Details preservation inspired blind quality metric of tone mapping methods

Ke Gu, Guangtao Zhai, Min Liu, Xiaokang Yang, and Wenjun Zhang
Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, Shanghai, China, 200240
Shanghai Key Laboratory of Digital Media Processing and Transmissions
{gukesjtuee, zhaiguangtao, liumin_merry, xkyang, zhangwenjun}@sjtu.edu.cn

***Abstract*—High dynamic range (HDR) images are extremely meaningful, especially in the space and medical fields. For visualization of HDR images on standard low dynamic range (LDR) display devices, how to convert HDR to LDR images naturally becomes a valuable issue, which has aroused a variety of tone-mapping operators (TMOs). To compare different LDR images created by distinct TMOs, researchers have recently provided a subject-rated tone-mapped image database, and then developed a full-reference objective tone-mapped image quality index (TMQI) based on the measurement of multi-scale signal fidelity and statistical naturalness. Instead, the basic property of HDR images about details preservation is studied in this paper. With it, a natural inference is that higher-quality tone-mapped images are capable of displaying much more details. We therefore propose a blind quality metric by estimating the amount of details in images generated by darkening/brightening an original tone-mapped images. Experimental results on the above tone-mapped image database confirm that the proposed method, despite of no reference, is robust and statistically superior to the currently optimal full-reference TMQI algorithm, and remarkably outperforms state-of-the-art no-reference IQA metrics.**

***Index Terms*—Image quality assessment (IQA), no-reference (NR), high dynamic range (HDR), tone mapping operator (TMO), details preservation**

I. Introduction

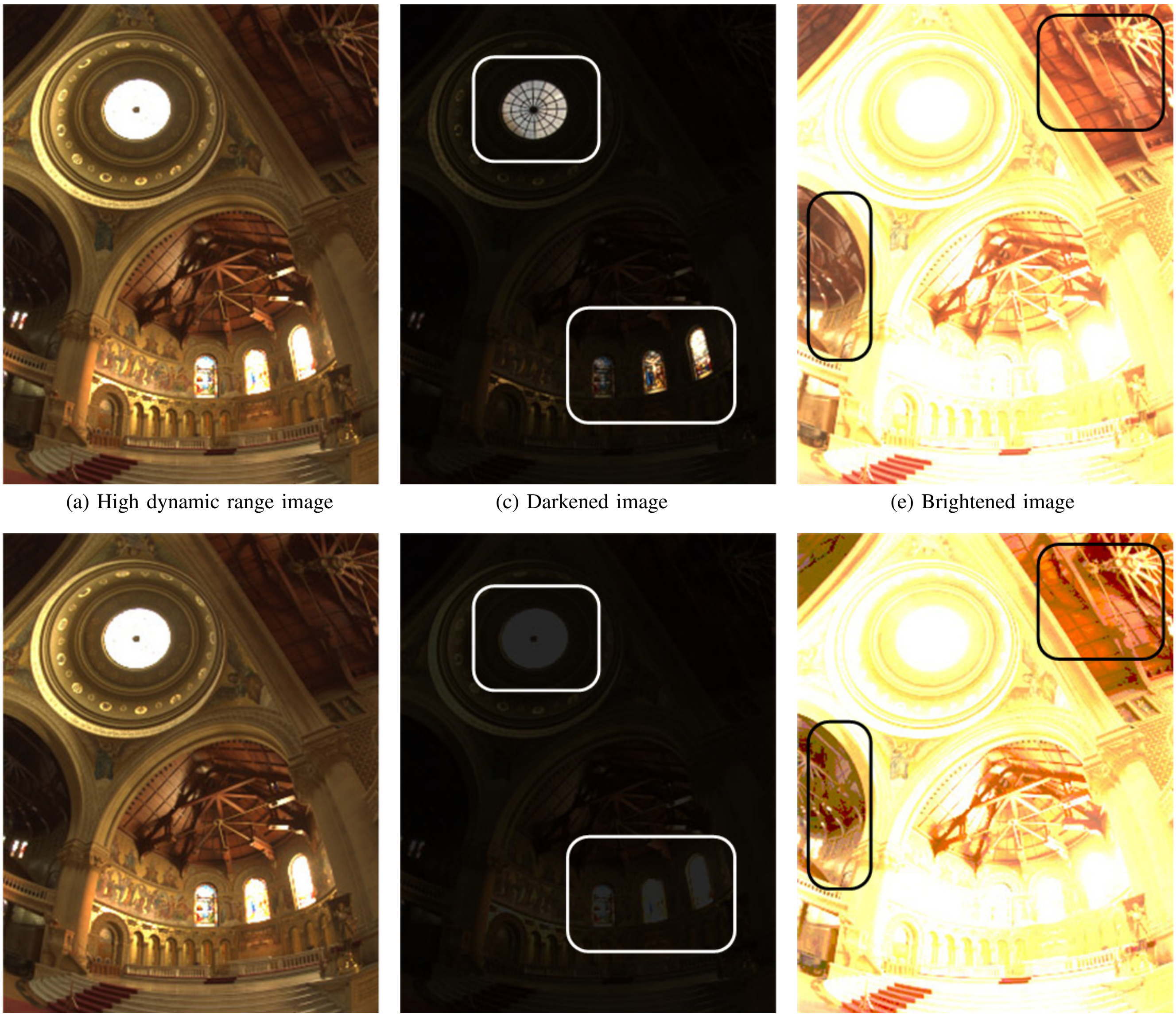
Nowadays, numerous people still make use of low dynamic range (LDR) images and the related 8-bit monitors. Due to only 256 intensity levels involved, LDR images probably lead to many important details missing, particularly in some specific fields, such as space and medicine. Researchers therefore expect high dynamic range (HDR) images to accurately represent the luminance variations, ranging from direct sunlight to faint starlight [1], and adequately protect detailed information. Until now, HDR images have been widely permeated into various kinds of fields. A commonly encountered problem in practical applications is how to well visualize HDR images on standard display devices. Aiming to solve this problem, a growing amount of tone mapping operators (TMOs) converting HDR to LDR images have been exploited so far, such as some previous works in [2]-[4]. Because of the reduction in dynamic range and unideal performance of existing algorithms, TMOs inevitably cause information loss. Therefore, the best tone-mapped images that are transferred from HDR images is still require a human-assisted step, in which subjects compare a large set of distinct LDR products created by different TMOs with different coefficients to pick the most satisfied tone-mapped image.

The method that TMO assessment generally dependeds on human subjective evaluations has lasted for several years. However, the authors in [5] recently pointed out a pair of remarkable limitations of the subjective assessment: 1) it is usually laborious, expensive and time-consuming; 2) it is hard to be used to automatically pick the optimal parameters so as to validly improve TMOs and obtain the best-quality LDR images. The authors therefore provided a subject-rated tone-mapped image database, as well as a novel full-reference (FR) objective tone-mapped image quality index (TMQI) that is illuminated by two successful design principles in image quality assessment (IQA). The first is the modified multi-scale structural similarity (MS-SSIM) [6]. The structural similarity (SSIM) [7] was proposed under the assumption that human visual system (HVS) is highly adapted for extracting structural information from a scene, and has aroused hundreds of IQA metrics [8]-[12]. The second is the statistical naturalness inspired by the important natural scene statistics (NSS) model, which characterizes fundamental attributions of natural images (e.g. their power spectrum is a function of frequency), and in recent years have been devoted to many NSS based no-reference (NR) quality metrics [13]-[15].

In this paper, the basic property of HDR images about details preservation is mainly taken in account. On this basis, we concluded that better tone-mapped images should maintain much more detailed information. Further, it is reasonable to assume that the high-quality converted LDR images are capable of showing more details (i.e. larger entropy value), particularly in the over-dark or over-bright conditions. This paper accordingly proposes a blind quality metric by measuring the entropies in nine images transferred from a tone-mapped image by darkening/brightening its original brightness, followed by implementing support vector machine (SVM) regressor (SVR) [19] to map the above nine entropies to a quality score. Testing on the tone-mapped image database in [5], our blind algorithm leads to the statistically better performance than the optimal FR TMQI method [5] (to date), and clearly outperforms recent NR metrics [13]-[18].

The rest of this paper is arranged as follows. Section II illustrates the basic idea of the proposed blind quality evaluation method in detail. In Section III, a comparison of our algorithm with FR TMQI method and state-of-the-art NR IQA

978-1-4799-3432-4/14/$31.00 ©2014 IEEE



(a) High dynamic range image (c) Darkened image (e) Brightened image

(b) Normal 8-bit image (d) Darkened image (f) Brightened image

Fig. 1. The illustration of the advantage of high dynamic range image [20]: (a)-(b) HDR image and its related normal 8-bit image; (c)-(d) Darkened images of (a)-(b) with 1/64th original brightness; (e)-(f) Brightened images of (a)-(b) with 32 times original brightness.

metrics on the tone-mapped image database [5] is reported and analyzed. Finally, Section IV provides a conclusion and some possible future works.

II. The Proposed Blind Quality Metric of Tone-Mapped Images

The HDR shop [20] offers a comparison between a high dynamic range image and a normal 8-bit image, as shown in Fig. 1. Despite the great difficulty in discerning the difference between an original HDR image and the relavant 8-bit image, their darkened and brightened images, which are darkened to 1/64th and brightened to 32 times original brightness respectively, clearly reveal the fundamental property of HDR images about details preservation. For a convenient display and comparison, we have labeled some noticeable distinguished regions with white rectangles in darkened images (c)-(d) and black rectangles in brightened images (e)-(f).

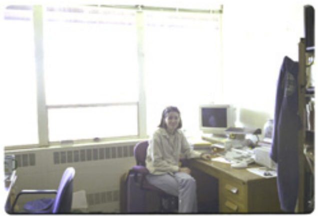
Motivated by the observation above, a direct and simple idea to evaluate the quality of a tone-mapped image is to estimate the volume of detailed information in itself and its converted images by darkening/brightening its original brightness. The transferred images are generated by

$$\mathcal{I}_j = \min(\max(\mathcal{I} \times mul_j, 0), 255) \tag{1}$$

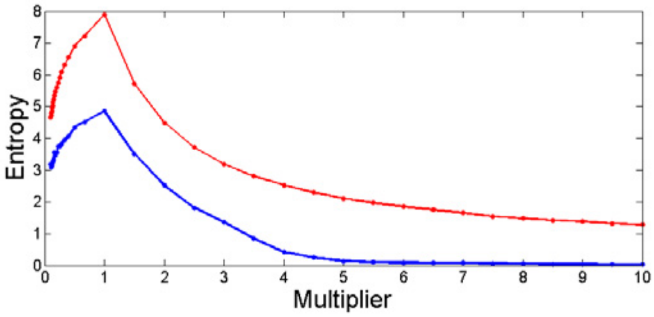
where $\mathcal{I}$ is a tone-mapped image, and mul_j indicates the j-th multiplier. The $\max$ and $\min$ operators are used to clip the transferred image into the range of $0 \sim 255$.

Next, we will show how to measure the details' amount. It is widely known that information entropy is an important concept in statistics [21]. By measuring the average unpredictability

(a)



(b)



(c)

Fig. 2. The illustration of how the entropies $\mathcal{H}$ vary with the changes of the multipliers mul: (a) A high-quality tone-mapped image; (b) A low-quality tone-mapped image; (c) The relationship between the multipliers mul and their corresponding entropies $\mathcal{H}$ in (a) and (b). The red one is for (a), and the blue is for (b).

of an arbitrary signal, entropy represents its disorderly degree. We accordingly apply entropy to quantify the volume of details in the tone-mapped image signal $\mathcal{I}$ and its transferred image signals $\mathcal{I}_j$ as follows

$$\mathcal{H}_j = -\sum_{i=0}^{255} p_i(\mathcal{I}_j) \cdot \log_2 \ p_i(\mathcal{I}_j) \tag{2}$$

where $\mathcal{H}_j$ indicates $\mathcal{I}_j$'s entropy, and $p_i(\mathcal{I}_j)$ is the probability density of i-th grayscale in the transferred image $\mathcal{I}_j$.

Here, we display two relevant tone-mapped images in Fig. 2(a)-(b). Among them, (a) shows a high-quality tone-mapped image, while (b) indicates an over-bright LDR image. We then, for each tone-mapped image, create 37 transferred images with $mul = \{1, n, \frac{1}{n}|n = 1.5, 2.0, ..., 9.5, 10\}$, and compute corresponding 37 entropies. Fig. 2(c) shows how the entropies $\mathcal{H}$ varies with the changes of the multipliers mul. The red and blue curves separately correspond to (a) and (b). It is easy to find that, with a small decrease/increase of luminance in (b), the entropy value quickly falls down to a quite low level, indicating its weak ability for details maintenance. In contrast, (a) presents a good performance to resist the fast fading of entropy and preserve details. Referring to the subjective ratings in [5], (a) really has a higher subjective score than (b).

It is obvious that using 37 entropies as features is too much, making the quality metric being difficult to implement in real time. In our test, we choose only 9 entropies that are measured with $mul = \{1, n, \frac{1}{n}|n = 3.5, 5.5, 7.5, 9.5\}$ as features. In practice, it is observed that increasing the number of features cannot lead to remarkable performance improvement.

A mapping is finally learned from the feature space to quality scores using a regression module, yielding an estimation of tone-mapped image quality. Inspired by recent NR IQA metrics in [13]-[15], this paper also applies the SVR [19] with a radial basis function (RBF) kernel. As a result, the objective quality score of our algorithm is given by

$$q = SVR(\mathcal{H}_j, model) \tag{3}$$

where $model$ is a trained model for regression. More contents about how well the performance of the q score is in terms of the correlation with human opinions will be described in the next section.

III. Experimental Results and Analysis

In this paper, the tone-mapped image database [5] is used as the testing bed. The database is consisted by 8 HDR images and their converted 120 LDR images. And the MOS score for each tone-mapped image is the average value of subjective rankings provided by twenty inexperienced observers. We test and compare the proposed NR IQA method with TMQI [5], DIIVINE [13], BLIINDS-II [14], BRISQUE [15], NFSDM [16], NIQE [17] and QAC [18]. In order to ensure that our approach is robust across image content and is not limited to specific train-test split, this paper reuses random 80%-20% training-test split without overlap for 1000 times. The train-test split in each time is different from each other. Fig. 3 plots standard variances of SROCC results of the proposed blind IQA metric for 1000 training-test iterations on the tone-mapped image database. It is obvious that the performance becomes small and stable after 800 training-test iterations, illustrating that the proposed NR IQA metric is robust across image content, and its performance can be accurately presented by the obtained median performance evaluations.

As suggested by VQEG [22], three significant performance indices of Spearman's rank-order correlation coefficient (S-ROCC), Pearson linear correlation coefficient (PLCC) and Kendall's rank-order correlation coefficient (KROCC) are employed in our research. Note that the PLCC value is computed after nonlinear regression with a commonly used four-

TABLE I
SROCC, PLCC and KROCC values (after nonlinear regression) of TMQI, DIIVINE, BLIINDS-II, BRISQUE, NFSDM, NIQE, QAC, and the proposed algorithm on the tone-mapped image database. Proposed (median) is the median value for 1000 train-test splits.

Metrics	Type	SROCC	PLCC	KROCC
TMQI	FR	0.7715	0.7407	0.5585
DIIVINE	NR	0.3791	0.3681	0.3031
BLIINDS-II	NR	0.5326	0.4683	0.3265
BRISQUE	NR	0.5481	0.4810	0.3351
NFSDM	NR	0.2488	0.2193	0.1794
NIQE	NR	0.5652	0.4968	0.3495
QAC	NR	0.7148	0.5185	0.3595
Proposed (median)	NR	0.8106	0.7683	0.5865

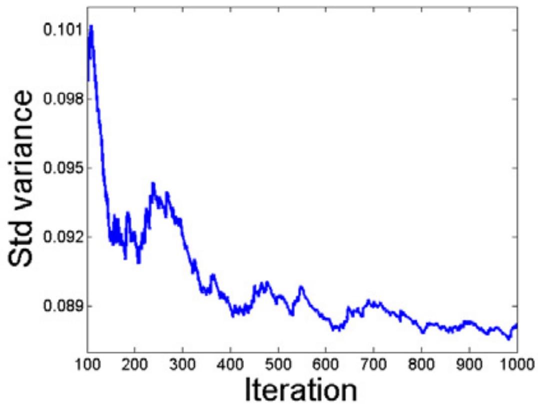


Fig. 3. Standard variances of SROCC values of the proposed blind IQA metric for n-th ($n = 100, ..., 1000$) training-test iterations on the tone-mapped image database.

parameter logistic function like the way in [11]-[12]:

$$Quality(q) = \frac{\gamma_1 - \gamma_2}{1 + \exp(-(q - \gamma_3)/\gamma_4)} + \gamma_2 \tag{4}$$

where q indicates the input score, $Quality(q)$ is the mapped score, and $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ are free parameters to be confirmed during the curve fitting process. A value close to 1 for SROCC, PLCC and KROCC indicates superior correlation with subjective human ratings. Table 1 tabulates the median SROCC, PLCC and KROCC values of our approach for 1000 tain-test splits, and the performance measures of TMQI, DIIVINE, BLIINDS-II, BRISQUE, NFSDM, NIQE and QAC. Clearly, our blind algorithm has resulted in statistically performance than the currently optimal FR TMQI, and surpassed state-of-the-art NR DIIVINE, BLIINDS-II and BRISQUE. Further, the proposed metric depends on merely 9 features (i.e. entropy) with a quite small amount computational loads, making it have a largely possibility to perform in real-time image/video processing systems.

IV. Conclusion

In this paper, we propose a new blind IQA metric in assessing the qualities of tone-mapped images. Based on the basic property of HDR images about details preservation, we extract 9 entropies as features in images transferred from a tone-mapped image by darkening/brightening its original brightness, and then map these features to a quality score using the classical SVR module. Results of experiments on the tone-mapped image database with 1000 training-test splits demonstrate the robustness of the proposed algorithm across image content, and show that it has achieved the statistically superior performance over the currently optimal full-reference TMQI, and noticeable higher prediction accuracy than state-of-the-art no-reference metrics. Further, with only 9 features with very small computational complexity, our algorithm is probably capable of implementing in real time. Despite of its simplicity and effectiveness, our method only considers entropy to measure quality of the tone-mapped images, ignoring the influence of fundamental attributions of natural images, e.g. luminance and contras). The future work therefore will be devoted to a higher-performance blind metric by modifying our algorithm with above important factors.

Acknowledgment

This work was supported in part by NSFC (61221001, 61371146).

References

[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting.* San Mateo, CA: Morgan Kaufmann, 2010.

[2] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Visual. Comput. Graph.*, vol. 3, no. 4, pp. 291-306, 1997.

[3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *Proc. Annu. Conf. Comput. Graph. Interact. Tech.*, vol. 21, pp. 267-276, 2002.

[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419-426, 2003.

[5] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657-667, February 2013.

[6] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conference Signals, Systems and Computers*, November 2003.

[7] 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.

[8] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193-201, April 2009.

[9] 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.

[10] A. Guo, D. Zhao, S. Liu, X. Fan, and W. Gao, "Visual attention based image quality assessment," *Proc. IEEE Int. Conf. Image Process.*, pp. 3297-3300, September 2011.

[11] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Self-adaptive scale transform for IQA metric," *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 2365-2368, May 2013.

[12] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Structural similarity weighting for image quality assessment," *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, pp. 1-6, July 2013.

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

[14] 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.*, pp. 3339-3352, vol. 21, no. 8, August 2012.

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

[16] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, "No-reference image quality assessment metric by combining free energy theory and structural degradation model," *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1-6, July 2013.

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

[18] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," *IEEE CVPR*, 2013.

[19] C-C. Chang and C-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intelligent Systems and Technology*, vol. 2, no. 3, 2011. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

[20] HDR shop. http://www.hdrshop.com/.

[21] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, Octber 1948.

[22] 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/.