IEICE TRANS. INF. & SYST., VOL.E100–D, NO.3 MARCH 2017

LETTER

Naturalization of Screen Content Images for Enhanced Quality Evaluation

Xingge GUO[†], Liping HUANG[†], Ke GU[††], Leida LI[†a)], Zhili ZHOU[†††], *Nonmembers*, *and* Lu TANG[†], *Member*

SUMMARY The quality assessment of screen content images (SCIs) has been attractive recently. Different from natural images, SCI is usually a mixture of picture and text. Traditional quality metrics are mainly designed for natural images, which do not fit well into the SCIs. Motivated by this, this letter presents a simple and effective method to naturalize SCIs, so that the traditional quality models can be applied for SCI quality prediction. Specifically, bicubic interpolation-based up-sampling is proposed to achieve this goal. Extensive experiments and comparisons demonstrate the effectiveness of the proposed method.

key words: *image quality assessment, screen content image, naturalization, bicubic interpolation*

1. Introduction

Screen content images (SCIs) have been increasingly prevalent in modern multi-client communication systems, such as remote education, video conference, cloud gaming, etc [1]. In such applications, the complicated graphical interfaces are transmitted to the end users so that even thin-clients can enjoy the computationally intensive and graphically rich services. In the generation of SCIs, distortions can be easily introduced, including blurring, contrast, compression artifacts, etc. Therefore, the objective quality assessment of SCIs is highly desirable.

Extensive image quality assessment (IQA) metrics have been proposed in the past several years [2]. According to the amount of information needed from reference images, IQA metrics can be classified into full-reference (FR), reduced-reference (RR) and no-reference (NR). The existing quality metrics are mainly designed for natural images, which have quite different statistical properties from SCIs. So the traditional quality metrics do not perform well on SCIs. Recently, several SCI quality models have been reported. Yang *et al.* [1] first proposed the quality assessment of screen content images. A Screen Image Quality Assessment Database (SIQAD) was first constructed with subjective experiments. Then an objective quality model was proposed by considering the difference between textual and pictorial regions. In [3], Gu *et al.* proposed a method by combining the measurements of structural distortions and visual saliency. Wang *et al.* [4] proposed a metric by incorporating visual field adaptation and information content weighting into the evaluation of structural similarity. In [5], Ni *et al.* first extracted gradient direction in SCIs. Then a deviation based pooling strategy was proposed to generate the quality score for SCIs. While these metrics have achieved very encouraging results, another interesting problem in parallel is that: is it possible to adapt the current natural image quality models to the characteristics of SCIs. If yes, we have a new choice for SCI quality evaluation. In this letter, we address this less investigated problem.

SCIs often consist of picture and text, so the statistical properties of SCIs are different from those of natural images. This is the main reason why the traditional natural image quality metrics do not work well on SCIs. Inspired by this, this letter presents a unified naturalization method, which can be applied on SCIs and make them have more similar statistical properties with natural images. By this means, natural image quality models are adapted to SCIs. The proposed method is extremely simple and can be achieved by bicubic interpolation-based up-sampling. The performance of the proposed method is verified by experiments.

2. Statistical Differences between Screen Content and Natural Images

In order to have an intuitive understanding of the proposed method, we first illustrate the different statistical properties of screen content and natural images. The SCIs differ from natural images mainly in that the textual regions in SCIs are sharp and have significant variations. By contrast, natural images are relatively smooth. In [6], Yang *et al.* proposed an image activity measure (IAM) for screen content image segmentation. IAM measures image activity and it reflects the complexity of an image. For an image block **b**, the block activity measure (BAM) is defined as [6]:

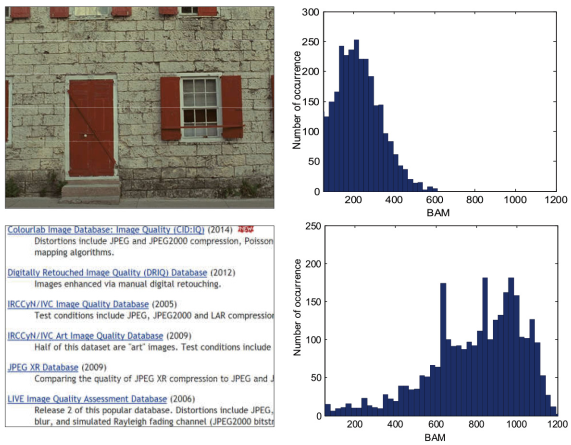
$$\mathcal{BAM} = \frac{\alpha\sqrt{\mathcal{V}_1} + (1-\alpha)\sqrt{\mathcal{V}_2}}{m \times n}, \tag{1}$$

where $m \times n$ denotes the block size, α is a weighting factor (equals to 0.5 in this work), $\mathcal{V}_1$ is the sum of the 1-distance down-left diagonal and down-right diagonal variances, and $\mathcal{V}_2$ is the sum of the 2-distance horizontal and vertical variances. $\mathcal{V}_1$ and $\mathcal{V}_2$ are defined as:

Manuscript received August 30, 2016.
Manuscript revised October 30, 2016.
Manuscript publicized November 24, 2016.

[†]The authors are with the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China.
[††]The author is with the School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore.
[†††]The author is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China.
a) E-mail: lileida@cumt.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2016EDL8178



(a) (b)

Fig. 1 Sample images and their BAM histograms. (a) Natural and textual images; (b) BAM histograms.

$$\mathcal{V}_1 = \sum_{i=1}^{m}\sum_{j=1}^{n-1}(\mathbf{b}_{i,j} - \mathbf{b}_{i-1,j+1})^2 + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(\mathbf{b}_{i,j} - \mathbf{b}_{i+1,j+1})^2, \tag{2}$$

$$\mathcal{V}_2 = \sum_{i=2}^{m-1}\sum_{j=1}^{n}(\mathbf{b}_{i-1,j} - \mathbf{b}_{i+1,j})^2 + \sum_{i=1}^{m}\sum_{j=2}^{n-1}(\mathbf{b}_{i,j-1} - \mathbf{b}_{i,j+1})^2. \tag{3}$$

where $\mathbf{b}_{ij},\ i = 1,\cdots,m,\ j = 1,\cdots,n$, denotes the intensity value.

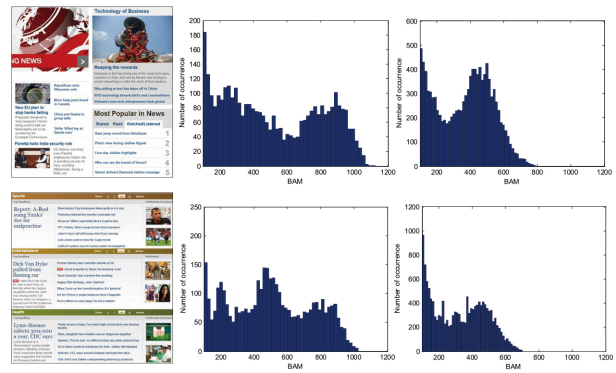
Figure 1 shows a natural image and a textual image, together with their BAM histograms. In this experiment, the images are divided into non-overlapping 8×8 blocks to compute the BAMs. It is observed from the figure that most BAM values of the natural image are smaller than those of the textual image. We have done this experiment on extensive natural and screen content images with diversified contents, and we find this property holds consistently. Therefore, the BAM histograms of natural images typically have left-side distribution, while those of textual images typically have right-side distribution.

3. Proposed SCI Naturalization Method

Inspired by the characteristics of BAM histograms, we propose a simple method for the naturalization of SCIs. The underlying idea is to make SCIs more natural. Specifically, we propose to use bicubic interpolation-based up-sampling for SCI naturalization.

In bicubic interpolation, the interpolation result of a pixel (x, y) is defined as the the weighted average of its 16 (4×4) neighboring pixels. It approximates the local intensity values using a bicubic polynomial surface. The general form for bicubic interpolation is as follows:

$$f(x, y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij} x^i y^j, \tag{4}$$



(a) (b) (c)

Fig. 2 BAM histograms of sample SCIs before and after naturalization. (a) Two SCIs; (b) original BAM histograms; (c) BAM histograms after applying the proposed naturalization.

where a_{ij} is the weight for a neighboring pixel. Further details for determining the coefficients can be found in [7].

It should be noted that other interpolation (say nearest and bilinear) and image smoothing techniques (e.g. Gaussian low-pass filtering) can also be utilized. However in our experiments, we find bicubic interpolation produces consistently the best results, which will be shown in the experiment section. So we use bicubic interpolation in this work.

In order to demonstrate the effectiveness of the proposed method for naturalizing SCIs, we show an example in Fig. 2. In this example, the BAM histograms are computed for the original and naturalized SCIs. It is easily observed from the figure that the original BAM histograms spread a wide range, which do not fall into either left-side or right-side distribution. After naturalization, the BAM histograms more resemble the left-side distribution, which is a key property of natural images. Therefore, the proposed method is effective for the naturalization of SCIs. In this work, we first apply the proposed naturalization method to process the SCIs. Then the traditional natural image quality metrics are used to evaluate the quality of SCIs.

4. Experimental Results

To verify the effectiveness of the proposed method, we use it as a pre-processing step for SCIs. Then the traditional image quality metrics are tested using the processed SCIs. We employ the SIQAD database [1] to conduct the experiments. SIQAD consists of 20 reference SCIs and the corresponding 980 distorted images. Each reference SCI is subject to 7 distortions at 7 levels. The subjective quality is denoted by the difference mean-opinion-score (DMOS). Two criteria are utilized to measure the metric performances, namely Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank order Correlation Coefficient (SRCC). Before computing them, a five-parameter logistic mapping is conducted between the objective and subjective scores:

$$f(x) = \tau_1\left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2(x-\tau_3)}}\right) + \tau_4 x + \tau_5, \tag{5}$$

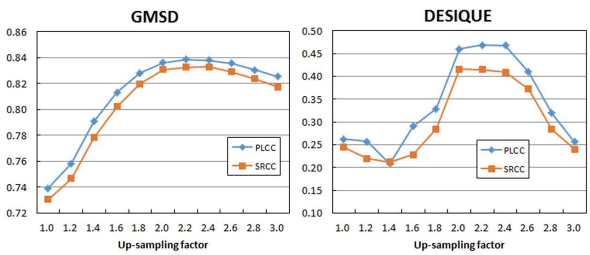


Fig. 3 Impact of up-samping factors on the metric performances.

Table 1 Performances of FR image quality metrics before and after incorporating the proposed naturalization module, together with a statistics of the performance gains in percentage.

Metric	PLCC			SRCC		
	Before	After	Gain (%)	Before	After	Gain (%)
MS-SSIM [8]	0.6195	0.7293	↑ 17.72	0.6112	0.7206	↑ 17.90
IW-SSIM [9]	0.6536	0.8310	↑ 27.14	0.6546	0.8116	↑ 23.98
VIF [10]	0.8488	0.8624	↑ 1.60	0.8451	0.8545	↑ 1.12
MAD [11]	0.6192	0.6520	↑ 5.30	0.6069	0.6415	↑ 5.70
FSIM [12]	0.5902	0.6108	↑ 3.49	0.5819	0.6042	↑ 3.83
GSIM [13]	0.5686	0.5775	↑ 1.57	0.5483	0.5513	↑ 55.00
GMSD [14]	0.7391	0.8381	↑ 13.39	0.7305	0.8329	↑ 14.02
LTG [15]	0.7519	0.8443	↑ 12.29	0.7430	0.8349	↑ 12.36
SFF [16]	0.6928	0.7917	↑ 14.28	0.6870	0.7830	↑ 13.97

where $\tau_1, \tau_2, \cdots, \tau_5$ are the fitting parameters.

Before conducting the experiments, the up-sampling factor of bicubic interpolation is first determined. Figure 3 shows the performances of two state-of-the-art natural image quality metrics (FR GMSD [14] and NR DESIQUE [20]) when different up-sampling factors are used. It is observed from the figure that the best performances are mainly achieved when the up-sampling factors are around 2 to 2.5. In the subsequent experiments, the up-sampling factor is set to 2.4, based on which most quality metrics can achieve very good performances.

4.1 Performances of FR Natural Image Quality Metrics

We first test the performances of the state-of-the-art FR natural image quality metrics, including MS-SSIM [8], IW-SSIM [9], VIF [10], MAD [11], FSIM [12], GSIM [13], GMSD [14], LTG [15] and SFF [16]. Table 1 summarizes the experimental results before and after incorporating the proposed naturalization module, together with a statistics of the performance gains in percentage.

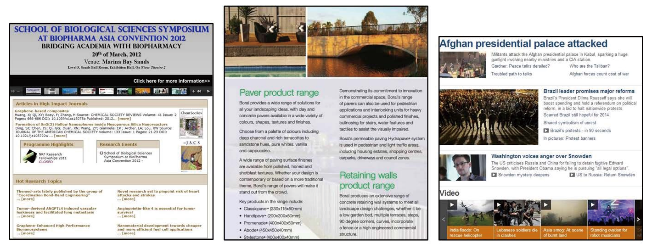
It is observed from Table 1 that the performances of all FR quality metrics are improved after incorporating the proposed naturalization module. Furthermore, many FR metrics achieve performance gains higher than 10% for both PLCC and SRCC, which are not trivial. Therefore, the proposed method is very effective to adapt the traditional natural image quality metrics to SCI quality assessment.

4.2 Performances of NR Natural Image Quality Metrics

In addition to the FR metrics, we further test the effectiveness method using the state-of-the-art NR natural image quality metrics, including BIQI [17], DIIVINE [18], BRISQUE [19], DESIQUE [20] and NIQE [21]. These metrics have been designed for evaluating image quality without knowing the exact distortion types. The experimental results are listed in Table 2.

Table 2 Performances of NR image quality metrics before and after incorporating the proposed naturalization module, together with a statistics of the performance gains in percentage.

Metric	PLCC			SRCC		
	Before	After	Gain (%)	Before	After	Gain (%)
BIQI [17]	0.0432	0.1682	↑ 289.35	0.0457	0.1700	↑ 271.99
DIVINE [18]	0.2946	0.4760	↑ 61.58	0.2750	0.4067	↑ 47.89
BRISQUE [19]	0.0510	0.3303	↑ 548.14	0.1063	0.2861	↑ 169.09
DESIQUE [20]	0.2621	0.4680	↑ 78.57	0.2443	0.4089	↑ 67.37
NIQE [21]	0.3005	0.5566	↑ 85.22	0.3675	0.5472	↑ 48.90



(a) DMOS=41.9984 (b) DMOS=43.8690 (c) DMOS=42.9456

Fig. 4 Three distorted screen content images. From (a) to (c), the images have increasing picture-to-text ratios but similar visual quality.

Table 3 Predicted scores of NR quality metrics before and after incorporating the proposed naturalization module for images shown in Fig. 3.

Metric	Before			After		
	(a)	(b)	(c)	(a)	(b)	(c)
BIQI [17]	56.8257	41.9546	33.1808	25.9913	28.0189	26.3609
DIVINE [18]	8.71150	27.6436	0.9015	59.2894	42.7398	49.0099
BRISQUE [19]	96.1845	25.5710	55.4717	64.0690	57.8123	58.9319
DESIQUE [20]	49.6898	26.1281	61.7303	38.3342	27.2375	31.1476
NIQE [21]	7.78240	3.9563	4.2130	2.6671	2.2822	2.2356

It is known from Table 2 that all the tested metrics achieve obvious performance improvement, which are much more significant than FR metrics. This further demonstrates the effectiveness of the proposed method.

Figure 4 shows three SCIs with similar quality but different picture-to-text ratios. Table 3 summarizes the predicted scores by the state-of-the-art NR quality metrics before and after integrating the proposed naturalization module. Since the images have similar visual quality, a good metric is also expected to produce similar scores. It is known from the table that the scores predicted by the original metrics differ greatly. After incorporating the proposed method, the scores become more similar. This also demonstrates that the proposed method is not sensitive to picture-to-text ratios, so it can be used for SCI quality assessment across images.

4.3 Evaluation of Other Approaches with Smoothing Effect

In image processing, there are other techniques that can pro-

Table 4 Performance comparisons of different approaches with smoothing effect.

Metric	Nearest interpolation				Bilinear interpolation				Gaussian low-pass filtering (3×3, 0.2)			
	PLCC	Gain (%)	SRCC	Gain (%)	PLCC	Gain (%)	SRCC	Gain (%)	PLCC	Gain (%)	SRCC	Gain (%)
MS-SSIM [8]	0.7246	↑ 16.97	0.7182	↑ 17.51	0.7139	↑ 15.23	0.7060	↑ 15.51	0.6192	↓ 0.05	0.6120	↑ 0.13
MAD [11]	0.7078	↑ 14.31	0.6968	↑ 14.81	0.6520	↑ 5.29	0.6446	↑ 6.21	0.6172	↓ 0.32	0.6070	↑ 0.02
FSIM [12]	0.6168	↑ 4.51	0.6109	↑ 4.98	0.6071	↑ 2.86	0.5979	↑ 2.75	0.5923	↑ 0.36	0.5824	↑ 0.09
GMSD [14]	0.8329	↑ 21.18	0.8295	↑ 21.01	0.8232	↑ 19.76	0.8213	↑ 19.8	0.7388	↓ 0.04	0.7305	0
BIQI [17]	0.1928	↑ 346.3	0.1832	↑ 300.88	0.1097	↑ 153.93	0.1242	↑ 171.87	0.0987	↑ 56.23	0.0898	↑ 49.11
DIVINE [18]	0.4290	↑ 45.62	0.3551	↑ 29.13	0.5564	↑ 88.87	0.4817	↑ 75.17	0.2946	0	0.2751	↑ 0.03
BRISQUE [19]	0.2042	↑ 300.66	0.1334	↑ 25.48	0.2805	↑ 450.44	0.2998	↑ 181.99	0.2480	↑ 386.60	0.2120	↑ 49.85
NIQE [21]	0.3991	↑ 32.81	0.3820	↑ 3.95	0.5016	↑ 66.92	0.4423	↑ 20.35	0.3319	↑ 10.45	0.3702	↑ 0.74

duce smoothing effect, among which is the most popular Gaussian low-pass filtering. Here, we also test the performance of the proposed method using Gaussian low-pass filtering. Further, nearest and bicubic interpolations are also included for comparison. The results are listed in Table 4.

It is known from Table 4 that in most cases these smoothing approaches are also effective for image naturalization. However, interpolation is advantageous over Gaussian filtering in this task, which can be seen from the results. Further, by comparing Tables 1, 2 and 4, bicubic interpolation is more effective than nearest and bicubic interpolations. So we adopt bicubic interpolation in this work.

5. Conclusion

In this letter, we have addressed the problem of screen content image quality assessment. Specifically, we have proposed a naturalization method for processing screen content images, with the aim to make SCIs more similar to natural images and the existing natural image quality metrics can be applied. We have demonstrated the effectiveness of the proposed method based on both FR and NR image quality metrics in a public screen content image quality database.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities (2015XKMS032) and the Qing Lan Project of Jiangsu Province.

References

[1] 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.

[2] W. Lin and C.-C.J. Kuo, "Perceptual visual quality metrics: a survey," J. Vis. Commun. Image Represent., vol.22, no.4. pp.297–312, May 2011.

[3] 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, June 2016.

[4] S.Q. Wang, K. Gu, K. Zeng, Z. Wang, and W.S. Lin, "Objective quality assessment and perceptual compression of screen content images," IEEE Comput. Graph., DOI: 10.1109/MCG.2016.46, 2016.

[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] H. Yang, W.S. Lin, and C.W. Deng, "Image activity measure (IAM) for screen image segmentation," Proc. IEEE Int. Conf. Image Process. (ICIP), pp.1569–1572, 2012.

[7] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, Numerical recipes in C++: The art of scientific computing, 2nd ed., Cambridge University Press, New York, NY, 2002.

[8] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multi-scale structural similarity for image quality assessment," in Proc. IEEE Asilomar Conference on Signals, Systems and Computers, pp.1398–1402, 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] 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.

[11] E.C. Larson and D.M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," J. Electronic Imaging, vol.19, no.1, Article ID 001006, pp.1–21, March 2010.

[12] 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.

[13] 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, April 2012.

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

[15] 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. (ICIP), pp.506–510, 2014.

[16] H.-W. Chang, H. Yang, Y. Gan, and M.-H. Wang, "Sparse feature fidelity for perceptual image quality assessment," IEEE Trans. Image Process., vol.22, no.10, pp.4007–4018, Oct. 2013.

[17] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," IEEE Signal Process. Lett.,, vol.17, no.5, pp.513–516, May 2010.

[18] A.K. Moorthy and A.C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," IEEE Trans. Image Process., vol.20, no.12, pp.3350–3364, Dec. 2011.

[19] A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process., vol.21, no.12, pp.4695–4708, Dec. 2012.

[20] Y. Zhang and D.M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," J. Electronic Imaging, vol.22, no.4, Article ID 043025, pp.1–22, 2013.

[21] 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, March 2013.