Objective Quality Assessment of Screen Content Images by Structure Information

Yuming Fang[1], Jiebin Yan[1], Jiaying Liu[2], Shiqi Wang[3], Qiaohong Li[3], and Zongming Guo[2⋆]

[1]Jiangxi University of Finance and Economics, Nanchang, China
[2]Institute of Computer Science and Technology, Peking University, Beijing, China
[3]School of Computer Science and Engineering, Nanyang Technological University, Singapore

Abstract. In this paper, we propose a novel full-reference objective quality assessment metric of screen content images by structure information. The input screen content image is first divided into textual and pictorial regions. The visual quality of textual regions is predicted based on perceptual structural similarity, where the gradient information is used as the feature. To estimate the visual quality of pictorial regions, we extract the luminance and structure features as feature representation. The overall quality of the screen content image is measured by fusing those of textual and pictorial parts. Experimental results show that the proposed method can obtain better performance of visual quality prediction of SCIs than other existing ones.

Keywords: Visual quality assessment; screen content image; full-reference quality assessment

1 Introduction

Recently, there is one type of images emerging over Internet, which is called screen content images (SCIs). Generally, the SCI includes different forms of visual content, such as texts, pictures, and graphics. It has been emerging and widely used in various multimedia applications, including information sharing system between computer and smart devices [1], cloud computing systems [2] [3], remote conference, product advertising, *etc.* With the popularity of smart phones, more and more users would like to share different information with each other by rendering various visual content as the form of SCIs, where various multimedia processing methods might be involved, such as coding [7] [8] [9], transmission [3], *etc.* There are a large number of image processing algorithms proposed for SCIs, including SCI compression [4], SCI quality assessment [5], SCI segmentation [6], *etc.*

⋆ Corresponding Author. This work was supported by the NSFC (No. 61571212), NSF of Beijing (No.4142021), and NSF of Jiangxi Province (No. 20151BDH80003, 20161ACB21014).

During SCI processing such as acquisition, processing and transmission, there might be various visual distortions generated. When SCIs are created by the camera from smart phones, the noise and blurring distortions might be involved due to the camera motion. For the transmission of SCIs over Internet, the compression distortion might be generated. To evaluate the visual quality of these distorted SCIs, it is highly desired to design effective objective visual quality assessment metrics for various multimedia processing applications.

In the past decades, there have been various quality assessment methods designed for visual content. Traditional signal fidelity methods such as PSNR (Peak Signal-to-Noise Ratio), MSE (Mean Square Error) and MAE (Mean Absolute Error) predict visual quality of images by simply computing pixel differences between the reference and distorted images. These signal fidelity methods are widely used for visual quality assessment (VQA) in both industry and academia due to their simple and efficient implementation. However, they do not consider the properties of the Human Visual System (HVS), and thus, they might not obtain accurate quality prediction results as human beings perceive [14] [15]. To overcome the drawbacks of these existing metrics, many advanced perceptual IQA metrics have been proposed during the past decade [14].

Wang *et al.* proposed the well-known SSIM (structural similarity) by considering the characteristics of human beings' perception on image structure [16]. Following this perceptual VQA metric, there are various types of full-reference VQA metrics proposed in recent ten years [14]: VIF (visual information fidelity) [17], IGM (internal generative mechanism) [18], GSM (gradient similarity metric) [19], *etc.* Also, many reduced-reference metrics [10] and no-reference metrics [11] [12] [13] have been designed in the past decades. These IQA metrics are mainly designed for VQA of general images and they are not effective in VQA of SCIs. Recently, Yang *et al.* conducted an user study for VQA of SCIs [5]. Based on the detailed analysis of the subjective data on the constructed SCI database, the authors proposed an objective VQA metric to predict visual quality of SCIs [5]. However, the performance of that metric in [5] can be further improved. Thus, it is highly desired to design VQA metrics for SCIs for various potential SCI processing applications.

In this study, we propose a full-reference (FR) VQA metric for SCIs based on structure features. We first divide the SCI into textual and pitorial parts by the text segmentation method. The visual distortion of textual regions is predicted by structural similarity, where gradient information is used to extract the structure feature. For the pictorial regions, we calculate the luminance and structure features to estimate the visual quality. The overall quality of the SCIs can be predicted by fusing those of pictorial and textual regions. Experimental results show that the proposed method can obtain good performance in visual quality prediction of SCIs.

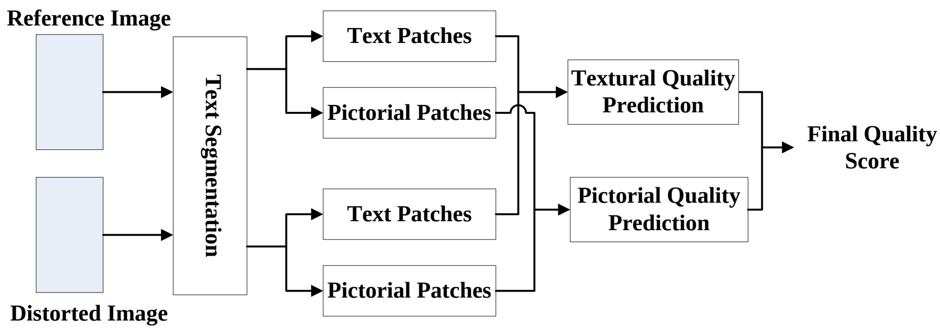

Fig. 1. The proposed framework.

2 The Proposed Method

Our previous study [5] has shown that the statistical features of natural and textual images are totally different. The detailed analysis of subjective data demonstrate that human perception on pictorial and textual regions is different with each other. Specifically, the distortion in textual regions are perceived differently from that of the overall SCI. Observers would be sensitive to the luminance and contrast change in pictorial regions, while for textual regions, they are more sensitive to blurring distortion than other types of distortion. Thus, it is reasonable to design different methods for visual quality assessment of pictorial and textual regions.

The framework of the proposed method is shown in Fig. 1. We first use the texture segmentation method in [20] to segment the SCI into pictorial and textual regions. Then the image is divided into two types of image patches: pictorial and textual patches. For textual patches, we extract the gradient to represent the feature for similarity computation. The luminance and texture features are used for similarity calculation of pictorial patches. The final quality of SCIs is obtained by fusing the visual quality of textual and pictorial patches. We will introduce the proposed method in detail in the following.

The content in the textual region of SCIs mainly includes various characters. Since characters are composed of various edges, we use the gradient to represent the structure feature of textual regions in SCIs. The following filters with two directions are adopted to compute the gradient of textual region: $h_x = [-1/2\ 0\ 1/2]$ and $h_y = [-1/2\ 0\ 1/2]'$. With these two filters, we can compute the structure feature for reference and distorted SCIs as follows.

$$g_{rx} = h_x \bigotimes T_r, \tag{1}$$

$$g_{ry} = h_y \bigotimes T_r, \tag{2}$$

$$g_{dx} = h_x \bigotimes T_d, \tag{3}$$

$$g_{dy} = h_y \bigotimes T_d, \tag{4}$$

where T_r and T_d represent textual patches in the reference and distorted images, respectively. (g_{rx}, g_{ry}) and (g_{dx}, g_{dy}) denote the gradient features with two directions for the textual patches in the reference and distorted images, respectively.

With the computed structure features in Eqs. (1)-(4), we calculate the similarity between the textual patches from the reference and distorted images as follows [16]:

$$S_i(g_{rk}, g_{dk}) = \frac{2\mu_{g_{rk}}\mu_{g_{dk}} + C_1}{\mu_{g_{rk}}^2 + \mu_{g_{dk}}^2 + C_1} \frac{2\sigma_{g_{rk}g_{dk}} + C_2}{\sigma_{g_{rk}}^2 + \sigma_{g_{dk}}^2 + C_2} \tag{5}$$

where $k \in \{x, y\}$; $S_i(g_{rk}, g_{dk})$ denotes the similarity between gradient features g_{rk} and g_{dk} for image patch i; μ_{rk} and μ_{dk} are mean values of the gradient g_{rk} and g_{dk}, respectively; σ_{rk} and σ_{dk} denote the standard variance values of the gradient g_{rk} and g_{dk}, respectively; $\sigma_{g_{rk}g_{dk}}$ is the covariance of the gradient g_{rk} and g_{dk}; C_1 and C_2 are two constant values.

After we calculate the similarity between gradient features with two directions for textual patches, we estimate the visual score of textual regions S_t in SCI as follows:

$$S_t = \frac{1}{N}\sum_{i=1}^{N}(\alpha S_i(g_{rx}, g_{dx}) + (1-\alpha)S_i(g_{ry}, g_{dy})) \tag{6}$$

where N is the number of textual patches in the SCI; α is a weighting parameter.

For pictorial patches, we use the structure and luminance features to predict the visual quality. First, the local contrast normalization is applied to the pictorial regions of SCIs to mimic early visual system and remove the redundancy information in the visual scene. The normalization operation can be implemented as [21].

$$P'(i,j) = \frac{P(i,j) - \mu_p}{\sigma_p + C} \tag{7}$$

where $P'(i,j)$ and $P(i,j)$ represent the normalized and original values at location (i,j) in pictorial regions; μ_p and σ_p denote the mean and standard variance values of pictorial regions; C is a constant parameter.

With the local normalized $P'(x,y)$, we can extract the structure feature of pictorial regions by using the rotation invariant uniform LBP descriptor [22]. The general LBP representation can be formulated as follows.

$$LBP_{K,R} = \sum_{i=0}^{K-1} t(p_i - p_c)2^i, \tag{8}$$

$$t(p_i - p_c) = \begin{cases} 1, (p_i - p_c) \geq 0 \\ 0, (p_i - p_c) < 0 \end{cases} \tag{9}$$

where K and R denote the number of neighbors and the radius of the neighborhood; p_c is the normalized luminance value of the center pixel in the local patch; $(p_0, p_1, ..., p_{K-1})$ represent the normalized luminance values of K circularly symmetric neighborhood. Based on the study [22], we can define the local rotation invariant uniform LBP operator as:

$$LBP'_{K,R} = \begin{cases} \sum_{i=0}^{K-1} t(p_i - p_c), U(LBP_{K,R}) \leq 2 \\ K+1, Otherwise \end{cases} \tag{10}$$

$$U(LBP_{K,R}) = \|t(p_{K-1} - p_c) - t(p_0 - p_c)\| + \sum_{i=0}^{K-1} \|t(p_i - p_c) - t(p_{i-1} - p_c)\| \tag{11}$$

where U is computed as the number of bitwise transitions.

After extracting the LBP features by using the LBP descriptor in Eq. (10), we further calculate the histogram of LBP features in each pictorial patch. Here, we set the bin of the histogram as 10 and thus obtain the structure feature with 10 elements $\{f_1, f_2, ..., f_{10}\}$. Meanwhile, we also calculate the normalized luminance histogram as the luminance feature in the proposed method. Similarly, we set the bin of the histogram as 10, and thus, there are 10 elements $\{f_{11}, f_{12}, ..., f_{20}\}$ in the luminance feature. In total, there is one feature vector with 20 elements for structure and luminance features for each pictorial patch i: $f_i = \{f_1, f_2, ..., f_{20}\}$. We predict the visual quality of pictorial regions in the SCI by the difference between pictorial patches from the reference and distorted SCIs.

$$S_p = \frac{1}{N} \sum_{i}^{N} e^{(-q_i)} \tag{12}$$

$$q_i = \sqrt{\sum_{j=1}^{20} (f_i - f'_i)^2} \tag{13}$$

where f_i and f'_i denote the used luminance and structure features from the reference and distorted SCIs.

After computing the visual quality of textual and pictorial regions in SCIs, we predict the final visual quality of each input SCI by combing them as follows.

$$S = \beta S_t + (1 - \beta) S_p \tag{14}$$

where β is a weighting parameter with the range [0,1].

3 Experimental Results

To demonstrate the advantages of the proposed method, we use the image database in [5] to conduct the comparison experiments. This database includes 20 reference SCIs in total. For each SCI in this database, there are seven distortion types (Gaussian Noise, Gaussian Blur, Motion Blur, Contrast Change, JPEG, JPEG2000, and Layer Segmentation Based Coding) with seven degradation levels and thus there are 980 distorted SCIs. These reference images obtained from webpages, slides, PDF files and digital magazines are diverse with the visual content.

The 11-category Absolute Category Rating (ACR) is used in the subjective experiment. In total, there were 96 subjects involved in the test and each image was rated by at least 30 subjects. The participants' ages range from 19 to 38 years. After the raw subjective scores were obtained, outliers were removed to obtain the DMOS.

Here, we adopt two commonly used methods to compute the correlation between the subjective and objective scores: SRCC (Spearman Rank-order Correlation Coefficient), and PLCC (Pearson Linear Correlation Coefficient). SRCC can be used to evaluate the prediction monotonicity, while PLCC can be adopted to assess the prediction accuracy. Generally, a better visual quality assessment method has higher SRCC and PLCC values. Given the ith image in the database (with N images in total), its objective and subjective scores are o_i and s_i. We can estimate the PLCC as follows.

$$PLCC = \frac{\sum_{i=1}^{N}(o_i - \overline{o})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(o_i - \overline{o}) * \sum_{i=1}^{N}(s_i - \overline{s})}} \tag{15}$$

where $\overline{o}$ and $\overline{s}$ denote the mean values of o_i and s_i, respectively.

SRCC can be computed as follows.

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2 - 1)} \tag{16}$$

where e_i is the difference between the ith image's ranks in subjective and objective results.

Table 1. Experimental results of the proposed method and other existing methods.

Components	PSNR	SSIM	VIF	IFC	MAD	GMSD	SPQA	Proposed
PLCC	0.5869	0.5912	0.8206	0.6395	0.6191	0.7259	0.8584	0.8656
SRCC	0.5608	0.5836	0.8069	0.6011	0.6067	0.7305	0.8416	0.8642

In this experiment, we perform the comparison experiments by using the proposed method and the following existing visual quality metrics: PSNR, SSIM [16], VIF [17], IFC [23], MAD [25], GMSD [24], SPQA [5]. We compute the correlations in terms of PLCC and SRCC values between the subjective scores and the predicted objective scores from the used compared metrics. The experimental results are shown in Table 1.

From Table 1, we can observe that GMSD can obtain better performance than PSNR, SSIM, IFC and MAD in visual quality prediction of SCIs. Obviously, SPQA an VIF can obtain much better performance than other existing studies. Among all the compared metrics, the proposed method can obtain best performance on visual quality prediction of SCIs, which can be demonstrated by the highest PLCC and SRCC values in Table 1.

4 Conclusion

In this study, we have proposed a new full-reference VQA metric for SCIs based on structural information. For textual regions in SCIs, we extract the structure features by using the gradient information for visual quality prediction of textual regions. For pictorial regions in SCIs, the luminance and structure features are computed by intensity and LBP information, respectively, for visual quality prediction of pictorial regions. The final visual quality of SCIs is estimated by fusing these of textual and pictorial parts. In the future, we will further investigate how to fuse the visual quality scores of textual and pictorial regions to obtain more reasonable quality scores of SCIs.

References

1. T. Chang and Y. Li, Deep Shot: A Framework for Migrating Tasks Across Devices Using Mobile Phone Cameras, *in ACM Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
2. Y. Lu, S. Li, and H. Shen, Virtualized Screen: A Third Element for Cloud-Mobile Convergence, in *IEEE Multimedia*, Vol. 18, No. 2, PP. 4-11, 2011.
3. H. Hu, Y. Wen, T.-S. Chua, and X. Li, Toward scalable systems for big data analytics: A technology toturial, *IEEE Access*, Vol. 2, PP. 652-687, 2014.
4. C. Lan, G. Shi, and F. Wu, Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation, *IEEE Transactions on Image Processing*, vol. 19, pp. 946957, 2010.
5. H. Yang, Y. Fang, and W. Lin, Perceptual quality assessment of screen content images, *IEEE Transactions on Image Processing*, 24(11), 4408-4421, 2015.
6. S. Minaee, and Y. Wang, Screen content image segmentation using least absolute deviation fitting, *IEEE International Conference on Image Processing*, 2015.
7. J. Lei, S. Li, C. Zhu, M.-T. Sun, and C. Hou, Depth coding based on depth-texture motion and structure similarities, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.2, pp.275-286, 2015.
8. J. Lei, J. Sun, Z. Pan, S. Kwong, J. Duan, and C. Hou, Fast mode decision using inter-view and inter-component correlations for multiview depth video coding, *IEEE Transactions on Industrial Informatics*, vol.11, no.4, pp.978-986, 2015.
9. S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, Nonlocal In-Loop Filter: The Way Toward Next-Generation Video Coding? *IEEE MultiMedia* 23(2), 16-26, 2016.
10. J. Wu, W. Lin, G. Shi, and A. Liu, Reduced-Reference Image Quality Assessment with Visual Information Fidelity, *IEEE Trans. on Multimedia*, Vol. 12, No. 7, PP. 1700-1705, Nov. 2013.
11. K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, No-Reference Image Sharpness Assessment in Autoregressive Parameter Space. *IEEE Trans. Image Processing*, 24(10): 3218-3231, 2015.
12. Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, No-reference quality assessment for contrast-distorted images based on Natural scene statistics, *IEEE Signal Processing Letter*. Vol. 22(7): 838-842, 2015.
13. L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments. *IEEE Trans. Cybernetics* 46(1): 39-50, 2016.

14. W. Lin, and C. C. J. Kuo, Perceptual visual quality metric: a survey. *Journal of Visual Communication and Image Representation*, 22(4), 297-312, 2011.
15. Z. Wang, and A. C. Bovik, Mean squared error: love it or leave it? a new look at fidelity measures. *IEEE Signal Processing Magazine*, 26(1), 98-117, 2009.
16. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, Vol. 13(4), 600-612, 2004.
17. H. R. Sheikh, and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, vol. 15 (2), 430-444, 2006.
18. J. Wu, W. Lin, G. Shi, and A. Liu. Perceptual Quality Metric With Internal Generative Mechanism. *IEEE Transactions on Image Processing*, 22(1): 43-54, 2013.
19. A. Liu, W. Lin, and M. Narwaria. Image Quality Assessment Based on Gradient Similarity. *IEEE Transactions on Image Processing*, 21(4): 1500-1512, 2012.
20. H. Yang, W. Lin, and C. Deng, Image acitivity measure (IAM) for screen image segmentation, *IEEE International Conference on Image Processing*, 2012.
21. L. Zhang, L. Zhang, and A. Bovik, 'A feature-enriched completely blind image quality evaluator,' *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 25792591, 2015.
22. T. Ojala, M. Pietikainen, and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971987, 2002.
23. H. R. Sheikh, A. C. Bovik, and G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117-2128, 2005.
24. W. Xue, L. Zhang, X. Mou, and A. C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684-695, 2014.
25. E. C. Larson and D. M. Chandler, Most apparent distortion: fullreference image quality assessment and the role of strategy, *Journal of Electronic Imaging*, vol. 19, no. 011006, 2010.