SUBJECTIVE QUALITY ASSESSMENT OF SCREEN CONTENT IMAGES

Huan Yang[1], *Yuming Fang*[2], *Weisi Lin*[1], *Zhou Wang*[3]

[1]School of Computer Engineering, Nanyang Technological University, 639798, Singapore.
[2]School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032, China.
[3]Department of Electrical and Computer Engineering, University of Waterloo, N2L3G1 Canada.
Emails: {hyang3, fa0001ng, wslin}@ntu.edu.sg; z.wang@ece.uwaterloo.ca

ABSTRACT

Research on *Screen Content Images* (SCIs) becomes important as they are increasingly used in multi-device communication applications. In this paper, we present a study of subjective quality assessment for distorted SCIs, and investigate which part (text or picture) contributes more to the overall visual quality. We construct a large-scale *Screen Image Quality Assessment Database* (*SIQAD*) consisting of 20 source and 980 distorted SCIs. The 11-category *Absolute Category Rating* (ACR) is employed to obtain three subjective quality scores corresponding to the entire image, textual and pictorial regions respectively. Based on the subjective data, we investigate the applicability of 12 state-of-the-art *Image Quality Assessment* (IQA) methods for objectively assessing the quality of SCIs. The results indicate that existing IQA methods are limited in predicting human quality judgement of SCIs. Moreover, we propose a prediction model to account for the correlation between the subjective scores of textual and pictorial regions and the entire image. The current results make an initial move towards objective quality assessment of SCIs.

1. INTRODUCTION

Inspired by various Internet-based applications [1–3], such as virtual screen sharing, cloud computing and gaming, video conferencing, etc., an increasing amount of visual content is shared between different digital devices (computers, tablets or smart phones). In these applications, visual content (e.g., web pages, slide files and computer screens) is typically presented in the form of *Screen Content Images* (SCIs), which render texts, graphics and natural pictures together. For efficient sharing among different devices, it is important to efficiently acquire, compress, store or transmit SCIs. Numerous solutions have been proposed for processing SCIs, especially for SCI compression [4–8]. Lately, MPEG/VCEG calls for proposals to efficiently compress screen content image/videos as an extension of the HEVC standard [9].

When processing SCIs, various distortions may be involved, such as blurring and compression artifacts. Generally, *Peak Signal-to-Noise Ratio* (PSNR) is adopted to evaluate the quality of the processed images. However, it is know that PSNR is not consistent with human visual perception [10–12]. Although other many IQA methods have been proposed to evaluate quality of distorted natural images [13], whether these IQA methods are applicable to distorted SCIs is still an open question, since SCIs are a specific type of images including texts and pictures concurrently. In real applications, specified objective metrics are more desired to predict quality of processed SCIs, based on which we can control the processing of SCIs more efficiently. Before using the objective metrics, we need to verify whether these metrics are consistent with human visual perception when judging SCI quality. Hence, it is meaningful to investigate both subjective and objective methods in the quality evaluation of distorted SCIs. To the best of our knowledge, this has not yet been carefully studied in the literature.

In this work, we aim to carry out the first in-depth study on subjective quality assessment of SCIs by building a large-scale *Screen Image Quality Assessment Database* (*SIQAD*). Based on the user study on this database, we propose a prediction model to investigate the impact of textual and pictorial regions to the overall image quality. In particular, 20 reference images are selected from the Internet with various content styles, and 980 distorted images are generated from seven distortion processes at seven degradation levels: *Gaussian Noising* (GN), *Gaussian Blurring* (GB), *Contrast Change* (CC), JPEG, JPEG2000 and *Layer Segmentation based Compression* (LSC) [7]. The 11-category *Absolute Category Rating* (ACR) method [14] is adopted to obtain the subjective quality scores of images in *SIQAD*. Three subjective quality scores are obtained for the entire, textual and pictorial regions of each image. Based on these scores, a prediction model is constructed to account for the correlation between the three parts. Finally, to investigate the applicability of existing objective IQA metrics, 12 advanced IQA approaches are employed to evaluate the quality of images in *SIQAD*. Through detailed analysis, we found that existing IQA methods are limited in predicting the quality of the distorted images. The results and observations inspire the development of new objective quality assessment models for SCIs.

978-1-4799-6536-6/14/$31.00 ©2014 IEEE

2. THE SCREEN IMAGE QUALITY ASSESSMENT DATABASE (*SIQAD*)

To investigate quality evaluation for SCIs, we construct a large-scale screen content image database (i.e., *SIQAD*) with seven distortion types, each with seven degradation levels. Totally, 20 reference and 980 distorted SCIs are included in the *SIQAD*. Subjective evaluation of these SCIs is then conducted by human subjects.

2.1. Introduction of the *SIQAD*

We select reference SCIs with various layout styles, including different sizes, positions and ways of textual/pictorial region combination. Meanwhile, pictorial or textual regions are also diverse in contents. In total, twenty SCIs are collected from webpages, slides, PDF files and digital magazines through screen snapshot. The reference SCIs are cropped from these twenty images to proper sizes for natively displaying on computer screens in the subjective test that follows.

Seven distortion types which usually appear on SCIs are applied to generate distorted images. *Gaussian Noise* (GN) is often involved in image acquisition, and is included in most existing image quality databases [15, 16]. *Gaussian Blur* (GB) and *Motion Blur* (MB) are also considered due to their common present in practical applications. For example, when capturing SCIs using digital cameras, hand-shake, out-of-focus or object moving would bring blur into images. *Contrast Change* (CC) is also an important item affecting peculiarities of the HVS. Different settings of brightness and contrast of screens will result in various visual experiences of viewers. As compression of SCIs is an crucial issue in most multimedia processing applications, three commonly used compression algorithms are utilized to encode the reference SCIs: JPEG, JPEG2000 and *Layer Segmentation based Coding* (LSC) [7]. The JPEG and JPEG2000 are two widely used methods to encode images, and have been introduced into many quality assessment databases. We include LSC as another codec due to its efficient compression for SCIs.

For all distortion types, seven levels are set to generate images from low to high degradation levels. These distortions are meant to creat a broad range of image impairment types, such as blurring, blocking, structured distortion and misclassification artifacts. The detailed configuration of these algorithms is given in the related supporting files in *SIQAD* [17].

2.2. Subjective Testing Methodology

Subjective testing methodologies for assessing image quality have been recommended by ITU-R BT.500-13 [14], including *Absolute Category Rating* (ACR), double-stimulus impairment scale and paired comparison. In this study, 11-category ACR is employed. Given one image displayed on the screen, a human subject is asked to give one score (from 0 to 10: 0 is the worst, and 10 is the best) on the image quality based on her/his visual perception. This methodology is chosen because the viewing experience of subjects is close to that in practice, where there is no access to the reference images. The subjective tests are performed using identical desktops, each of which has 16 GB RAM and 64-bit Windows operating system. The desktops with calibrated 24-inch LED monitors are placed in a laboratory with normal indoor light.

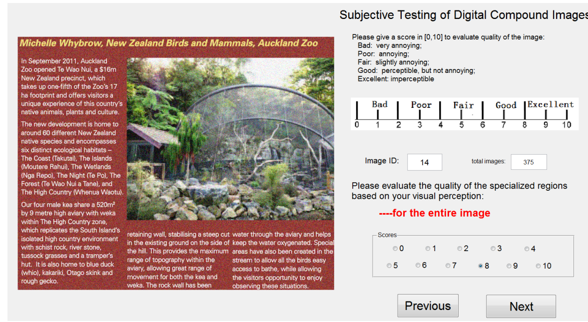


Fig. 1. Graphical user interface in the subjective test. The red tooltip will change if subjects need to judge different regions.

In this study, we would like to investigate which part (text or picture) contributes more to the overall visual quality. Hence, subjects were required to give three scores to each test image, corresponding to overall, textual and pictorial regions, respectively. In this test, each image was shown three times, and subjects gave one score to one specific region at a time. The graphical user interface is shown in Fig.1. When judging one image, three aspects are mainly considered: content recognizability, clearity and viewing comfortability. All the reference images are also included in the test. We generate a random permutation of 1000 images for each round, and make sure that every two consecutive images are not generated from the same reference image. According to [14], the execution time of one test session should not exceed 30 minutes to avoid fatigue. Thus, we split each permutation into 8 groups and assign one group of images to one subject at a time. Each subject finished the evaluation of several groups. Totally, 96 subjects took part in the study, and each image is evaluated by at least 30 subjects.

3. CORRELATION ANALYSIS OF QUALITY SCORES OF DIFFERENT REGIONS

The raw scores given by subjects are used to compute *Difference of Mean Opinion Scores* (DMOS) values of test images [15]. More detailed interpretations of the computation results will be reported in the experimental session. For each test image, we obtain three DMOS values (QE, QT and QP), corresponding to the quality of the entire image, textual and pictorial regions, respectively. The problem we would like to investigate is how the three scores are correlated, or which partial evaluation (QT or QP) contributes more to the overall quality (QE). Through in-depth investigation of this correlation, more efficient objective metrics for assessing quality of SCIs can be carried out. Here we make some initial investigations on the combination method of QT and QP and propose a prediction model QE_p, which is of good correlation with the subjective score QE .

There are many factors affecting human vision when viewing SCIs, including area ratio and region distribution of textual regions, size of characters, and content of pictorial regions, etc. In the proposed model, we investigate a statistical property of SCIs that reflects impairments of test images, rather than any specific factor. Image activity reflects the variation of image contents, which is not only useful in differentiating images, but also important to quality estimation [18, 19]. Based on the activity measure and the segmentation algorithm proposed in [20], we propose a novel model to compute two weights ($\mathcal{W}_t$ and $\mathcal{W}_p$) that can measure the effect of textual and pictorial regions to the quality of the entire image. In particular, given one reference SCI, based on its activity map, the segmentation algorithm can separate textual regions from pictorial regions with an index map I_t in which textual pixels are marked by one and pictorial pixels by zero. Meanwhile, we calculate the activity map A of the distorted SCI. Based on I_t and A, the activity map M_t and M_p for the textual and pictorial regions are obtained. Considering the viewing characteristic of human vision (Points closed to the center are important, and points far away are relatively insignificant), a Gaussian mask G is used to weight the activity values. Based on the weighted activity map, we obtain two activity values for the textual and pictorial parts respectively, which are subsequently employed as weights to combine the quality scores of the two parts.

The prediction model is constructed as a linear combination of QT and QP as follows.

$$QE_p = \mathcal{W}_t * QT + \mathcal{W}_p * QP \tag{1}$$

where

$$\mathcal{W}_t = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(A * I_t * G)_{i,j}}{\sum_{i=1}^{M}\sum_{j=1}^{N}(I_t)_{i,j}} \tag{2}$$

$$\mathcal{W}_p = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(A * (1 - I_t) * G)_{i,j}}{\sum_{i=1}^{M}\sum_{j=1}^{N}(1 - I_t)_{i,j}} \tag{3}$$

are weights for textual and pictorial regions perspectively. M and N represent the sizes of the test image. The performance of the proposed model is assessed by calculating the correlation between the predicted score QE_p and QE.

4. EXPERIMENTAL RESULTS

In this session, we first verify the reliability of the subjective DMOS values, and then test the effectiveness of the proposed prediction model. Finally, 12 existing IQA methods are applied to images in *SIQAD* to investigate whether existing objective quality metrics designed for natural images are applicable to SCIs.

4.1. Reliability of DMOS

When processing the raw subjective scores, we examine the consistency of all subjects' judgements for each image. According to [14], the consistency can be measured by the confidence interval that is derived from the number and standard deviation of scores for each image. Generally, with a probability of 95% confidence level, the distribution of the scores can be regarded as reliable. After outlier rejection, DMOS values of all images are computed and their confidence intervals are obtained. In Fig.2, two examples of DMOS distribution with 95% confidence interval are shown, which demonstrate the agreement of subjects on the visual quality of images. The DMOS values may be further regarded as the ground truth for performance evaluation of objective quality metrics.

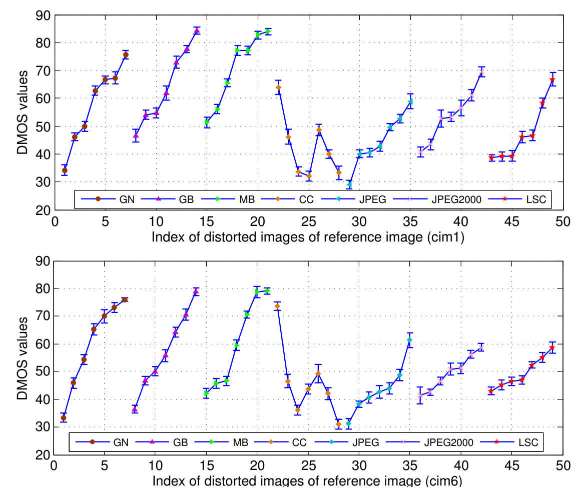


Fig. 2. Distribution of DMOS values of two examples. The error bars indicate the confidence intervals of related scores.

Generally, the quality scales of the distorted SCIs in the database should exhibit good separation of perceptual quality and span the entire range of visual quality (from distortion imperceptible to severely annoying) [21]. Fig.3 shows the histogram of the DMOS values (0:100) of all distorted images in the database. It can be observed that the DMOS values of images range from low to high, and have a good spread at different levels.

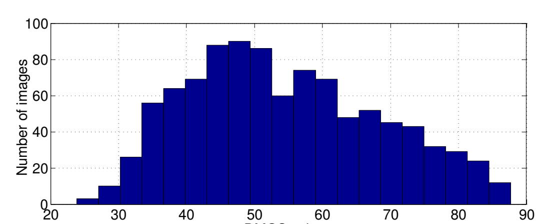


Fig. 3. Histogram of DMOS values of images in the SIQAD.

Table 1. Correlation analysis of the obtained quality scores for the entire images, textual and pictorial regions.

	QE and QT			QE and QP		
Distortions	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC
GN	0.9749	2.7974	0.9571	0.9777	2.7885	0.9393
GB	0.9835	2.3815	0.9571	0.9665	3.0782	0.9482
MB	0.9749	2.1825	0.9475	0.9380	3.1350	0.9032
CC	0.9217	3.8243	0.8446	0.9457	3.0667	0.8746
JPEG	0.9542	2.3801	0.9000	0.8967	3.1158	0.8596
JPEG2000	0.9144	3.1033	0.8625	0.9082	3.1530	0.8589
LSC	0.9187	2.6754	0.8196	0.9002	2.9464	0.8464
Overall	0.9338	4.8067	0.9148	0.8833	6.2471	0.8620

4.2. Verification of the Proposed Prediction Model

Firstly, we analyze the correlations of the obtained three quality scores (QE, QT and QP) in terms of *Pearson Linear Correlation Coefficient* (PLCC), *Root Mean Squared Error* (RMSE) and *Spearman rank-order correlation coefficient* (SROCC) [22]. As such, we can roughly know which part attracts more attention of observers. Meanwhile, correlations for each distortion type are also calculated to estimate human visual perception to different distortion types. The correlation measures are reported in Table 1. From Table 1, we can observe that the textual part has higher overall correlation with the entire image than the pictorial part. However, for different distortion types, the results vary to some extent. For example, in the case of contrast change (CC), the contrast variation of pictorial regions affect human vision more compared to that of textual regions. The reason may be that, observers prefer to give high scores to texts of high shape integrity and clearity, even though their colors change significantly. For pictorial regions, severe contrast change would result in uncomfortable viewing experience. Therefore, in this case, pictorial regions contribute more to the quality of the entire image. By contrast, in the case of motion blurring (MB), textual regions attract more attention. The integrity and clearity of texts are easier to be affected by motion blurring. For other distortions, the correlation results also vary from case to case. Consequently, it is a challenging problem to build an unified formula to account for the correlation among the three scores.

As an initial attempt towards solving this problem, we propose a prediction model for estimating the quality of the entire image based on the quality of textual and pictorial regions, as described in Sec.3. The performance of the proposed model is measured by computing the correlation between the estimated and ground truth scores. Meanwhile, we compare with a simple averaging combination method of textual and pictorial scores. Table 2 reports the comparison results. It shows that the results of the proposed model are more consistent with visual perception. Although there is still space to improve the performance, the proposed prediction model reflects the contributions of textual and pictorial regions with a high reliability.

Table 2. Comparison of two combination methods

	Average combination			Proposed prediction model		
Distortions	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC
GN	0.9048	5.2572	0.8707	0.9847	2.2691	0.9607
GB	0.9032	5.4064	0.8654	0.9833	2.3888	0.9554
MB	0.9005	5.8983	0.8464	0.9798	1.9622	0.9464
CC	0.8577	6.0412	0.8168	0.9573	2.7777	0.8732
JPEG	0.8609	6.0150	0.8382	0.9458	2.4178	0.9018
JPEG2000	0.8373	6.6196	0.8329	0.9372	2.6148	0.8946
LSC	0.8120	6.9176	0.8136	0.9141	2.6838	0.8536
Overall	0.8674	5.9514	0.8433	0.9472	3.8577	0.9234

4.3. Applicability of Traditional IQA Methods to SCIs

Aiming to investigate the effectiveness of state-of-the-art objective IQA methods in quality evaluation of distorted SCIs, the following 12 IQA metrics [13] are applied to *SIQAD*: PSNR, SSIM, MSSIM, VIF, IFC, UQI, NQM, VSNR, WSNR, FSIM, GSIM and GMSD. Most of them are implemented using the toolbox [23] and the codes of others are from their public websites. We apply all the metrics to the grayscale version of images, and compute the correlations between the predicted values and the DMOS values in terms of PLCC, RMSE and SROCC. Meanwhile, the correlations for specific distortions are calculated, to investigate the effectiveness of IQA methods for different distortion types. We report the correlation results in Table 3, where the ones of the best performance are marked with **bold** fonts.

It is shown from Tables 3 that the VIF achieves the highest correlation with the DMOS values in terms of the three measures. Correlations between the VIF and DMOS scores for different distortion types are distinct from each other, as most of the other metrics. Particularly, it has much higher values for the first three distortions (i.e., GN, GB and MB) than others. The reason is that observers are sensitive to such kinds of distortions that are allocated in the entire image, and are able to distinguish the images with different distortion levels. Meanwhile, most IQA metrics are effective to detect these three distortions. However, for the remaining four types, especially for the CC case, the correlation results of the VIF scores and the DMOS values are not as good. For example, the SROCC value of VIF for the CC case is only 0.7607, which indicates the severe inconsistency between the predicted scores and the visual quality of the contrast changed SCIs. The reason may be that contrast change only affects the intensity of texts, but not the integrity of texts about which subjects care more. By contrast, the IQA metrics take the intensity variation into account, resulting in the inconsistency with DMOS values.

From Tables 3, we can also find that the overall correlation results are much lower than the distortion specified results. Although the VIF method achieves the highest overall correlation with the DMOS values (PLCC = 0.8429, SROCC = 0.8183 and RMSE = 7.2295), this result only represents a limited success in predicting human visual perception.

Table 3. Correlation results of the DMOS values and the objective scores given by 12 IQA methods.

	Distortions	PSNR	SSIM	MSSIM	VIF	IFC	UQI	NQM	WSNR	VSNR	FSIM	GSIM	GMSD
PLCC	PLCC	**0.9748**	0.9668	0.9626	0.9682	0.9727	0.9707	0.9717	**0.9748**	0.9722	0.9476	0.9636	0.9640
	GB	0.9802	0.9780	0.9755	0.9797	0.9788	**0.9811**	0.9803	0.9800	0.9802	0.9771	0.9757	0.9808
	MB	0.9631	0.9648	0.9604	0.9664	0.9676	0.9656	**0.9689**	0.9678	0.9657	0.9039	0.9596	0.9660
	CC	0.8542	0.9284	0.9276	0.8806	**0.9321**	0.9300	0.8891	0.8462	0.8710	0.8632	0.9203	0.9225
	JPEG	**0.9403**	0.9231	0.9169	0.9245	0.9301	0.9154	0.9032	0.9332	0.9006	0.9208	0.9207	0.9194
	J2K	0.9096	0.9063	0.9103	0.9090	0.9113	0.9066	**0.9176**	0.9003	0.8852	0.9068	0.9095	0.9082
	LSC	0.9169	0.9192	0.9169	0.9275	**0.9292**	0.9189	0.8396	0.9013	0.9075	0.9002	0.9147	0.9164
	Overall	0.6244	0.7977	0.6508	**0.8429**	0.6736	0.5351	0.6346	0.6787	0.6217	0.6073	0.6161	0.7542
SROCC	GN	0.9375	0.9393	0.9411	0.9393	0.9321	0.9375	0.9321	0.9357	0.9321	0.9286	**0.9446**	0.9393
	GB	0.9411	0.9464	**0.9536**	0.9429	0.9411	0.9482	0.9429	0.9411	0.9411	0.9357	0.9375	0.9411
	MB	0.9375	0.9393	0.9018	0.9393	0.9393	0.9357	**0.9411**	0.9393	0.9393	0.8804	0.9268	**0.9411**
	CC	0.7589	0.7196	0.7821	0.7607	**0.8304**	0.7804	0.7554	0.7107	0.7321	0.7071	0.7625	0.8071
	JPEG	0.8625	0.8554	0.8482	0.8536	0.8536	0.8393	0.8107	**0.8661**	0.8321	0.8589	0.8536	0.8482
	J2K	0.8696	0.8679	**0.8714**	0.8661	0.8661	**0.8714**	0.8482	**0.8714**	0.8607	0.8339	0.8429	0.8679
	LSC	0.8268	0.8250	0.8196	**0.8268**	0.8161	0.8214	0.7268	0.7893	0.8036	0.8232	0.7982	0.8071
	Overall	0.6020	0.7897	0.6345	**0.8183**	0.6347	0.4607	0.6377	0.6947	0.5933	0.5669	0.5832	0.7243
RMSE	GN	2.8622	3.2204	3.4264	3.1413	2.9685	3.0238	2.9522	**2.8470**	2.9364	3.7938	3.3648	3.3401
	GB	2.5150	2.6001	2.7949	2.5691	2.6505	**2.4115**	2.5284	2.5482	2.5417	2.6528	2.7525	2.4724
	MB	2.8209	2.7770	2.8931	2.7163	2.6476	2.7484	**2.6265**	2.6463	2.7122	3.5739	2.9539	2.7289
	CC	5.4663	3.9241	3.6132	4.8653	3.7642	3.7398	4.4100	5.4407	4.6796	4.8120	3.8398	**3.5697**
	JPEG	**2.6938**	2.8237	2.8978	2.8235	2.7733	2.9504	2.9246	2.7616	3.0444	2.8437	2.8771	2.8636
	J2K	3.1672	3.2362	3.1532	3.1791	**3.1239**	3.2382	3.0965	3.2240	3.3721	3.2697	3.1761	3.2101
	LSC	2.5319	2.5445	2.5881	2.4075	**2.3897**	2.6098	3.3314	2.8701	2.8048	2.6829	2.6014	2.5996
	Overall	10.6303	8.1220	10.3000	**7.2295**	9.9235	11.3322	10.3900	9.9495	10.6568	10.4559	10.6430	8.7898

The objective metrics generally capture the practical variations occurring in the distorted images, without considering human's perception when viewing SCIs with different distortions. For instance, in the subjective test, most subjects prefer to give low scores to blurred images. This phenomenon can be observed from Fig.2, where most of the DMOS values for blurred images (from the first eight to the twenty-one points) are higher than other images. Some image examples with their related quality scores are shown in Fig.4 to illustrate this phenomenon. Comparing (c)(d) with (f)(g), although there are no obvious noise artifacts appear in (c) and (d), most subjects have a bad impression to the blurring effect at first sight, and give low scores to the blurred images. Besides, we can observe that the three measures (PSNR, SSIM and VIF) cannot achieve high consistency with the DMOS values. In (b) and (c), there is not much visual quality difference between these two images, but the SSIM gives a much lower score to (b). This inconsistency also appears in (e) to (h): the visual quality of (e) is much better than the other three images in (f)-(g), but the PSNR and SSIM give lower scores to (e). In conclusion, there is a large room to improve and objective measures that can accurately predict the quality of SCIs are still yet to be developed.

5. CONCLUSION

In this paper, we constructed a new large-scale image database, *SIQAD*, to investigate the subjective quality assessment of SCIs. DMOS values of images in the database are obtained via subjective testing, and their reliability is verified. In the subjective test, three scores were given to the entire image and the textual and pictorial regions, respectively, based on which we find that textual regions contributes more to the quality of the entire image in most of the distortion cases. In addition, a prediction model is proposed to account for this relationship. Through the correlation analysis of 12 IQA models (designed for natural images) and the obtained DMOS values, we found that existing IQA methods cannot achieve high consistency with human visual perception when judging the quality of SCIs. In the future, we will investigate the prediction model and use it to guide the construction of objective assessment metrics for distorted SCIs.

References

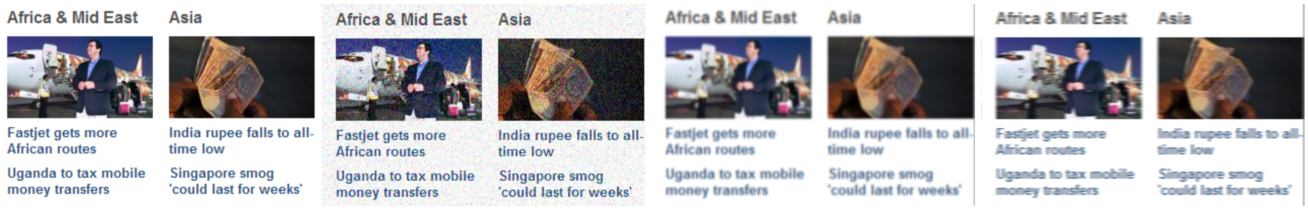
[1] H. Shen, Y. Lu, F. Wu, and S. Li, "A High-Performanance Remote Computing Platform," in *IEEE PerCom*, 2009.

[2] Y. Lu, S. Li, and H. Shen, "Virtualized Screen: A Third Element for Cloud-Mobile Convergence," in *IEEE Multimedia*, 2011.

[3] T. Chang and Y. Li, "Deep Shot: A Framework for Migrating Tasks Across Devices Using Mobile Phone Cameras," in *ACM CHI*, 2011.

[4] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," *IEEE T-IP*, vol. 14, no. 8, pp. 993–1005, 2005.

[5] C. Lan, G. Shi, and F. Wu, "Compress compound im-

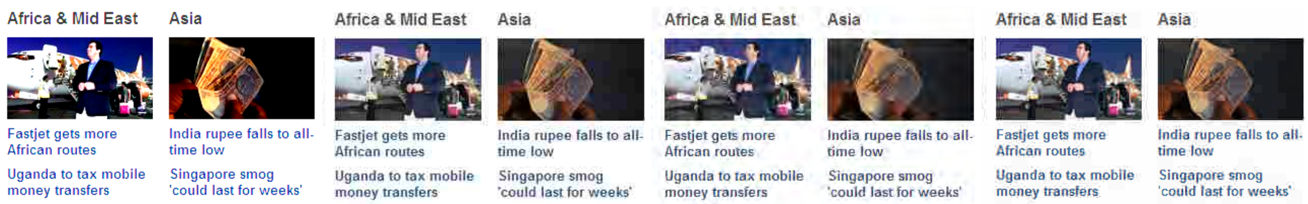


(a) Reference image (cropped from 'cim13' in *SIQAD*)

(b) DMOS: 64.7171, PSNR: 24.4163, SSIM: 0.6302, VIF: 0.4900

(c) DMOS: 67.4922, PSNR: 20.8381, SSIM: 0.8500, VIF: 0.4266

(d) DMOS: 68.8890, PSNR: 20.9106, SSIM: 0.8743, VIF: 0.5359



(e) DMOS: 38.4027, PSNR: 20.5616, SSIM: 0.8595, VIF: 0.5850

(f) DMOS: 57.7172, PSNR: 25.8389, SSIM: 0.8914, VIF: 0.5270

(g) DMOS: 54.9158, PSNR: 26.2688, SSIM: 0.8726, VIF: 0.4134

(h) DMOS: 61.0588, PSNR: 24.8590, SSIM: 0.8718, VIF: 0.4604

Fig. 4. Image quality comparison and quality scores computed by four different methods: DMOS, PSNR, SSIM and VIF. Images in (b)-(h) correspond to seven distortions (GN, GB, MB, CC, JPEG, JPEG2000 and LSC), respectively.

ages in H.264/MPGE-4 AVC by exploiting spatial correlation," *IEEE T-IP*, vol. 19, pp. 946–957, 2010.

[6] H. Yang, W. Lin, and C. Deng, "Learning based screen image compression," in *IEEE MMSP*, 2012.

[7] Z. Pan, H. Shen, S. Li, and N. Yu, "A low-complexity screen compression scheme for interactive screen sharing," *IEEE T-CSVT*, vol. 23, no. 6, pp. 949–960, 2013.

[8] Z. Pan, H. Shen, and Y. Lu, "Brower-friendly hybrid codec for compound image compression," in *IEEE ISCAS*, 2011.

[9] ISO/IEC JTC 1/SC 29/WG 11 Requirements subgroup, "Requirements for an extension of HEVC for coding of screen content," in *MPEG 109 meeting*, 2014.

[10] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[11] Z. Wang and A.C. Bovik, "Mean Squared Error: Love It or Leave It?," *IEEE Signal Processing Magazine*, vol. 26, pp. 98–117, 2009.

[12] W. Lin and C.-C. Jay Kuo, "Perceptual Visual Quality Metrics: A Survey," *Journal of Visual Communication and Image Representation*, vol. 22, pp. 297–312, 2011.

[13] Damon M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," *ISRN Signal Processing*, 2013.

[14] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," in *Int. Telecommunications Union*, 2012.

[15] H.R. Sheikh, F.M. Sabir, and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3441–3452, 2006.

[16] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[17] "SIQAD," `https://sites.google.com/site/subjectiveqa/`.

[18] L. Li and Z.S. Wang, "Compression Quality Prediction Model for JPEG2000," *IEEE Trans. Image Processing*, vol. 19, no. 2, pp. 384–398, 2010.

[19] Y.H. Lee, J.F. Yang, and J.F. Huang, "Perceptual activity measures compouted from blocks in the transform domain," *Signal Processing*, vol. 82, pp. 693–707, 2002.

[20] H. Yang, W. Lin, and C. Deng, "Image Acitivity Measure (IAM) for Screen Image Segmentation," in *IEEE International Conference on Image Processing*, 2012.

[21] K. Soundararajan, NR. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 14271441, 2010.

[22] "Final report from the video quality experts group on the validation of objective models of video quality assessment," `http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx`.

[23] "MeTriX Mux," `http://foulard.ece.cornell.edu/gaubatz/metrix_mux/`.