Towards efficient blind quality evaluation of screen content images based on edge-preserving filter

Jiansheng Qian✉, Lijuan Tang, Vinit Jakhetiya, Zhifang Xia, Ke Gu and Hong Lu

The problem of quality assessment of screen content images from two respects is addressed, the edge-preserving filter based free energy and structural degradation model. For screen content images which always contain texts, edge information plays an important role in the process of evaluating the screen images quality. Inspired by this, the edge-preserving filter based free energy entropy and structural degradation model were combined to extract the quality features and accordingly the statistical model of screen content images was established. Experimental results prove that the proposed method can produce highly consistent with human perception, particularly superior to state-of-the-art full- and no-reference quality metrics on the SIQAD database dedicated to the quality assessment of screen images.

Introduction: With the soaring development of electronic business and network, objective image quality assessment (IQA) issue becomes increasingly important for both the practical application and scientific research in digital image processing systems. Objective IQA can be divided into three types, full reference (FR), reduced reference (RR) and no reference (NR). FR IQA metrics evaluate the quality of images by comparing the distorted image with the original image, some popular FR IQA metrics have been devised in [1–4]. RR IQA metrics utilise partial original image information. However, in most practical applications, the original image is not available. NR IQA metrics have been intensively studied in [5–8]. In this Letter, we focus our attention on NR IQA metric for the degradation measurement of screen content images.

Screen content image quality assessment and compression have been recently emerged as an active research topic due to the rapid development of Internet technology and cloud computing. Generally, screen content images are composed of natural images and computer generated graphical/textual content. It is hard to design NR IQA model for screen content images. However, it is important to evaluate screen content images due to its good capability for optimisation in different cloud and remote computing processing systems.

One direct strategy to solve the quality evaluation of screen content images is to distinguish pictorial and textual regions, before integrating the scores of the two parts to infer the overall quality scores, such as [1]. Another strategy is to combine visual saliency to solve the quality assessment of screen content images, such as [4]. Very recently, a blind quality evaluation method of screen content images was developed upon perceptual features [8]. However, these methods from effectiveness and efficiency two aspects are not able to produce satisfactory results. In this Letter, we propose an effective and fast NR IQA model for assessing the quality of screen content images.

Proposed blind quality measure: We exploit the edge-preserving filter based free energy entropy [9] and structural degradation measure [10] to extract effective perception features in our proposed method. As for screen content images that always contain texts, edge information plays an important role in the process of assessing the screen images quality, while the exiting methods have not considered it [8]. A new bilateral filter was defined as [9]

$$\hat{I} = \frac{I_{\mathrm{Obs}} + \lambda \sum_{K=1}^{N} \omega_k I_k}{1 + \lambda \sum_{K=1}^{N} \omega_k} \tag{1}$$

where $\hat{I}$ and I_{Obs} are the prediction and the observed values of the pixel of interest I, respectively. I_k are the neighbouring pixels in the window used in prediction. λ controls the trade-off between the likelihood and the prior. ω_k are the weights. To generate sharp edges, kernel ridge regression based least squares optimisation is utilised to approximate the I as I_{Obs} [9]. The kind of edge-preserving filter is very helpful to describe the properties of screen content images. Thus, we integrate the edge-preserving filter based prior information and likelihood into our proposed method [9]. The free energy theory lies in the assumption that there always exists a gap between an input visual signal and its processed one by human brain [11]. Moreover, an internal generative model manipulates this process, which can infer predictions of the input visual signal and avoid the residual uncertainty information. Therefore, the psychovisual quality of a scene is defined by both the scene itself and the output of the internal generative model. We define an input screen content image and the image after processed by the edge-preserving filter version as I and $\hat{I}$. The free energy of I can be approximated as the entropy of residual uncertainty information:

$$E(I) = -\sum_i P_i(\bar{I}) \log P_i(\bar{I}) \tag{2}$$

where $\bar{I} = I - \hat{I}$ is the error between the input image and its predicted one, $P_i(\bar{I})$ is the probability density. According to the analysis in [10], different types of distorted images appear distinct degrees of spatial frequency reduction after the lowpass filtering. Hence, the structural degradation information can be defined by

$$S_\mu(I) = A\left(\frac{\eta_{(\hat{\mu}_I \check{\mu}_I)+\kappa}}{\eta_{\hat{\mu}_I}\eta_{\check{\mu}_I} + \kappa}\right) \tag{3}$$

$$S_\sigma(I) = A\left(\frac{\eta_{(\hat{\sigma}_I \check{\sigma}_I)+\kappa}}{\eta_{\hat{\sigma}_I}\eta_{\check{\sigma}_I} + \kappa}\right) \tag{4}$$

where $\hat{\mu}_I$ and $\hat{\sigma}_I$ denote the local mean and variance of I defined by Gaussian kernel, respectively, and $\check{\mu}_I$ and $\check{\sigma}_I$ are defined by the impulse function, $A(\cdot)$ is the global average function, $\eta_{(\hat{\mu}_I \check{\mu}_I)}$ denotes the local convariance, κ is a small constant to avoid the denominator becomes zero [10].

Next, the three couples of (U, V), which are set as (1, 1), (3, 3) and (5, 5), for the normalised Gaussian kernel $\boldsymbol{w} = \{w(u, v)|u = -U, \ldots, U, v = -V, \ldots, V\}$ are used to describe a different amount of neighbouring information. To acquire highly consistent with human perception, $S_\mu(I)$ and $S_\sigma(I)$ are reversed when $E(I)$ is larger than the threshold value T. Finally, $S_\mu(I)$ and $S_\sigma(I)$ are calculated in the inner 6×6 part and outer block edge part to measure the degree of frequency change caused by JPEG compression.

To unearth the relationship between the free energy and structure degradation features, we select 1000 high quality images from the Internet to verify it. We find that there exists a linear dependence between the original image's free energy and structure degradation features. The linear regression model is defined as

$$E(I_0) = a_\lambda \cdot S_{\mu_\lambda}(I_0) + b_\lambda \tag{5}$$

$$E(I_0) = c_\lambda \cdot S_{\sigma_\lambda}(I_0) + d_\lambda \tag{6}$$

where I_0 is the original image, a_λ, b_λ, c_λ, d_λ with $\lambda = \{i_1, i_2, i_3, o_1, o_2, o_3\}$ are attained by the least square method and the values are listed in Table 1. This linear dependence can be used to evaluate the quality of the distorted screen content image without the reference image. For the values $E(I_0) - a_\lambda \cdot S_{\mu_\lambda}(I_0) + b_\lambda$ and $E(I_0) - c_\lambda \cdot S_{\sigma_\lambda}(I_0) + d_\lambda$ of high-quality images without distortion are close to zero, while the values of corrupted images become far away from zero.

Table 1: Parameters a_λ, b_λ, c_λ and d_λ for S_{μ_λ} and S_{σ_λ} are obtained by the least square method

	a_λ	b_λ		c_λ	d_λ
$S_{\mu_{i1}}$	−5.7843	5.7674	$S_{\sigma_{i1}}$	−3.6281	3.4906
$S_{\mu_{i3}}$	−2.3789	2.3077	$S_{\sigma_{i3}}$	−5.5892	5.8585
$S_{\mu_{i5}}$	−6.0227	5.9080	$S_{\sigma_{i5}}$	−2.4260	2.3898
$S_{\mu_{o1}}$	−5.9242	5.8749	$S_{\sigma_{o1}}$	−3.5363	3.4305
$S_{\mu_{o3}}$	−2.3820	2.3105	$S_{\sigma_{o3}}$	−5.2679	5.5900
$S_{\mu_{o5}}$	−6.0495	5.9285	$S_{\sigma_{o5}}$	−2.4399	2.3965

After feature extraction, a proper way is needed to map the feature space to subjective quality scores. To avoid the problem of over-fitting, we adopted the method used in [8] to train the model. To specify, SQMS method [4] is utilised to produce subjective ratings, which was demonstrated validly for screen content IQA. Then we use a support vector regression [12] to generate a proper mapping.

Experimental results and analysis: To validate the effectiveness of the proposed metric, we compared our model with state-of-the-art IQA metrics, feature similarity index for image (FSIM) [2], visual saliency-induced index (VSI) [3], saliency-guided quality measure of SCIs (SQMS) [4], natural image quality evaluator (NIQE) [5], blind/

referenceless image spatial quality evaluator (BRISQUE) [6], NR Free Energy based Robust Metric (NFERM) [7], blind quality measure for SCIs (BQMS) [8] on the recently released screen image quality assessment database (SIQAD) [1]. In this Letter, we follow the video quality experts group's suggestion and employ a five-parameter non-linear fitting function to map objective quality scores to subjective human ratings first. Spearman's rank ordered correlation coefficient (SRCC) and Kendall's rank correlation coefficient (KRCC) criteria are used to evaluate the prediction monotonicity. And Pearson's correlation coefficient (PLCC) and RMS error (RMSE) are selected for the prediction accuracy. An excellent IQA model is expected to produce high SRCC, KRCC and PLCC values, and low RMSE values.

We validate the performance of our proposed model in two manners. The first is to compare our proposed model with state-of-the-art IQA methods based on the model from a large number of 100,000 distorted screen images and the corresponding scores produced by the SQMS method. The result is listed in Table 2. It can be seen that the proposed method obtains the optimal performance. The second evaluation is to compare our method with NR IQA metrics based on the typical training process. The images in the SIQAD database are randomly divided into two parts, 80% for training and 20% for testing. This process repeats 1000 times and reports the median results. Table 3 lists the result. As expected, the proposed method acquires the best performance.

Table 2: Performance of our proposed metric and state-of-art FR IQA, NR IQA metrics. We highlight the top two metrics with boldface

Metrics	Type	PLCC	SRCC	KRCC	RMSE
FSIM [2]	FR	0.5906	0.5824	0.4253	11.551
VSI [3]	FR	0.5568	0.5381	0.3874	11.890
SQMS [4]	FR	**0.8872**	**0.8803**	**0.6936**	**6.6039**
NIQE [5]	NR	0.3758	0.3742	0.2543	13.265
Proposed metric	NR	**0.7704**	**0.7406**	**0.5451**	**9.1258**

Table 3: Performance of our proposed metric and state-of-art blind metrics on SIQAD database. We bold the best performed metric

Metrics	PLCC	SRCC	KRCC
BRISQUE [6]	0.7708	0.7237	0.5382
NFERM [7]	0.8007	0.7717	0.5815
BQMS [8]	0.8115	0.8005	0.6056
Proposed metric	**0.8187**	**0.8093**	**0.6258**

To directly measure the correlation performance, Fig. 1 shows the scatter plots of subjective DMOS (differential mean opinion score) values and objective quality predictions of state-of-art FR FSIM [2], FR VSI [3], NR NIQE [5] and our proposed method. Obviously, our method has acquired the impressive convergency and monotonicity, much better than recently designed FSIM, VSI and NIQE metrics.

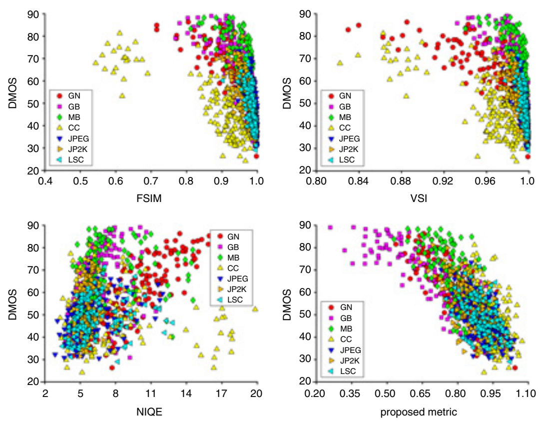

Fig. 1 *Scatter plots of DMOS versus FSIM [2], VSI [3] and NIQE [5] scores on the SIQAD database [1]*

In many screen content images practical applications, such as remote computing, cloud gaming and cloud-guided enhancement, it is desired to evaluate the quality of a screen content image in real time. Hence, we compare the computational complexity of our proposed algorithm and the recently developed BQMS model for blindly assessing screen content images [4]. Experiments are conducted using the laptop with Intel(R) core(TM) i5-4200U CPU 1.60 GHZ and 4 GB RAM. The feature extraction time consumed by the proposed method and BQMS model when applied to a screen content image of size 672 × 682 are 1.67s and 95.21s.

Conclusion: We have devised a new NR IQA metric for screen content images. A comparison of our proposed method with state-of-the-art FR IQA and NR IQA methods is conducted on the SIQAD database. The experiment results have proved the superior performance of the proposed blind quality method for screen content images. Apart from the substantially high prediction accuracy, it is worthy to emphasise two spotlights; i.e. the proposed method has very good generalisation ability, and it is more efficient than existing popular screen content images quality metrics.

Submitted: *25 January 2017* E-first: *30 March 2017*
doi: 10.1049/el.2017.0325
One or more of the Figures in this Letter are available in colour online.

Jiansheng Qian and Lijuan Tang (*China University of Mining and Technology, Xuzhou, Jiangsu 221116, People's Republic of China*)

✉ E-mail: qianzhangiqa@163.com

Vinit Jakhetiya (*Hong Kong University of Science and Technology, Hong Kong, Hong Kong*)

Zhifang Xia (*State Information Center of P.R.China, Beijing 100084, People's Republic of China*)

Ke Gu (*Nanyang Technological University, Singapore 639798, Singapore*)

Hong Lu (*Nanjing Institute of Technology, Nanjing 211167, People's Republic of China*)

Lijuan Tang: Also with Jiangsu Vocational College of Business, Nantong, Jiangsu, People's Republic of China

References

1 Yang, H., Yuming, F., and Weisi, L.: 'Perceptual quality assessment of screen content images', *IEEE Trans. Image Process.*, 2015, **24**, pp. 4408–4421
2 Zhang, L., Zhang, L., Mou, X., and Zhang, D.: 'FSIM: a feature similarity index for image quality assessment', *IEEE Trans. Image Process.*, 2011, **20**, pp. 2378–2386
3 Zhang, L., Shen, Y., and Li, H.: 'VSI: A visual saliency induced index for perceptual image quality assessment', *IEEE Trans. Image Process.*, 2014, **23**, pp. 4270–4281
4 Gu, K., Wang, S., Yang, H., *et al.*: 'Saliency-guided quality assessment of screen content images', *Trans. Multimedia*, 2016, **18**, pp. 1098–1110
5 Mittal, A., Soundararajan, R., and Bovik, A.C.: 'Making a 'completely blind' image quality analyzer', *IEEE SPL*, 2013, **22**, pp. 209–212
6 Mittal, A., Moorthy, A.K., and Bovik, A.C.: 'No-reference image quality assessment in the spatial domain', *IEEE Trans. Image Process.*, 2012, **21**, pp. 4695–4708
7 Gu, K., Zhai, G., Yang, X., and Zhang, W.: 'Using free energy principle for blind image quality assessment', *IEEE Trans. Multimedia*, 2015, **17**, pp. 50–63
8 Gu, K., Zhai, G., Lin, W., Yang, X., and Zhang, W.: 'Learning a blind quality evaluation engine of screen content images', *Neurocomputing*, 2016, **196**, pp. 140–149
9 Jakhetiya, V., Au, O.C., Jaiswal, S., Jia, L., and Zhang, H.: 'Fast and efficient intra-frame deinterlacing using observation model based bilateral filter'. ICASSP, Florence, Italy, May 2014, pp. 5819–5823
10 Gu, K., Zhai, G., Yang, X., and Zhang, W.: 'A new reduced-reference image quality assessment using structural degradation model'. Proc. IEEE Int. Symp. Circuits and Systems, Beijing, China, May 2013, pp. 1095–1098
11 Friston, K.: 'The free-energy principle: a unified brain theory?', *Nat. Rev. Neurosci.*, 2010, **11**, pp. 127–138
12 Gu, K., Tao, D., Qiao, J.-F., *et al.*: 'Learning a no-reference quality assessment model of enhanced images with big data', *IEEE Trans. Neural Netw. Learning Syst.*, 2017, **PP**, (99), pp. 1–13, doi: 10.1109/TNNLS.2017.2649101