HDR2014 - A HIGH DYNAMIC RANGE IMAGE QUALITY DATABASE

Min Liu, Guangtao Zhai, Shen Tan, Zhili Zhang , Ke Gu, and Xiaokang Yang

Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong Univ., Shanghai, China, 200240
{liumin_merry, zhaiguangtao, tanshen, 345492947, gukesjtuee, xkyang}@sjtu.edu.cn

ABSTRACT

High dynamic range (HDR) imaging has attracted a lot of attention and enthusiasm in the last decades. With the quick advances of sensor technologies, even consumer level digital cameras are capable of capturing HDR images. However, a vast majority of nowadays displays still only have 8-bit color depth, and this leads to the widely studied topic of displaying HDR images on low dynamic range (LDR) devices or tone mapping. On the other hand, the recent emergence of 10-bit display devices brings the possibility of direct visualization of HDR images. So a natural question to ask is whether existing popular image quality metrics designed for and validated on LDR (8-bit) images perform equally well for HDR (10-bit) images. In this paper we propose a new and dedicated High Dynamic Range image quality database (HDR2014). That database is composed of 192 images with four kinds of distortions applied on 6 reference images. More specifically, we use 8 distortion levels for the artifacts of JPEG/JPEG2000 compression, white noise injection and Gaussian blurring. Twenty-five inexperienced viewers were involved in the subjective viewing test. Images were displayed on a pair of carefully calibrated 8-bit LDR and 10-bit HDR monitors and the subjective scores on both of which were recorded. We then tested some ubiquitous and state-of-the-art IQA metrics on the HDR2014 database. Experimental results show that HDR monitor indeed improved perceptual quality of the visual stimuli, as compared to LDR ones. And several existing IQA metrics are still doing well on HDR images, yet performance of some metrics drop significantly.

Index Terms— Image quality assessment (IQA), mean opinion score (MOS), high dynamic range (HDR), low dynamic range (LDR), 10-bit monitor, 8-bit monitor

1. INTRODUCTION

Images involved in current research of image quality assessment (IQA) are low dynamic range (LDR) images. However, in the real world, our visual system is presented with a wide range of light, spanning approximately ten orders of absolute range from direct sunlight to faint sunlight. So high dynamic range (HDR) imaging expressing the real scene has become a landmark in the history of imaging science. And in theory, the visual quality of HDR images are much higher than conventional LDR images for containing more extended range of contrast, detail, brightness and color. With the advances of sensor technology, the availability of HDR images becomes possible. It produces revolutionary change in digital photography, visual art performance, entertainment and games, even medical and security imaging [1].

Although HDR images are becoming increasingly more commonplace and important in computer graphics, the dynamic range of various common devices (monitors, printers, etc) is much smaller than the dynamic range actually found in real-world scenes. Therefore, tone mapping operators (TMOs) rendering HDR to LDR images has been an emerging field a long time ago. Excitingly, we have witnessed a significant advancement of display technology recently. A 10-bit technology invention has created a tipping point, making the direct visualization of HDR images possible. A natural question here is whether the existing image quality assessment (IQA) metrics which perform well on popular databases (LIVE [2], TID2008 [3], CSIQ [4], Toyama [5], IVC [6] and A57 [7]) are still suitable for HDR images. Therefore, we propose a new High Dynamic Range image quality database (HDR2014) in this paper, which contains 6 lossless images chosen from [8]-[10], and their derived 192 distorted images (four distortion types: JPEG, JPEG2000, white noise and Gaussian blur, and eight levels for each distortion). Subjective test for the database is conducted on 8-bit and 10-bit monitors under the environment instructed by ITU-R BT.500-12 [11].

As we all know, objective IQA can be divided into three kinds: full-reference (FR), reduced-reference (RR) and no-reference (NR), based on the availability of the original image[1]. In the last century, the mean-squared error (MSE) and its equivalent peak signal-to-noise ratio (PSNR) were a pair of commonly used benchmark methods, owing to their simplicity, strong portability and clear physical meaning. Yet both MSE and PSNR do not take the influence of image content into account, leading to their poor correlation with the human judgement of quality, namely the mean opinion score (MOS) [12]. Aiming at the shortage of MSE and PSNR, Wang *et al.* proposed a valid structural similarity index (SSIM) [13] based

[1] FR means the whole reference image is assumed to be known; RR is that the reference image is partially available; NR indicates the original image is not available.

on the hypothesis that the human visual system (HVS) is highly paying attention to structural information, and the SSIM was realized to compare the distinguishment in terms of luminance, contrast and structural information between the reference and distorted images. Inspired by this, numerous variants of SSIM have been proposed in succession [14]-[18]. Except FR IQA algorithms mentioned above, an increasing number of researchers have devoted to the exploitation of N-R IQA metrics [20]-[24] and RR algorithms [25]-[28]. The above-mentioned FR, RR and NR IQA algorithms are tested and compared on the new database to verify their effectiveness and applicability on HDR images.

The structure of this paper is as follows. Section 2 describes the details of our proposed HDR2014 database, along with the subjective testing method. In Section 3, several popular FR, RR, and NR metrics are tested on the HDR2014 database. Experimental results and comparative studies are then reported and analyzed. Finally, Section 4 concludes this paper and proposes some possible future work.

2. THE HDR2014 DATABASE

2.1. The content of HDR2014 database

Similar to LDR images, HDR images in real life application usually undergo various process stages, such as acquisition, compression, transmission, and presentation, before reaching the final observers. Consequently, various image distortions, such as compression artifacts, blurring, and noise injection could occur during the course. For the limited viewing distance, we elaborately selected six lossless natural HDR images (Scene 1-6) of size 512 $\times$ 384 from [8]-[10] (not restricted by copyright) and etc. All images are converted to 8-bit images by Adobe Photoshop, displayed in Figure. 1. Another 192 images were generated by contaminating the pristine ones with four distortions, JPEG, JPEG2000 (JP2K), Gaussian blur and white noise (WN), and the distortion level is eight respectively. We illustrate the details of four artifacts JPEG, JP2K, white noise and Gaussian blur as follows.

- JPEG: The $imwrite$ command from Matlab is used to compress HDR images into JPEG compressed images using Q parameters (70, 60, 50, 40, 30, 20, 10, 5). Then use $hdrwrite$ command to generate HDR images from JPEG images.
- JP2K: We also use Matlab command $imwrite$ to create JP2K compressed images according to $CompressionRatio$ parameters. The ratios we adopted are 20, 40, 60, 80, 100, 200, 300, and 400. Then these images are stored by $hdrwrite$ in HDR format.
- White noise: We add eight kinds of normal noise with μ = 0, σ_N^2 = (0.002, 0.004, 0.008, 0.01, 0.04, 0.08, 0.1, 0.3) to each of the three planes R, G and B by the Matlab $imnoise$ function.

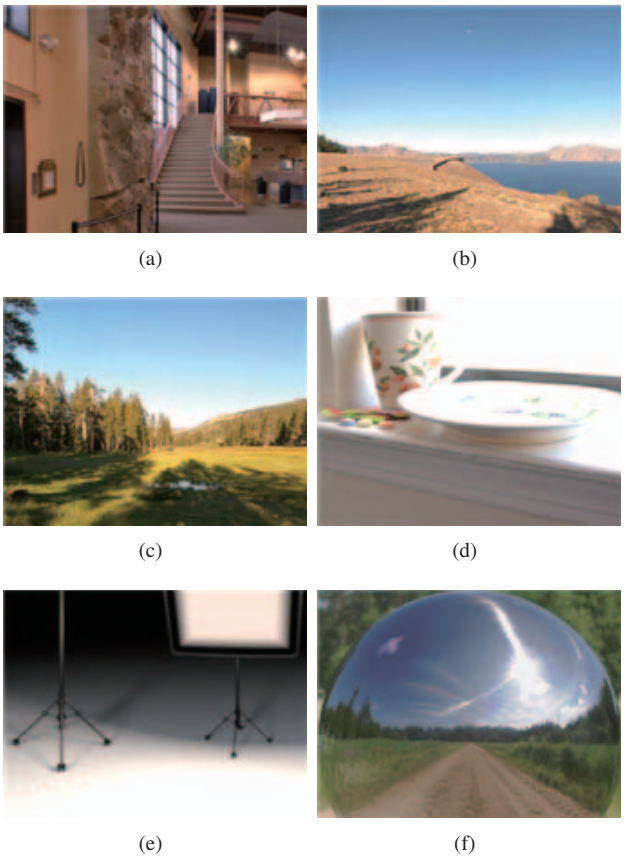

Fig. 1: The six original lossless natural color images of size 512 $\times$ 384 from websites. (a)-(f) are 8-bit format images of Scene 1-6 respectively.

- Gaussian blur: By using $fspecial$ and $imfilter$ commands, we adopt Gaussian kernels with a window of size $l_G \times l_G$ (l_G = 20), and the standard deviation is σ_G = (1, 2, 3, 4, 5, 6, 7, 8). The three panels R, G and B are all blurred using the same kernel.

For example, we show three different levels of Gaussian blur distorted image of Scene 2, three different levels of JPEG distorted images of Scene 3, three different levels of JP2K distorted images of Scene 4 and three different levels of white noise distorted images of Scene 5 in Figure. 2 (all in 8-bit format). And the distortion level increases from left to right. The database will be available on the author's website, which is omitted in this anonymous submission.

2.2. The subjective test of HDR2014 database

To illustrate the differences between 8-bit and 10-bit monitors, the images from HDR2014 database were displayed on a LCD (8-bit) monitor and a HDR (10-bit) monitor separately in our experiment (both displayed by the Adobe Photoshop software). The resolutions of LCD and HDR displays are 1920 $\times$ 1080 and 1920 $\times$ 1200, and their refresh rates

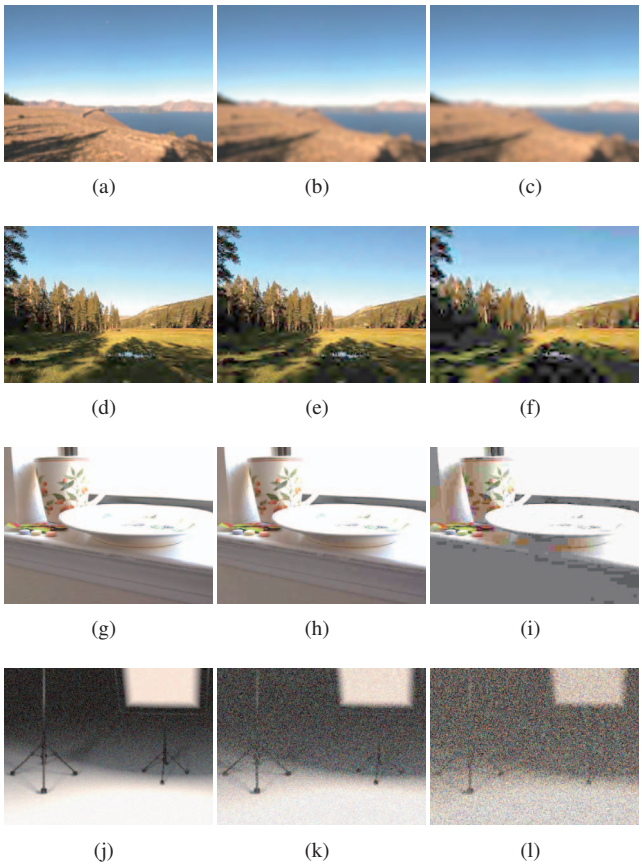

Fig. 2: (a)-(c) Three Gaussian blur distorted image of Scene 2; (d)-(f) three JPEG distorted images of Scene 3; (g)-(i) three JP2K distorted images of Scene 4; (j)-(l) three white noise distorted images of Scene 5.

are 60 Hz and 59 Hz. The experiment was conducted using a single-stimulus (SS) method refering to ITU-R BT.500-12 [11]. Twenty-five viewers (fourteen male, eleven female, age from twenty to thirty) were involved, most of them were college students, from different majors (computer science, electronic engineering, chemistry and etc). And all of them had

Table 1: Subjective test conditions and parameters.

Method	Single-stimulus (SS)
Evaluation scales	Continuous quality scale from 0 to 1
Color depth	30-bits/pixel color images
Image coder	High Dynamic Range (HDR)
Subjects	Twenty-five inexperienced observers
Image resolution	512×384
Viewing distance	Four times the image height
Room illuminance	Dark

Table 2: MOS of original images displayed on 8-bit and 10-bit monitors. The higher MOS values are bolded for every single image.

Display device	Scene 1	Scene 2	Scene 3
8-bit monitor	0.8924	0.6529	0.7136
10-bit monitor	**0.9061**	**0.7740**	**0.8229**
Display device	Scene 4	Scene 5	Scene 6
8-bit monitor	0.7829	0.6760	0.8006
10-bit monitor	**0.8383**	**0.7869**	**0.8331**

normal or corrected to normal vision. Each participant was presented with all of the HDR images displayed on two monitors separately, and no subject took part in more than one experiment. To eliminate order effect, the image presentation order was randomized. Some important parameters of the subjective testing condition are illustrated in Table I. After observing the images for seconds, participants are asked to grade the image on a continuous quality scale from 0 to 1 with precision up to 0.01%.

We tabulate the mean opinion score (MOS) of the original images displayed on 8-bit monitor and 10-bit monitor separately in Table 2. We find out that the quality scores of reference images on 10-bit monitor are higher than those on 8-bit monitor. To illustrate, for Scene 2, Scene 3 and Scene 5, the luminance and color are continuous when displayed on 10-bit monitor. But obvious color hopping can be detected on 8-bit display, which is more apparent on the sky (Fig. 1(b), (c)) and ground (Fig. 1(e)). For Scene 4, the presentation on 10-bit monitor has more details (more decorative pattern on the dish) in bright regions than on 8-bit monitor. For Fig. 1(a) and (f), the difference of presentation performance is not big. All these phenomena confirm that the HDR images have larger range of colors, contrast and intensities than LDR images. However, from Fig. 2, we can figure out that it becomes more difficult to tell the difference between distorted HDR images displayed on 10-bit and 8-bit monitors when the distortion level becomes higher. Then, we eliminate some outlier scores (far from other scores) and calculate the differential mean opinion score (DMOS) values between the original images and their derived distorted images, which will be used in the following IQA.

3. OBJECTIVE IMAGE QUALITY ASSESSMENT ON THE HDR2014 DATABASE

3.1. Objective testing IQA metrics

Following the original metrics MSE and PSNR, many FR quality assessment algorithms have been proposed successively during the last decade. In 2004, a classical algorithm structural similarity index (SSIM) [13] was proposed, which is a combination of the luminance comparison function l,

contrast comparison function c, and structure similarity function s. Following that, various FR IQA algorithms are proposed based on NSS model, image gradient, human brain science, and etc, including multi-scale SSIM (MS-SSIM) [14], information content weighted SSIM (IW-SSIM) [15], state-of-the-art feature similarity (FSIM) [16], gradient similarity (GSM) [17], and internal generative model (IGM) [18]. Besides, H. R. Sheikh proposed visual information fidelity (VIF) [19], on the basis of a statistical model for natural scene and information-theoretic setting.

The NR IQA metrics are more valuable than FR IQA algorithms on account of their wide usages in most applications, under the condition of which the original image signal cannot be acquired. In the early stage, there were three popular NR IQA algorithms Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [20], BLind Image Integrity Notator by using DCT Statistics (BLIINDS-II) [21], and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [22], which respectively work well using support vector machine (SVM) [29] in DCT, DWT and spatial domains. And furthermore, last year has seen another two novel NR metrics, which work without human scored images, prior knowledge of the image contents and distortion categories [23]-[24].

As a tradeoff between FR and NR IQA, a great amount of RR methods have also been proposed for high IQA performance [25]-[28]. For instance, free energy based distortion metric (FEDM) [25] was designed to approximate the internal generative model of the human brain when perceiving an input visual signal. And structural degradation model (SDM) [28] succeeded to modify the well-known SSIM into valid RR IQA algorithms.

To the best of our knowledge, the above FR, RR and NR methods all achieved favorable performance on the well-known databases, such as LIVE, CSIQ, TID2008, Toyama, IVC, and A57. But all of these databases are composed of LDR (8-bit) images. Naturally, an essential and unaddressed question here is $whether\ these\ objective\ image\ quality\ metrics\ sitll\ perform\ well\ on\ HDR\ images.$

It needs to highlight here that Adobe Photoshop adopts a gamma correction when displaying HDR images. So we use a gamma correction to preprocess those images before using IQA metrics. It is common to adopt a gamma coefficient γ = 2.2 for displays [30], the gamma correction function is

$$L_d = L_w^{1/\gamma} \tag{1}$$

where L_w is the actual world luminance, and L_d is the displayed luminance.

3.2. Experimental results

After obtaining the objective prediction scores of every metric, we adopt the nonlinear regression to map the scores to subjective ratings using a four-parameter logistic function:

$$q(\varepsilon) = \frac{\xi_1 - \xi_2}{1 + \exp(-\frac{\varepsilon - \xi_3}{\xi_4})} + \xi_2 \tag{2}$$

where ε is the input score, while $q(\varepsilon)$ is the mapped score, and ξ_1 to ξ_4 are free parameters to be determined during the curve fitting process. Four commonly used performance measures, as suggested by VQEG [31], Pearson's linear correlation coefficient (PLCC), Spearman's rank ordered correlation (SROCC), Kendall's rank-order correlation coefficient (KROCC), and Root mean-squared error (RMSE), are adopted to evaluate and compare those competing FR, RR, and NR IQA metrics.

Table 3 tabulates the performance evaluations of PLCC, SROCC, KROCC, and RMSE of nine FR metrics on the HDR2014 database. The evaluation measures for four RR metrics and five NR algorithms on the HDR2014 database

Table 3: PLCC, SROCC, KROCC and RMSE values (after nonlinear regression) of FR metrics: PSNR, IW-PSNR, SSIM, MS-SSIM, IW-SSIM, VIF, FSIM, GSM and IGM on HDR2014 database (198 images). We bold the best performance for specific evaluation.

Metrics	Type	PLCC	SROCC	KROCC	RMSE
PSNR	FR	**0.8851**	0.8876	0.7020	8.4029
IW-PSNR	FR	0.8251	0.8251	0.6307	4.3736
SSIM	FR	0.7603	0.8472	0.6556	0.0964
MS-SSIM	FR	0.8309	0.9045	0.7325	0.0240
IW-SSIM	FR	0.7184	0.8491	0.6591	0.0827
VIF	FR	0.3779	0.3844	0.2669	0.2912
FSIM	FR	0.7049	0.7741	0.5835	0.0231
GSM	FR	0.7545	**0.9320**	**0.7744**	**0.0001**
IGM	FR	0.8091	0.8253	0.6283	0.0048

Table 4: PLCC, SROCC, KROCC and RMSE values (after nonlinear regression) of RR metrics: FEDM, RRED, QFTB, SDM and NN metrics: DIIVINE, BLIINDS-II, BRISQUE, NIQE, QAC on HDR2014 database (198 images). We bold the best performance for specific evaluation.

Metrics	Type	PLCC	SROCC	KROCC	RMSE
FEDM	RR	0.7859	0.2934	0.2396	**0.0438**
RRED	RR	**0.8117**	**0.8737**	**0.6896**	0.0529
QFTB	RR	0.7607	0.7486	0.5650	0.1924
SDM	RR	0.7096	0.5263	0.3893	0.0909
DIIVINE	NR	0.3980	0.3537	0.2896	14.2491
BLIINDS-II	NR	0.1620	0.1436	0.1025	23.4643
BRISQUE	NR	**0.6166**	**0.4353**	**0.3157**	12.9395
NIQE	NR	0.2127	0.1711	0.1165	2.3116
QAC	NR	0.6166	0.1787	0.1184	**0.0935**

Table 5: SROCC values (after nonlinear regression) of FR metrics: PSNR, IW-PSNR, SSIM, MS-SSIM, IW-SSIM, VIF, FSIM, GSM and IGM on every group of content-specific images (Scene 1-6, 32 distorted images for each scene) and every group of distortion-specific types (Blur, JP2K, JPEG, WN, 48 distorted images for each distortion type) from HDR2014 database (198 images). We bold the best performance for every single content and distortion of images for FR algorithms.

Metrics	Type	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6	Blur	JP2K	JPEG	WN
PSNR	FR	0.8783	0.9270	0.9388	0.8966	0.9154	0.9674	0.8376	0.7862	0.8770	0.8413
IW-PSNR	FR	0.8101	0.8177	0.9043	0.8021	0.8938	0.9080	0.7942	0.8213	0.9168	0.8983
SSIM	FR	0.7830	**0.9631**	0.8625	0.9021	0.8395	0.9296	0.7821	0.7598	0.8143	0.9391
MS-SSIM	FR	0.8985	0.9435	0.9373	0.9340	0.8769	0.9578	0.8748	0.8280	0.9050	0.8843
IW-SSIM	FR	0.8559	0.8250	0.8952	0.8244	0.9264	0.9274	0.8915	0.8657	0.9196	0.8878
VIF	FR	0.7430	0.1452	0.8061	0.4534	0.2442	0.2914	0.8196	0.7958	0.0933	0.4206
FSIM	FR	**0.9355**	0.9022	**0.9395**	0.9377	0.6474	0.9344	**0.8983**	0.5928	0.6247	0.6787
GSM	FR	0.9102	0.9284	0.9351	**0.9399**	**0.9385**	0.9556	0.8907	**0.8858**	**0.9406**	**0.9555**
IGM	FR	0.8735	0.8758	0.8790	0.9282	0.9161	**0.9714**	0.8078	0.6366	0.4940	0.8698

are tabulated in Table 4. Table 5 shows the SROCC values of nine FR metrics on each group of content-specific images (Scene 1-6) and each group of distortion-specific types (Blur, JP2K, JPEG and WN). Table 6 presents the SROCC values of four RR metrics and five NR algorithms on each group of content-specific image (Scene 1-6) and each group of distortion-specific types (Blur, JP2K, JPEG and WN).

The higher scores of PLCC, SROCC, KROCC, and the lower score of RMSE indicate the better performance of the quality metrics. From Table 3-6, most FR algorithms perform well on HDR2014 database, especially the GSM which performs the best on the whole database and some distortion types (e.g. JP2K, JPEG and white noise). Unfortunately, VIF has a bad correlation with subjective evaluations. In detail, for Scene 1 and Scene 3, VIF has a high correlation with subjective scores, but for Scene 2, 4-6, the correlation is low, which can be observed from Table 5. So we guess that VIF is a highly content-related and distortion-related metric for HDR images. However, the performance of most RR and NR metrics fall sharply, except the RRED that performs still well for every group of images corresponding to the same content. Observing Table 6, we notice that RR metrics (FEDM, SDM) and NR metrics (DIIVINE, BLIINDS-II, BRISQUE, and NIQE) are highly sensitive to image content. In addition, FEDM, SDM, BLIINDS-II, BRISQUE and QAC have various evaluation effects for different distortion types. The performances of QAC on each group of the same image content are generally very low. Surprisingly, we find that most RR and NR methods are very effective for white noise distorted HDR images.

4. CONCLUSIONS

With the advances of HDR images and the evolution of displaying technique, we propose a new High Dynamic Range image quality database (HDR2014) in this paper. It is composed of 198 HDR images, among which 6 are original lossless natural images and 192 are derived distorted images (contaminated by JPEG, JPEG2000, gaussian blur and white noise). Twenty-five viewers are invited to grade these HDR images displayed on 8-bit monitor and 10-bit monitor in accordance with the requirements provided by ITU-R BT.500-12. Subjective results show that HDR images have superior performance than LDR images. A total of eighteen classical and state-of-the-art IQA algorithms (nine FR metrics, four R-R methods, and five NR algorithms) are tested and compared on this database. Some algorithms still perform well, while others drop sharply.

In the future, we intend to enrich our HDR2014 database by bringing in some medical, entertainment or artistic HDR images, making it have a vast diversity of image contents. Besides, objective IQA algorithms specialized for HDR images should be studied further.

Acknowledgment

This work was supported in part by NSFC (61025005, 61371146, 61221001), 973 Program (2010CB731401) and FANEDD (201339).

5. REFERENCES

[1] Erik Reinhard, Greg Ward, Sumanta Pattanaik, Paul Debevec, "High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting," San Mateo, CA: Morgan Kaufmann, 2010.

[2] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality

[3] 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.

Table 6: SROCC values (after nonlinear regression) of RR metrics: FEDM, RRED, QFTB, SDM and NN metrics: DIIVINE, BLIINDS-II, BRISQUE, NIQE, QAC on every group of content-specific image (Scene 1-6, 32 distorted images for each scene) and every group of distortion-specific types (Blur, JP2K, JPEG, WN, 48 distorted images for each distortion type) from HDR2014 database (198 images). We bold the best performance for every single content and distortion of images for RR and NR algorithms.

Metrics	Type	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6	Blur	JP2K	JPEG	WN
FEDM	RR	0.0492	0.6182	0.0718	0.4989	0.6130	0.7484	0.4642	0.3698	0.5034	0.9291
RRED	RR	**0.8919**	**0.9405**	**0.9424**	**0.9421**	**0.9172**	**0.9516**	0.8064	**0.8349**	**0.8318**	0.7066
QFTB	RR	0.8809	0.8578	0.9410	0.8695	0.9161	0.9194	**0.8207**	0.5733	0.5109	0.7327
SDM	RR	0.3952	0.4425	0.1712	0.7610	0.7706	0.6125	0.7045	0.3823	0.6118	**0.9595**
DIIVINE	NR	0.0594	0.5985	**0.7661**	0.1961	0.4557	0.4688	0.1052	**0.3658**	0.3120	0.1639
BLIINDS-II	NR	**0.7938**	0.2314	0.6457	0.1639	0.3426	0.7281	0.1499	0.0009	**0.4659**	0.7485
BRISQUE	NR	0.0268	**0.6298**	0.1485	**0.8948**	0.5431	**0.8017**	0.0468	0.0450	0.2695	**0.9006**
NIQE	NR	0.0114	0.1571	0.5062	0.6180	**0.5719**	0.3827	0.2100	0.2327	0.4118	0.1864
QAC	NR	0.0084	0.2058	0.2093	0.1056	0.3794	0.2793	**0.2891**	0.0318	0.0028	0.8263

[4] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," [Online]. Available: http://vision.okstate.edu/csiq

[5] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "MICT image quality evaluation database," [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[6] A. Ninassi, P. Le Callet, and F. Autrusseau, "Subjective quality assessment-IVC database," [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[7] 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.

[8] [online] Available: http://bbs.redocn.com/viewthread.php?tid=93305&extra=page%5C%5C%5C%5C%3D1&page=1

[9] [online] Available: http://www.nipic.com/psd/

[10] [online] Available: http://www.cs.utah.edu//∼reinhard/cdrom/hdr/

[11] ITU-R Recommendation BT. 500, "Methodology for the subjective assessment of the quality of television pictures," 1974-2004.

[12] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?-A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98-117, January 2009.

[13] 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, April 2004.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1398-1402, November 2003.

[15] 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.

[16] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378-2386, August 2011.

[17] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512, April 2012.

[18] 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, January 2013.

[19] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, February 2006.

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

[21] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339-3352, August 2012.

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

[23] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 22, no. 3, pp. 209-212, March 2013.

[24] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," *IEEE Int. Conf. Computer Vision and Patter Recognition*, July 2013.

[25] G. Zhai, X. Wu, X. Yang, W. Lin and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41-52, January 2012.

[26] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced-reference entropic difference for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517-526, February 2012.

[27] M. Narwaria, W. Lin, I. V. McLoughlin, S. Emmanuel, and L. T. Chia, "Fourier transform-based scalable image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364-3377, August 2012.

[28] K. Gu, G. Zhai, X. Yang and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," *IEEE Int. Symp. Circuits and Syst.*, pp. 1095-1098, May 2013.

[29] B. Schökopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207-1245, 2000.

[30] F. Drago, K. Myszkowski, T. Anne n, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419C426, 2003.

[31] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," March 2000, http://www.vqeg.org/