IL-NIQE [25]	NR	0.388	0.322	0.228	13.206	20.994
BQMS [13]	NR	0.755	0.722	0.530	9.304	9.038
Proposed	NR	**0.768**	**0.734**	**0.545**	**9.173**	**0.016**

From the data, we have the following observations. First, those methods designed for NSIs (i.e., FSIM, MAD and GSIM) merely obtain general performance and exhibit their powerlessness in evaluating SCIs, even though they are reference-based. In contrast, it is obvious to find that these reference-based IQA models (i.e., GSS, SFUW, SIRR and Wang' method) specifically proposed for SCIs are more competent for addressing the quality assessment problem of SCIs with good performance. Second, these reference-based methods (specifically designed for SCIs) outmatch those reference-free methods (i.e., NIQE, IL-NIQE, BQMS and the proposed method). Third, the proposed NR method is superior to other competing NR methods. More specifically, it leaves a large room to NIQE and IL-NIQE (which is designed for NSIs without the help of human opinion scores) with the SRCC increment more than 0.35. Moreover, the proposed method still gains the upper hand even compared with the learning-based NR method BQMS. Reasonable explanations about above phenomena can be attributed to the following aspects. On the one hand, SCIs contain more complex contents than NSIs. It is acceptable that these methods designed for NSIs fail to evaluate SCIs by only considering the characteristics of NSIs. Compared with NR methods, FR methods can use the reference information, thereby containing more advantages and accordingly achieving higher performance. Theoretically, a higher performance could be obtained when more reference information is in hand. Therefore, it can be observed that FR methods are superior to RR methods. On the other hand, NIQE and IL-NIQE are built with the hypothesis that high-quality NSIs satisfy a certain regularity, which can be corrupted by distortions. The distance between the extracted features of the distorted image and the built model from high-quality images can quantify the distortion degree. Unfortunately, SCIs do not obey such regularity [10]. As a result, NIQE and IL-NIQE obtain unsatisfied performance. As for BQMS, it extracts a number of quality-sensitive features based on the characteristics of SCIs and builds the prediction model by connecting the features and quality labels. However, the quality labels are estimated by an FR method. There is no doubt that such method is restricted to and affected by the performance of the selected FR method [8]. In contrast, the proposed method aims to address the quality assessment problem from the analysis of main characteristics of SCIs. Specifically, it observes that the main edges in graphic line, text and picture regions are more easily corrupted and captures the structure distortion by calculating the self-similarity between the distorted image and its shifted versions. Therefore, it obtains promising performance.

Apart from effectiveness comparison, we further investigate the run time of the proposed method and make comparisons with competing methods. To ensure the fairness, all algorithms are implemented on the MATLAB2016b software, which is operated on a computer with Intel E5-1650 CPU @3.20 GHz and 16 GB RAM. In the last column of Table 1, we present the run time (recorded by *tic* and *toc* functions in seconds) costed on an image (i.e., 'lena') with the resolution 256×256. Clearly, our method is quite time-saving (requiring less than 0.02s) and superior to all competing methods.

4. CONCLUSION

This study proposes a new blind IQA algorithm for SCIs without the need of human opinion scores. The key strategy of the proposed method is to estimate structural distortion and generate the weighting map by analyzing the characteristics of SCIs. Based on the fact that the edge region is more sensitive to quality degradation, we quantify structural distortion by calculating the gradient similarity map between the distorted image and its translated ones. Also, the weighting map is obtained by calculating the gradient similarity map between the distorted image and its blurred version. Experimental results illustrate the effectiveness and efficiency of the proposed method.

5. REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[2] G. Yue, C. Hou, G. Ke, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2722–2732, 2018.

[3] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3784–3793, 2019.

[4] 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, 2015.

[5] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1394–1398, 2016.

[6] X. Min, K. Gu, G. Zhai, M. Hu, and X. Yang, "Saliency-induced reduced-reference quality index for natural scene and screen content images," *Signal Process.*, vol. 145, pp. 127–136, 2018.

[7] V. Jakhetiya, K. Gu, W. Lin, Q. Li, and S. P. Jaiswal, "A prediction backed model for quality assessment of screen content and 3-D synthesized images," *IEEE Trans. Ind. Inf.*, vol. 14, no. 2, pp. 652–660, 2018.

[8] F. Shao, Y. Gao, F. Li, and G. Jiang, "Toward a blind quality predictor for screen content images," *IEEE Trans. Syst. Man Cybern.*, vol. 48, no. 9, pp. 1521 – 1530, 2018.

[9] G. Yue, C. Hou, G. Ke, S. Mao, and W. Zhang, "Biologically inspired blind quality assessment of tone-mapped images," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2525–2536, 2018.

[10] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, 2018.

[11] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for DIBR-synthesized image quality evaluation," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2075–2088, 2019.

[12] N. Lu and G. Li, "Blind quality assessment for screen content images by orientation selectivity mechanism," *Signal Process.*, vol. 145, pp. 225–232, 2018.

[13] 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, 2016.

[14] X. Ran and N. Farvardin, "A perceptually motivated three-component image model-Part I: description of the model," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 401–415, 1995.

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

[16] 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. 1098–1110, 2016.

[17] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, 2017.

[18] B. Jähne, H. Haussecker, and P. Geissler, *Handbook of computer vision and applications*. Citeseer, 1999, vol. 2.

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

[20] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.

[21] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of objective models of video quality assessment," 2003. [Online]. Available: http://www.vqeg.org/2003

[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 Syst. Video Technol.*, vol. 28, no. 1, pp. 1–14, 2018.

[23] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011 006–1–011 006–21, 2010.

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

[25] 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, 2015.