Screen Image Quality Assessment Incorporating Structural Degradation Measurement

Ke Gu[†‡§], *Shiqi Wang*[‡], *Guangtao Zhai*[†], *Siwei Ma*[♯], *and Weisi Lin*[§]
[†]Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong Univ., Shanghai, China, 200240
[‡]Dept. of Electrical and Computer Engineering, Univ. of Waterloo, Waterloo, ON, N2L 3G1, Canada
[♯]School of Electronic Engineering & Computer Science, Peking Univ., Beijing, China, 100871
[§]School of Computer Engineering, Nanyang Technological Univ., Singapore, 639798

***Abstract*—Screen content is typically composed of computer generated text and graphics. The contents shown on the screen exhibit various unnatural properties, such as sharp edges and thin lines with few color variations. In this paper we design a novel structure-induced quality metric (SIQM) for assessing the screen image quality. The proposed SIQM works by weighting the benchmark structural similarity index (SSIM) with the structural degradation measurement that is computed using SSIM as well. Experimental results conducted on the newly released subjective quality database concerning screen images show that on one hand the proposed technique is superior to existing quality measures, and on the other hand our model is able to optimize screen video coding and thus introduce remarkable visual quality improvement.**

***Index Terms*—Screen content, image quality assessment (IQA), structural similarity, structural degradation, screen video coding**

I. Introduction

Recent advances in cloud computing and mobile computing have brought new challenges to the quality assessment community. In many scenarios, e.g. cloud-mobile convergence [1], cloud gaming [2], and remote computing platforms [3], the remote computing is facilitated by users' interaction with the local display interface, which is typically realized by computer generated screen images. Usually, the screen images are generated and compressed at the server, and then transmitted to the thin client side. This process inevitably introduces distortions. The quality of screen images directly determines the interactivity performance and users' experience. Hence accurately predicting the screen quality with an objective model plays a variety of roles in cloud and remote computing applications. First, it can be used to dynamically monitor the screen image quality and adjust resources to improve remote computing experience. Second, it can be used for optimization, e.g. screen content compression to attain better rate distortion performance. Third, it can work as a benchmark in the quality evaluation of the remote computing system.

In general, the screen image is a mixture of computer generated graphical content and natural images. Graphical content generated by computer usually contains textual and graphics. There are quite a lot of differences between screen content and natural image content. Capturing natural video will introduce noise due to the limitations of image sensors. The screen content, however, is completely noise free and can be directly captured by screen recording tools. The camera captured natural content is mostly featured with thick lines, rich color variations and complex texture content, while the computer generated screen content has thin lines, limited colors and usually uncomplicated shapes [4]. During the past decade, a large set of image quality assessment (IQA) metrics have been proposed for natural images corrupted by, e.g. JPEG and JPEG2000 compressions [5]-[10], contrast change [11]-[12], and multiple distortions [13]. However, to our best knowledge, few of them were specifically developed for screen content. In [14]-[15], a data set of distorted screen images and associated subjective quality ranking results have been provided, showing that classical/state-of-the-art IQA methods do not work effectively in predicting the subjective quality of screen images.

In this paper we address the issue based on the following four considerations. First, the classical structural similarity index (SSIM) [5] was found to capture an individual category of artifacts from distorted screen images successfully. Second, the different importance between screen contents (e.g. computer generated text and graphics) and others cannot be well discriminated by a direct average pooling, and this limits the prediction performance of SSIM to a large extent. Third, in all perceptual IQA algorithms, only SSIM has been broadly incorporated into compression tools [16]-[17] to date. Fourth, a constant weighting map (in the pooling stage) for one frame is highly desired in efficient video coding. Hence we design a simple yet valid structure-induced quality metric (SIQM), which weights SSIM with structural degradation measurement computed using SSIM on the source image and its degraded one by a low-pass filter [18]-[19]. The proposed technique is shown to be effective on screen image quality assessment database (SIQAD) and screen video.

The rest of this conference paper is arranged as follows. Section II first analyzes the difficulty of screen image quality assessment before presenting the proposed SIQM. In Section III, a numerical comparison with classical and state-of-the-art relevant quality measures on SIQAD and with the optimization of various IQA metrics for screen video coding validate the superiority and practicability of our approach. We conclude the whole paper in Section IV.

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

II. Proposed Quality Measure

Though MSE has been used for various applications, it was found to poorly correlate with human judgement of quality, i.e. the mean opinion score (MOS) [20]. The most popular solution presently is perhaps SSIM, which has been embedded into a wide range of systems. Given a reference image block $\mathbf{r}$ and its distorted one $\mathbf{d}$, SSIM combines local luminance, contrast and structural similarities that are expressed by

$$l(\mathbf{r},\mathbf{d}) = \frac{2\mu_r\mu_d + c_1}{\mu_r^2 + \mu_d^2 + c_1} \tag{1}$$

$$c(\mathbf{r},\mathbf{d}) = \frac{2\sigma_r\sigma_d + c_2}{\sigma_r^2 + \sigma_d^2 + c_2} \tag{2}$$

$$s(\mathbf{r},\mathbf{d}) = \frac{\sigma_{rd} + c_3}{\sigma_r\sigma_d + c_3} \tag{3}$$

where μ_r and σ_r (μ_d and σ_d) indicate the mean intensity and standard deviation of r (d), σ_{rd} is the covariance between r and d, and c_1 to c_3 are three small fixed values for increasing the stability of numerical calculation. The overall score of SSIM is computed by

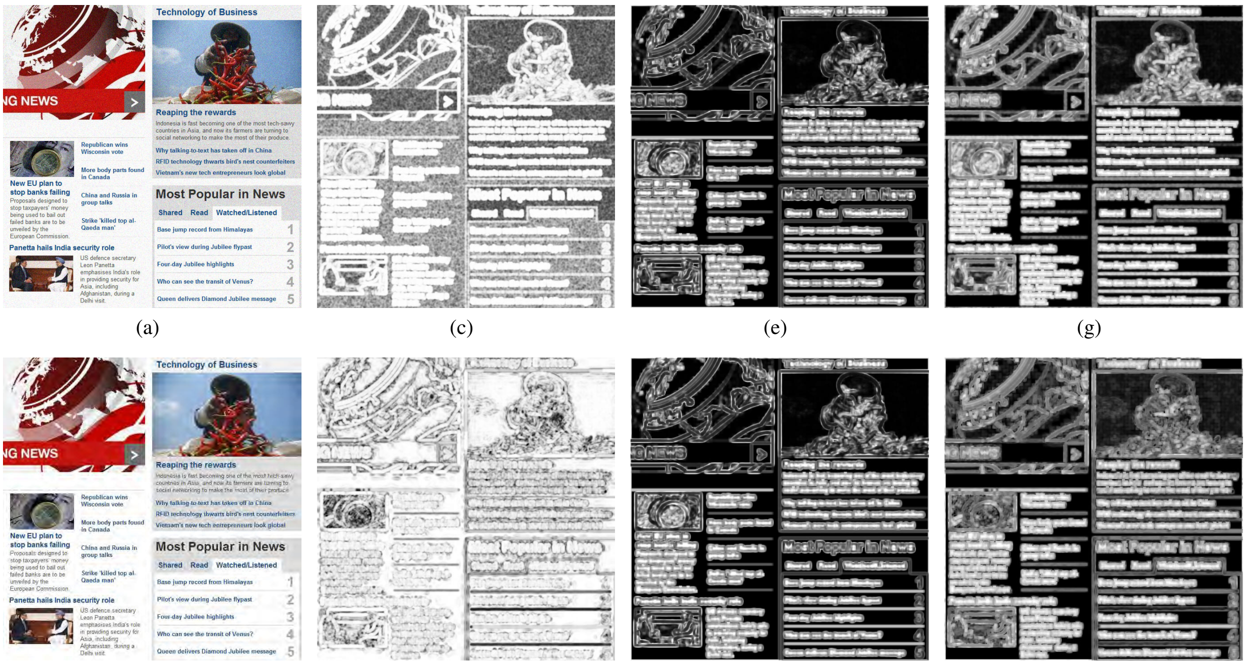
$$\begin{aligned} \text{SSIM}(\mathbf{r},\mathbf{d}) &= \frac{1}{M}\sum_{i=1}^{M} \text{SSIM\_MAP}(r_i, d_i) \\ &= \frac{1}{M}\sum_{i=1}^{M} l(r_i,d_i)\cdot c(r_i,d_i)\cdot s(r_i,d_i) \end{aligned} \tag{4}$$

where r_i and d_i represents the i-th pixel values in $\mathbf{r}$ and $\mathbf{d}$, and M is the number of local windows in the image.

We take two exemplary distorted screen images associated with the same reference in Fig. 1: (a) indicates a noised image and (b) indicates a JPEG2000 compressed one. As given in Figs. 1 (c)-(d), SSIM has detected artifacts accurately. With an average pooling, however, SSIM shows a pair of incorrect quality scores, whose order is completely contrary to the order of subjective quality scores[1]. Under the condition of artifacts detected, the problem is very probably caused by the average pooling, which does not consider the different influences of screen contents, such as computer generated text and graphics, and other contents, such as natural illustrations and photos, on subjective perception. As it is found, the screen contents have a stronger impact than natural contents in terms of visual quality, since screen contents generally show much more significant information than natural contents.

An effective pooling strategy is a straightforward solution. Saliency-based weighting is a good choice [21], but it needs the help of a human-assist step to collect fixations for forming the saliency map. Information content weighting (IW) [22] and nonlinear additive model based saliency weighting (S_NW) [23] are also fine options while they require both the reference and distorted images and this largely decreases their practicabilities. Note that, since only SSIM in all perceptual quality models has been inserted into many image processing systems up to now, the desired pooling (e.g. in efficient video coding) is only using SSIM for reference image/frame.

[1]DMOS is the differential version of MOS and is defined as the MOS value of the reference image minus that of the distorted version. Given an image, the smaller DMOS the better quality.



(a) (c) (e) (g)

(b) (d) (f) (h)

Fig. 1. Two exemplary distorted screen images associated with the same reference one: (a) White noise, DMOS = 46.042, SSIM = 0.7380, SIQM = 0.9738; (b) JPEG2000 compression, DMOS = 69.397, SSIM = 0.8405, SIQM = 0.9523; (c)-(d) SSIM distortion maps of (a)-(b); (e)-(f) Structural degradation maps of (a)-(b); (g)-(h) SSIM distortion maps weighted by structural degradation maps of (a)-(b).

TABLE I
PERFORMANCE CORRELATION OF TESTING IQA MODELS ON SIQAD. WE BOLD THE TOP METRIC ON EACH DISTORTION TYPE.

Metrics	SRC								PCC							
	GN	GB	MB	CC	JPEG	JP2K	LSC	ALL	GN	GB	MB	CC	JPEG	JP2K	LSC	ALL
PSNR	0.886	0.838	0.683	0.686	0.731	0.744	0.720	0.553	0.908	0.841	0.679	0.742	0.730	0.767	0.708	0.579
SSIM	0.874	0.886	0.786	0.654	0.718	0.733	0.663	0.745	0.888	0.887	0.790	0.760	0.710	0.754	0.665	0.750
MS-SSIM	0.875	0.878	0.785	**0.750**	0.754	0.746	0.717	0.597	0.887	0.883	0.790	**0.836**	0.753	0.761	0.718	0.605
LTG	0.873	0.895	0.822	0.611	0.673	0.712	0.661	0.645	0.887	0.895	0.829	0.760	0.668	0.731	0.679	0.656
GSIM	0.843	0.860	0.758	0.736	0.649	0.678	0.634	0.531	0.850	0.864	0.761	0.817	0.646	0.690	0.652	0.552
IGM	0.881	0.867	0.744	0.680	0.758	0.785	0.763	0.624	**0.901**	0.871	0.748	0.805	0.761	0.800	0.757	0.628
GMSM	**0.888**	0.890	0.782	0.652	0.756	**0.814**	**0.780**	0.666	0.898	0.895	0.783	0.773	0.760	**0.824**	**0.778**	0.669
GMSD	0.885	0.902	0.820	0.645	0.741	0.807	0.774	0.717	0.900	0.902	0.820	0.783	0.743	0.820	0.777	0.728
SIQM	0.871	**0.910**	**0.840**	0.705	**0.775**	0.777	0.725	**0.845**	0.892	**0.912**	**0.845**	0.790	**0.771**	0.794	0.720	**0.852**

In this paper we therefore find a good tradeoff between the practical value and quality prediction performance by simplifying the recently proposed structural degradation model in [18]-[19] to derive an effective pooling. More precisely, given the reference image signal $\mathbf{r}$, the structural degradation measurement (SDM) is defined as

$$\text{SDM}(\mathbf{r}) = 1 - \text{SSIM}(\mathbf{r}, \mathbf{r}_f) \tag{5}$$

where $\mathbf{r}_f$ is generated applying a low-pass filter to $\mathbf{r}$. A simple circular-symmetric Gaussian weighting function is used in this implementation:

$$\mathbf{r}_f = \frac{1}{N}\sum_{i=1}^{N} g_i r_i \tag{6}$$

where $\mathbf{g} = \{g_i | i = 1, 2, \ldots, N\}$ has the standard deviation of 2.5 and is normalized to unit sum ($\sum_{i=1}^{N} g_i = 1$). As a result, the proposed SIQM is defined by

$$\text{SIQM}(\mathbf{r}, \mathbf{d}) = \frac{\sum_{i=1}^{M} \text{SSIM\_MAP}(r_i, d_i) \cdot \text{SDM}(r_i)}{\sum_{i=1}^{M} \text{SDM}(r_i)}. \tag{7}$$

We present the weighting maps computed by the structural degradation measurement in Figs. 1(e)-(f). It is easy to find that the important regions, such as those around the text, are highlighted. Figs. 1(g)-(h) show the results of SSIM distortion maps weighted by structural degradation maps. The yielded SIQM scores provide the same order with those of subjective quality ratings, which illustrates the accuracy of the proposed pooling method. More detailed comparisons will be given in the next section.

III. RESULTS OF EXPERIMENTS

A. Screen Quality Assessment

A comparison of the proposed SIQM with classical PSNR, SSIM [5], and MS-SSIM [6], as well as state-of-the-art LTG [7], GSIM [8], IGM [9], GMSM [10], and GMSD [10] is conducted using the newly proposed SIQAD database [14]-[15]. This database is composed of 980 screen images created by corrupting 20 source images with 7 distortion types at 7 distortion levels. Those distortion types include Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG and JPEG2000 (JP2K) compressions, and Layer Segmentation based Coding (LSC).

According to the suggestion of video quality experts group (VQEG) [24], we first employ a five-parameter nonlinear regression to map the objective quality predictions to subjective scores:

$$q(s) = \beta_1\left(\frac{1}{2} - \frac{1}{1+\exp\left(\beta_2 \cdot (s - \beta_3)\right)}\right) + \beta_4 \cdot s + \beta_5 \tag{8}$$

where s and $q(s)$ mean the input and the mapped scores, and β_j $(j = 1, 2, 3, 4, 5)$ are free parameters to be determined during the curve fitting process. We then adopt two frequently used performance evaluations, Spearman rank order correlation coefficient (SRC) for prediction monotonicity and Pearson linear correlation coefficient (PCC) for prediction accuracy, to measure and compare our metric with the competing quality measures.

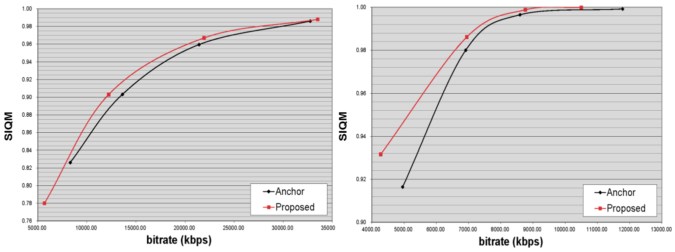
We report the performance correlations of all testing IQA models in Table I, and bold the top metric on each distortion type for helping readers to compare. On the overall database, the proposed SIQM has acquired a substantially high performance, and it is the only metric with the correlation coefficient of beyond 0.8. As compared to other quality measures, the performance gain of our technique is about 13.4% than the second-performer SSIM and 17.9% than the third GMSD. Furthmore, we observe that the proposed algorithm has the best correlation with the subjective perception to image quality on Gaussain and motion blurred images and JPEG compressed images. This implies that our algorithm is good at assessing screen video sequences, which are mostly corrupted by blur and blockiness artifacts.

B. Applications to Screen Content Coding

In this subsection, we further demonstrate that the potential application scope of the proposed quality assessment algorithm exceeds as a benchmark for quality prediction. In particular, it can be embedded into the video coding framework to optimize the whole compression process.

The main object of video coding is to minimize the perceptual distortion D of the reconstructed video with the number of used bits R subjected to a constraint R_c. This can be expressed as follows,

$$\min\{D\} \quad \text{subject to} \quad R \leq R_c. \tag{9}$$



(a) (b)

Fig. 2. R-SIQM curves for two screen content sequences. (a) Map; (b) PPT_Doc_Xls.

In practice, this constrained optimization problem can be solved by Lagrangian optimization technique,

$$\min\{J\} \quad \text{with} \quad J = D + \lambda \cdot R \tag{10}$$

where J is called the rate-distortion (RD) cost and λ is known as the Lagrange multiplier.

Since the ultimate receiver of screen images is the human visual system (HVS), the correct optimization goal should be perceptual quality. Based on the proposed SIQM and our previous work on perceptual video coding [17], the quality metric that is used for optimizing the compression quality of screen content is defined as the combination of the divisive normalization factor from SSIM and the structural degradation measurement. Specifically, the quality measure for pixel r_i is defined as follows,

$$D = \frac{(r_i - d_i)^2}{f^2} \cdot \text{SDM}(r_i) \tag{11}$$

where the normalization factor f is derived from the energy of the AC coefficients within a local window [17].

We perform the new screen coding optimization in the HEVC Range Extension codec for screen content coding (HM15.0+RExt-8.0+SCM-2.0rc1) [25]. The test sequences are in YUV4:4:4 and advanced coding techniques such as intra copy and palette mode used for compression. Intra only cases are tested to verify the proposed algorithm. The R-D curves in terms of the proposed metric are demonstrated in Fig. 2. It is observed that significant bit rate reduction is achieved for both low and high bit rate coding. This further illustrates the effectiveness of our SIQM in the codec optimization.

IV. Conclusion

In this paper we have investigated the problem of quality assessment of screen images as well as the application to screen video coding. The proposed SIQM works by weighting the classical SSIM with a very simple structural degradation measurement that is also computed by SSIM on the reference image only. Results of experiments using the SIQAD database confirm the superiority of our technique as compared to relevant classical and state-of-the-art quality measures. Furthermore, SIQM is also applicable to the existing screen content codec to provide more bandwidth savings.

Acknowledgment

This work was supported in part by NSFC (61025005, 61371146, 61221001, 61390514), FANEDD (201339) and SHEITC (140310).

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.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, "GamingAnywhere: An open cloud gaming system," *in ACM MSC*, pp. 36-47, 2013.
[3] H. Shen, Y. Lu, F. Wu, and S. Li, "A high-performanance remote computing platform," *in ICPCC*, pp. 1-6, Mar. 2009.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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, Apr. 2012.
[9] 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.
[10] 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.
[11] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, 2015.
[12] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybernetics*, 2015.
[13] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcasting*, vol. 60, no. 3, pp. 555-567, Sept. 2014.
[14] H. Yang, W. Lin, C. Deng, and L. Xu, "Study on subjective quality assessment of digital compound images," *in ISCAS*, pp. 2149-2152, Jun. 2014.
[15] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," *in QoMEX*, Sept. 2014.
[16] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516-529, Apr. 2012.
[17] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418-1429, Apr. 2013.
[18] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," *in Proc. IEEE Int. Symp. Circuits and Systems*, pp. 1095-1098, May 2013.
[19] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50-63, Jan. 2015.
[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] X. Min, G. Zhai, Z. Gao, and K. Gu, "Visual attention data for image quality assessment databases," *in ISCAS*, pp. 894-897, Jun. 2014.
[22] 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.
[23] 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 Workshop on Multimedia Signal Process.*, pp. 313-318, 2012.
[24] 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/.
[25] HM15.0+RExt-8.0+SCM-2.0rc1, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-15.0+RExt-8.0+SCM-2.0rc1