STUDY OF SUBJECTIVE AND OBJECTIVE QUALITY ASSESSMENT FOR SCREEN CONTENT IMAGES

Xu Wang[†] *Lei Cao*[†] *Yingying Zhu*[†*] *Yun Zhang*[#] *Jianmin Jiang*[†] *Sam Kwong*[‡]

[†] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China
[#]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China
[‡]Department of Computer Science, City University of Hong Kong, Hong Kong

ABSTRACT

In this paper, we present the results of a recent large-scale subjective study of image quality on a collection of screen contents distorted by a variety of application-relevant processes. With the development of multi-device interactive multimedia applications, metrics to predict the visual quality of screen content images (SCIs) as perceived by subjects are becoming fundamentally important. For developing the objective image quality assessment (IQA) method, there is a need for large-scale public database with diversity of distorted types and scene contents, and available subjective scores of distorted SCIs. The resulting Immersive Media Laboratory screen content image quality database (IML-SCIQD) contains 1250 distorted SCIs from 25 reference SCIs with 10 distortion types. Each image was rated by 35 human observers, and the different mean opinion scores (DMOS) were obtained after data processing. The performance comparison of 17 state-of-the-arts, publicly available IQA algorithms are evaluated on the new database. The database will be available online in our project website.

***Index Terms*—** Screen content image, Image Quality Assessment

1. INTRODUCTION

With the rapid development of mobile internet and communication technology, interactive multimedia applications among multi-devices [1] such as multi-screen display, video game and remote desktop accessing, have achieved great development. In these applications, users share screen content images (SCI) for education, entertainment and commercial purposes.

For the multi-device multimedia communication system, the artifacts are inevitably introduced due to the resource shortage in the processing chain including acquisition, compression, transmission, decompression and reconstruction, etc. Thus, it is critical to design high performance screen content image quality assessment (SCIQA) metrics to monitor and guarantee the user's quality of experiences (QoE). However, it is claimed that subject's quality preference on SCIs are significantly different from that on natural images [2]. Since SCIs possess more visual content types such as texts, images and graphics, existing image quality assessment (IQA) metrics [3–15] designed for natural images can not be directly applied to SCIs . Therefore, how to automatically and effectively evaluate the perceptual quality of SCI content becomes a challenging issue.

Recently, researchers have devoted there efforts on designing objective SCIQA metrics [16–19] and evaluated these metrics on the databases generated from subjective quality assessment of SCIs. According to our knowledge, currently there are only two public SCI quality databases named as SIQAD database [20] and SCD database [21]. The SIQAD database contains 20 reference SCIs and 980 distorted SCIs generated from 7 distortions. The reference SCIs are mainly collected from webpages, slides, PDF files and digital magazines. The SCD database contains 24 reference SCIs and 492 distorted SCIs generated from 2 distortions. The reference SCIs are mainly sourced from the visual content of concrete screen operation interface. As for distortion type, the diversity of the SCD database is obviously not enough. Due to the limitation of diversity of distortion types and scene contents, it is insufficient to for developing high performance objective SCIQA metrics. In other words, there is a need for build more public databases specific for SCIs.

In this paper, both of subjective and objective IQA for screen contents are studied. First, we build a SCI quality database (denoted as IML-SCIQD) with 25 reference SCIs and 1250 distorted SCIs in total. The reference SCIs are well selected to guarantee the diversity of visual content and layout style. The whole database contains 10 distortion types. Second, performance comparison of typical IQA metrics are

*Corresponding author. This work was supported in part by the National Natural Science Foundation of China under Grant 61501299, 61672443, 31670553, 61602314, 61620106008 and 61471348, in part by the Guangdong Nature Science Foundation under Grant 2016A030310058 and 2016A030313043, in part by the Shenzhen Emerging Industries of the Strategic Basic Research Project under Grants JCYJ20150525092941043, JCYJ20160226191842793, JCYJ20130326105637578 and in part by the Project 2016049 supported by SZU R/D Fund.

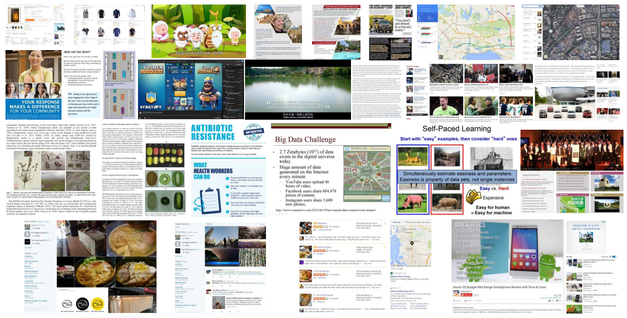

Fig. 1. Reference Images of IML-SCIQD database.

Fig. 2. Quality judgement interface of subjective test

implemented on our database. Experimental results show that the performance of existing objective IQA metrics are limited, there is still a lot of rom to further improve the accuracy of SCIs quality prediction especially for no reference (NR)-IQA metrics. Our database contains diversity of the scene contents and more distortion types, which is a good complement to existing public databases. The research community of SCIQA may benefit from the new database.

The rest of this paper is organized as follows. Section 2 describes the construction of the new database. Experimental results and discussion are provided in Section 3. Finally, conclusions are given in Section 4.

2. DESCRIPTION OF IML-SCIQD DATABASE BUILDING

In order to investigate quality evaluation of SCIs, we firstly build a screen content image quality database with 25 reference SCIs. Then, subjective experiment is conducted on the IML-SCIQD to obtain the DMOS of each distorted SCI. The details are provided as follows.

2.1. Generation of Image Database

In the database, all the reference SCIs are collected and generated through screen snapshot with diverse layout styles, visual contents and resolution. As shown in Fig. 1, the visual contents of reference SCIs contains social media, online shopping, news media, map navigation, digital magazines, slides, movies, and video games, etc. 10 distortion types are applied to generate distorted SCIs. For each type of distortion, 5 degradation levels that cover a wide range of visual quality are selected to generate distorted SCIs. The database contains $25 \times 10 \times 5 = 1250$ distorted SCIs in total.

The distortion types include JPEG compression (denoted as JPEG), JPEG2000 compression (denoted as JP2K), Gaussian blur (denoted as GB), motion blur (denoted as MB), white noise contamination (denoted as GWN), salt & pepper noise (denoted as SPN), multiplicative noise (denoted as MN), contrast change (denoted as CC), bit errors in JPEG2000 bit stream when transmitted over a simulated fast-fading Rayleigh channel (denoted as FF), screen content compression [22] (denoted as SCC). The SCC is specific to screen contents, which aims to improve the compression performance on screen contents. The detailed information of parameters will be public and available in our project website [23].

2.2. Protocol and Methodology of Subjective Quality Assessment Experiment

During the experiment conduct, the Web-Enabled Subjective Test (WEST) platform [24] is modified and employed as the experimental control software, which can offer a solution to gather subjective test data from multiple locations and multiple computing devices. Thus, parallel subjects tests can be conduct by start the local WEST server to save time. To collect quality scores of distorted SCIs from subjects, the single-stimulus methodology is employed for easy operation. For each time, the computer program shows only one image on the screen and the subject is asked to evaluate the the perceived visual quality. In the subjective experiment, all the reference SCIs are also included in the test. Therefore, there are 1275 images prepared for test in total. All these SCIs are grouped into 5 test sessions. Each session consists of 5 reference SCIs and their 5 (scene)$\times$10 (distortion type)$\times$5 (degradation level)$=$250 corresponding distorted ones. Every human subject must finish the quality judgement of all the 255 images in one session. The length of each session is around 30 minutes includes the time for training, demo and actual test. To avoid human subjects fatigue and affect their quality judgment. Each human subject is required to finish all the 5 sessions in 5 different days.

Before the start of actual test, researchers are briefed about the background of IQA and how to do quality judgment concretely. There are 35 subjects in total recruited in the experiment tests. They are all undergraduate or graduate students from Shenzhen University with no experience in image processing or quality assessment, aging from 18 to 28 years old. The test is parallel performed on four desktop computers with Windows 7 operating systems and Google Chrome web browser. The test images are displayed on Lenovo LT2223wA LED monitor with original resolution. After human subjects log into WEST server, the images will be shown on the screen consecutively and human subjects are required to give a score to each image based on their visual perception. As shown in Fig. 2, the quality judgement interface will appear on the screen once subjects finish the view of image. The quality scale is in the range of 1 to 100. The quality scale is divided into five equal portions for easy quality judgement, which are labeled bad, poor, fair, good, and excellent. The higher scores you give, it indicates that the better visual perception the image brings to you. Human subjects are asked firstly to decide which quality portion the image belongs roughly to and then finely tune the slider to a concrete position which they think the quality scores of the image should be. If they want to review the image shown just now, they could click the BACK button. If not, they should click the VOTE button. The position of the slider is then converted to an integer quality score in the range of 1 to 100 and the next test image in that session will be presented.

2.3. Data Processing

After obtaining raw subjective scores, the raw subjective scores are converted into DMOS according to the methods detailed in [25]. Let s_{ijk} denote the raw subjective score assigned by subject i to distorted image j in session k. We firstly convert the raw subjective scores s_{ijk} assigned by subject i in session k to difference score d_{ijk}.

$$d_{ijk} = s_{ij_{ref}k} - s_{ijk}, \tag{1}$$

where $s_{ij_{ref}k}$ denote the raw subjective scores assigned by subject i to the corresponding reference image of distorted image j in session k. Then, the difference score d_{ijk} corresponding to subject i in batch k will be converted to Z-scores z_{ijk}. Final, we can get a Z-score matrix, where the row is the subject index and column is the distorted SCI index. A subject rejection procedure is followed to discard scores from unreliable subjects. We first determine if the scores assigned by a subject i in the i-th row of Z-score matrix are normally distributed by computing the kurtosis of the scores. The scores are considered normally distributed if the kurtosis falls between the values of 2 and 4. If the scores are normally distributed, the procedure rejects a subject whenever more than 5% of scores assigned by him falls outside the range of two standard deviations from the mean scores. If the scores are not normally distributed, the subject is rejected whenever more than 5% of his scores falls outside the range of 4.47$\times$standard deviations from the mean scores. After a subject rejection procedure, 3 out of 35 subjects are rejected. The updated Z-scores z_{ij} are then rescaled to the range of 0 to 100 through a linear mapping. The rescaled Z-scores is denoted as z'_{ij}.The DMOS of each distorted SCI is then computed by

$$DMOS_j = \frac{1}{N}\sum_{i=1}^{N} z'_{ij}. \tag{2}$$

3. PERFORMANCE COMPARISON OF OBJECTIVE METRICS

In order to investigate the performance of existing IQA metrics on quality prediction of SCIs, 11 FR-IQA metrics including PSNR, SSIM [3], VIF [4], IFC [5], VSNR [6], FSIM [7], GMSD [8], GSIM [9], VSI [10], GSS [18] and SQMS [16], 1 reduced-reference (RR) metric RR_SCI [26] and 5 NR-IQA including BIQI [11] , DIIVINE [12], BLIINDS2 [13], BRISQUE [14], SSEQ [15] are evaluated on the IML-SCIQD. All the source codes are collected from the project websites of metrics. The GSS, SQMS and RR_SCI are the metrics designed for quality prediction of SCIs. Other metrics are designed for natural scenes only. To remove nonlinearity introduced by the subjective rating process and facilitate empirical comparison of different IQA metrics, the nonlinear least-squares regression function *nlinfit* of Matlab is employed to map the objective quality score q to the predicted subjective quality score $DMOS_p$. The mapping function is the five parameters logistic function

$$DMOS_p = \frac{p_1}{2} - \frac{p_1}{1 + exp(p_2 \cdot (q - p_3))} + p_4 \cdot q + p_5, \tag{3}$$

where p_1, p_2, p_3, p_4 and p_5 are the parameters of the logistic function. Three criteria are employed to evaluate the mapping performance: (1) correlation coefficient (CC): accuracy of objective metrics; (2) Spearman's rank order correlation coefficient (SROCC): monotonicity of objective metrics; and (3) root mean-squared-error (RMSE). For learning based NR-IQA metrics. The 80%-20% train-test partition scheme is employed and the partition is conducted 1000 times. The mean performance is obtained as the overall performance. Detailed experimental results are provided in Tables 1. The metric achieved the best performance for each distortion type is highlighted with bold font. The experimental results indicate that existing IQA metrics cannot achieve the similar high performance that they have achieved on natural images. In other words, existing IQA metrics are limited in predicting visual quality of SCIs. The results of natural scene statistics (NSS) based NR-IQA metrics on those compression distortions are especially notable. The reason may be that SCIs are not natural scenes and NR-IQA metrics cannot access the reference image. According to the performance comparison results, there is still a lot for room to further improve the accuracy of SCIs quality prediction especially for no reference (NR)-IQA metrics.

4. CONCLUSION

In this paper, we build a screen content image quality database for the study of quality evaluation of SCIs. Based on the performance evaluation and comparison of existing IQA metrics, it can be concluded that there is still a lot of room to further investigate on the SCIQA, especially for the NR-IQA metrics. In the future, we will design objective IQA

Table 1. Performance of the objective quality metrics on database in terms of CC, SROCC and RMSE

Metric	JPEG	JP2K	GB	MB	GWN	SPN	MN	CC	FF	SCC	ALL
PSNR	0.7199	0.8411	0.7777	0.7839	0.9494	0.8707	0.9177	0.7394	0.9158	0.8734	0.7776
SSIM	0.8915	0.9302	0.9577	0.9210	0.9447	0.8826	0.9194	0.3797	0.9401	0.8681	0.8861
VIF	**0.9413**	0.9568	0.9452	0.8765	**0.9601**	0.8933	**0.9489**	0.5052	0.9403	**0.9314**	0.8719
IFC	0.8051	0.8802	0.8787	0.7483	0.8333	0.8205	0.8237	0.4875	0.8046	0.7942	0.7689
VSNR	0.4374	0.4946	0.4851	0.8097	0.8940	0.6476	0.8681	0.6748	0.4767	0.3352	0.4327
FSIM	0.8720	0.9531	0.8857	0.8035	0.9341	0.8089	0.9239	0.6850	0.9518	0.9057	0.8156
GMSD	0.8580	0.9539	0.9052	0.8018	0.9551	0.8058	0.9353	0.3208	0.9449	0.9114	**0.9078**
GSIM	0.8751	0.9420	0.9157	0.8188	0.9082	0.7749	0.8346	**0.8013**	0.9304	0.9223	0.7833
VSI	0.8636	0.9447	0.9156	0.8149	0.9339	0.8202	0.9288	0.4598	**0.9594**	0.9031	0.8188
GSS	0.8370	0.8847	0.9525	0.8981	0.8727	0.6783	0.8800	0.2921	0.8600	0.8168	0.8638
SQMS	0.9023	**0.9614**	0.9324	0.8902	0.9590	**0.9011**	0.9403	0.5979	0.9426	0.8972	0.8975
RR_SCI	0.8256	0.8949	**0.9699**	**0.9506**	0.9545	0.8914	0.9450	0.4304	0.9149	0.8149	0.8563
BIQI	0.2665	0.2560	0.7806	0.6166	0.9422	0.8976	0.9017	0.6947	0.4226	0.2259	0.4551
DIIVINE	0.2242	0.2886	0.6224	0.5658	0.9022	0.8324	0.8481	0.5009	0.3982	0.2425	0.5671
BLIINDS2	0.4447	0.5941	0.3370	0.2829	0.7200	0.7244	0.6778	0.2260	0.4915	0.3171	0.5563
BRISQUE	0.2258	0.4051	0.7442	0.6557	0.9422	0.8952	0.8509	0.2517	0.4294	0.2733	0.6024
SSEQ	0.3426	0.4063	0.2313	0.2046	0.8709	0.8123	0.7863	0.2438	0.6439	0.2081	0.5921
PSNR	0.8510	0.9174	0.8920	0.7783	0.8828	0.7785	0.8043	**0.6067**	0.8398	0.7971	0.7621
SSIM	0.8991	0.9222	0.9384	0.9198	0.9040	0.8424	0.8240	0.2579	0.8578	0.7927	0.8814
VIF	**0.9063**	0.9457	0.9253	0.8475	0.9186	0.8479	0.8725	0.4853	0.8694	0.8091	0.8629
IFC	0.7817	0.8842	0.8756	0.7062	0.8226	0.7951	0.7939	0.3469	0.5582	0.7292	0.7290
VSNR	0.8393	0.9221	0.9047	0.7878	0.7836	0.6046	0.7480	0.5012	0.8155	0.7840	0.7113
FSIM	0.8630	0.9435	0.8870	0.7859	0.8496	0.7373	0.8279	0.5542	0.8771	0.7942	0.8079
GMSD	0.8550	0.9408	0.8961	0.7729	0.9051	0.6953	0.8477	0.1476	0.8919	**0.8208**	**0.9062**
GSIM	0.8727	0.9328	0.9081	0.8067	0.8043	0.6815	0.7521	0.6035	0.8143	0.8024	0.7803
VSI	0.8476	0.9346	0.9087	0.8044	0.8450	0.7357	0.8368	0.4001	**0.8998**	0.7990	0.8062
GSS	0.8210	0.8677	0.9325	0.8865	0.8472	0.2786	0.7960	0.2193	0.6807	0.7265	0.8331
SQMS	0.8880	**0.9463**	0.9188	0.8648	0.9169	**0.8623**	0.8620	0.4937	0.8605	0.8177	0.8906
RR_SCI	0.7880	0.8872	**0.9527**	**0.9449**	**0.9228**	**0.8623**	**0.8894**	0.4380	0.7639	0.7367	0.8423
BIQI	0.2438	0.1372	0.7444	0.6216	0.9077	0.8267	0.8566	0.5145	0.3225	0.1553	0.6656
DIIVINE	0.1105	0.2251	0.6074	0.5743	0.8108	0.7559	0.7290	0.3615	0.3290	0.1820	0.6303
BLIINDS2	0.3317	0.5898	0.2020	0.1636	0.6942	0.6597	0.6294	0.1667	0.3638	0.1780	0.5550
BRISQUE	0.1550	0.3852	0.6970	0.6232	0.8863	0.8160	0.7969	0.1895	0.4054	0.2258	0.6516
SSEQ	0.2533	0.3758	0.1377	0.1348	0.8197	0.7693	0.7184	0.2006	0.5170	0.1300	0.5597
PSNR	5.2827	6.6769	6.7154	5.1610	3.3735	3.9396	3.9541	4.4215	4.6251	3.7936	7.6999
SSIM	3.4487	4.5305	3.0744	3.2380	3.5209	3.7602	3.9144	6.0758	3.9235	3.8658	5.6758
VIF	**2.5704**	3.5874	3.4877	4.0007	**3.0017**	3.5960	**3.1401**	5.6678	3.9173	**2.8350**	5.9967
IFC	4.5149	5.8590	5.1003	5.5131	5.9367	4.5735	5.6422	5.7345	6.8363	4.7332	7.8296
VSNR	6.8448	10.7440	9.3572	4.8780	4.8102	6.0956	4.9413	4.8474	10.1200	7.3381	11.0400
FSIM	3.7258	3.7371	4.9596	4.9479	3.8345	4.7037	3.8074	4.7850	3.5328	3.3016	7.0863
GMSD	3.9096	3.7029	4.5405	4.9666	3.1815	4.7386	3.5224	6.2245	3.7688	3.2059	**5.1373**
GSIM	3.6836	4.1427	4.2924	4.7720	4.4937	5.0565	5.4820	**3.9299**	4.2214	3.0095	7.6128
VSI	3.8378	4.0474	4.2951	4.8177	3.8403	4.5768	3.6883	5.8323	**3.2479**	3.3441	7.0309
GSS	4.1649	5.7548	3.2540	3.6556	5.2429	5.8776	4.7257	6.2812	5.8751	4.4938	6.1708
SQMS	3.2811	**3.3966**	3.8610	3.7863	3.0426	3.4691	3.3856	5.2644	3.8452	3.4402	5.4017
RR_SCI	4.2954	5.5075	**2.6034**	**2.5812**	3.2038	3.6252	3.2546	5.9284	4.6471	4.5147	6.3257
BIQI	7.0806	11.6870	6.0538	6.0824	3.3231	**3.3609**	3.4808	4.1924	9.8149	7.4140	9.5890
DIIVINE	7.2114	11.3840	7.2856	6.0666	4.3819	4.1646	5.0475	4.9762	9.9571	7.2556	9.1863
BLIINDS2	6.5585	9.1931	9.7352	7.7541	6.8148	5.1455	6.8338	6.2473	9.3016	7.0503	9.4890
BRISQUE	7.2724	10.7510	6.6044	5.7673	3.1733	3.3930	4.5028	6.1439	9.6659	7.2081	8.9630
SSEQ	6.8723	10.7680	10.2180	7.9817	4.8147	4.2604	5.7079	6.1562	8.1352	7.4015	9.2651

metrics specific to SCIs based on existing database including our IML-SCIQD database.

5. 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, 2011.

[2] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4408–4421, 2015.

[3] 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, no. 4, pp. 600–612, 2004.

[4] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[5] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity creation for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[6] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images,"

IEEE Transactions on Image Processing, vol. 16, no. 9, pp. 2284–2298, 2007.

[7] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[8] 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.

[9] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, 2012.

[10] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.

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

[12] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 3350–3364, 2011.

[13] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 3339–3352, 2012.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 4695–4708, 2012.

[15] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[16] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1098–1110, 2016.

[17] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, pp. 140C149, 2016.

[18] Z. Ni, H. Zeng L. Ma, C. Cai, and K.-K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1394–1397, 2016.

[19] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 2016–2027, 2017.

[20] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX),*. IEEE, 2014, pp. 257–262.

[21] S. Shi, X. Zhang, S. Wang, R. Xiong, and S. Ma, "Study on subjective quality assessment of screen content images," in *Picture Coding Symposium (PCS), 2015*. IEEE, 2015, pp. 75–79.

[22] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Screen content coding based on HEVC framework," *IEEE Transactions on Multimedia*, vol. 16, pp. 1316–1326, 2014.

[23] Immersive Media Lab of Shenzhen University, "Project of screen content image quality assessment," http://csse.szu.edu.cn/staff/wangx/wordpress/project/sciqa/.

[24] "Web-enabled subjective test (WEST) software," http://www.its.bldrdoc.gov/resources/video-quality-research/web-enabled-subjective-test-west.aspx.

[25] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[26] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Transactions on Circuits and Systems for Video Technology*, in press.