Visual Attention Guided Quality Assessment of Tone-Mapped Images using Scene Statistics

Debarati Kundu and Brian L. Evans
Embedded Signal Processing Laboratory
Wireless Networking and Communications Group
The University of Texas at Austin, Austin, TX
Email: debarati@utexas.edu, bevans@ece.utexas.edu

***Abstract*—Measuring visual quality, as perceived by human observers, is becoming increasingly important in the many applications in which humans are the ultimate consumers of visual information. This paper assesses the visual quality in mapping of high dynamic range (HDR) images to standard dynamic range (SDR) images with 8 bits/color/pixel. In previous work, the Tone-Mapped image Quality Index (TMQI) compares the original HDR image with the rendered SDR image. TMQI quantifies distortions locally and pools them by uniform averaging, in addition to measuring naturalness of the SDR image. For SDR images, perceptual pooling strategies have improved correlation of image quality assessment (IQA) algorithms with subjective scores. The primary contributions of this paper are: (1) integrating local information-based pooling strategies in the TMQI IQA algorithm, (2) measuring image naturalness by using mean-subtracted contrast-normalized pixels, and (3) testing the proposed methods on JPEG compressed tone-mapped images and tone-mapped images for SDR displays using subjective scores.**

***Index Terms*—High Dynamic Range Imaging, Tone Mapped Quality Index, Structural Similarity, Visual Saliency, Natural Scene Statistics**

I. Introduction

Recent years have seen a huge growth in the popularity of High Dynamic Range (HDR) images due to their ability to accurately represent the wide range of variation of illumination in real scenes. Unlike traditional Standard Dynamic Range (SDR) scenes, the range of the luminance levels in HDR scenes can range from 10,000 to 1 [1]. HDR rendering also finds its use in computer graphics where the lighting calculations are performed over a wider dynamic range. This results in a better contrast variation leading to a higher degree of detail preservation. However, in order to visualize HDR images on standard displays meant for SDR images, they need to tone-mapped to an SDR image. Different tone-mapping operators (TMO) may result in different SDR visualization, so a natural question is how to evaluate the quality of the images obtained.

Subjective testing is important in evaluating the visual quality of images produced by different algorithms. A faster and less expensive alternative is objective quality evaluation. Recently, full-reference IQA (FR-IQA) algorithms [1] [2] [3] were proposed for evaluating tone-mapped SDR images in comparison to the reference HDR image. In [1], Yeganeh *et al.* carried out a subjective study with various tone-mapped SDR images and proposed the tone-mapped image quality index (TMQI) based on the structural similarity metric in order to ensure that the details in the original HDR image are represented faithfully in the SDR version. It is combined with a naturalness measure based on scene statistics in order to ensure that the rendered image looks realistic.

Tone-mappped FR-IQA metrics employ average pooling that weights all pixels equally. Using different pooling strategies for combining local quality scores to yield the final quality index of the processed SDR image is well-researched [4] [5] [6]. Using perceptual pooling methods for quality evaluation of HDR images is less well studied. In [7], the authors have proposed Saliency weighted Tone-Mapped Quality Index (STMQI) that employ an Attention based on Information Maximization [8] model to find the salient regions of the image. In [9], authors use Spectral Residual [6] and Saliency detection by combining simple priors [10] methods with TMQI.

Petit *et al.* [11] propose a variation of the SDR image saliency measure by Itti *et al.* [12] to make it suitable for HDR images. In this work, we use this method with TMQI for pooling the local quality scores. Although this is found to show good correlation with the ground-truth eye-tracking data obtained from human subjects, computing the saliency map using Gaussian Dyadic Pyramid and Gabor filters is computationaly expensive. To reduce complexity, we also propose simple local information content based pooling strategies that improve the performance of the TMQI algorithm.

We also investigate a natural scene satistics model based on mean-subtracted-contrast-normalized (MSCN) pixels that has been widely used for blind quality prediction of natural SDR images [13] [14]. This does not need any previous training on the corpus of natural images, unlike the model in [1] that fits a Gaussian and a Beta probability distribution to the histograms of the means and standard deviations of these images.

Quality evaluation of compressed tone-mapped images is an emerging problem that involves the joint optimization of tone-mapping and compression parameters. The proposed FR-IQA algorithms have not been evaluated until now for this application. In this paper, we evaluate the performance of the proposed algorithm on multiple artifacts arising from tone-mapping and JPEG compression of HDR images.

II. Tone Mapped Quality Index

The TMQI algorithm is based on the combination of two image quality indicators: 1) a multi-scale image fidelity metric based on a modified structural similarity (SSIM) [15] index and 2) a measure based on natural scene statistics (NSS): the mean and standard deviation of pixel intensities. Since the dynamic range of the HDR images is much higher than that of standard SDR displays, the TMOs cannot preserve all the details of the HDR versions; however it must ensure that the SDR image is structurally similar to the HDR version. The SSIM-inspired component takes into account this aspect of signal fidelity. In addition, the SDR image must also ensure that it looks natural because the human visual system is trained on NSS that appear irrespective of image content. The TMQI algorithm takes into account only the pixel luminances.

A. Structural Fidelity

The SSIM index (and its multi-scale version MS-SSIM) measures changes in luminance, structure and contrast between the images. Tone mapping operators change local intensity and contrast [1], so TMQI redefines the structural fidelity term as:

$$S_{local}(x,y) = \frac{2{\sigma_x}'{\sigma_y}' + C_1}{{\sigma_x}'^2 + {\sigma_y}'^2 + C_1}.\frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2} \tag{1}$$

where x and y are image patches in the HDR and the corresponding tone-mapped SDR image, σ_x, σ_y and σ_{xy} are the local standard deviations and cross-correlations between them, ${\sigma_x}'$ and ${\sigma_y}'$ are the nonlinearly mapped versions of σ_x and σ_y in (2). The algorithm penalizes only those cases where the signal strength is significant in one of the image patches, but insignificant in the other. To distinguish between significant and insignificant signal strength, the local standard deviation in mapped nonlinearly through a psychometric function (related to the visual sensitivity of contrast) which takes the form of a cumulative normal distribution, given by

$$\sigma' = \frac{1}{\sqrt{2\pi}\theta_\sigma}\int_{-\infty}^{\sigma} \exp\left[-\frac{(x-\tau_\sigma)^2}{2{\theta_\sigma}^2}\right] \mathrm{d}x \tag{2}$$

where σ' is the mapped version of σ, τ_σ is the modulation threshold, and θ_σ is the standard deviation of the normal distribution. It is bounded between 0 and 1. τ_σ is proportional to the inverse of the visual contrast sensitivity [16]. At each scale, the scaled version of map is pooled by averaging to output a single score and the overall structural similarity metric is obtained by multiplying the structural similarity scores from the various scales.

B. Image Naturalness

Apart from maintaining structural fidelity, the tone-mapped SDR images should also satisfy some criterion of natural fidelity. In [1], the authors have used naturalness measures based on brightness and contrast of the tone-mapped images. The histograms of the means and standard deviations of natural images have been found to fit a Gaussian and Beta probability distribution respectively. The naturalness measure is the product of these two distributions since natural scene statistics of brightness and contrast are largely independent quantities. The final Tone Mapped image Quality Index (TMQI) is given by:

$$Q = aS^\gamma + (1-a)N^\delta \tag{3}$$

where $0 \leq a \leq 1$ adjusts the relative importance of the structual measure (S) and the naturalness measure (N), and γ and δ control their sensitivities.

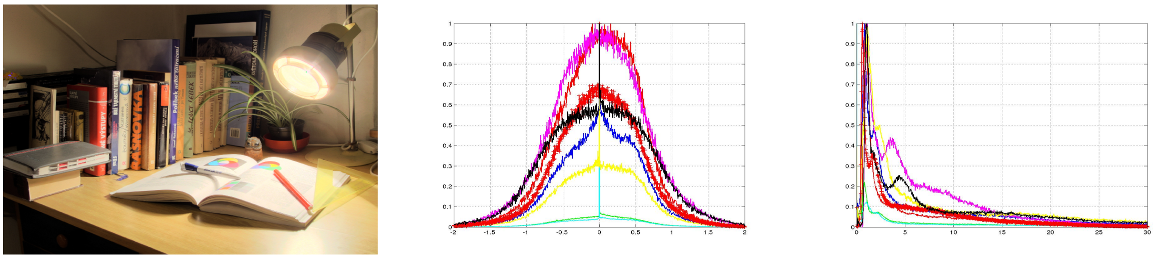
III. Proposed IQA Algorithm

This section outlines the modifications made to the TMQI algorithm to take into account perceptual pooling strategies and a NSS model based on the distribution of the MSCN pixels to quantify image naturalness.

A. Visual saliency measure

In [12], the authors build a master "saliency map" using features like color, intensity and orientations at different scales. Instead of just using intensity differences for HDR images, [11] uses intensity contrast between the scales and normalizes the orientation features also over the intensity channel. We use this method of saliency detection to improve performance of the TMQI algorithm. In addition, we explore the role of the local contrast of the tone-mapped images in pooling the quality scores since the quality of tone-mapped images on SDR displays depends on the degree of detail-preservation. Measures of edge density and local contrast tend to be greater at the points of fixation than at other locations [17] [18]. Regions of higher contrast in the tone-mapped SDR image should be given higher weight in pooling the local structural fidelity score at every scale.

The local contrast of the test image (the tone-mapped SDR image or the JPEG compressed tone-mapped image) is measured with two simple methods. (1) σ-map of the image obtained by (6), and (2) local entropy of the image at every pixel location (using a rectangular window). Since entropy is a measure of uncertainty of the random variables, it can be used to capture the local contrast also. For example, if a tone-mapping operator leads to over-exposed uniformly bright regions (such as the sky), these regions are expected to show higher entropy than a region having aesthetically rendered foliage.



(a) (b) (c)

Fig. 1: (a) An image from the TMQI database [1] and the corresponding histograms of (b) MSCN pixels and (c) σ-field of the tone-mapped SDR images. The figures show how different tone-mapping operators result in different distribution of the MSCN pixels and the σ-field, which can be quantified into the naturalness measure of FR-IQA algorithms.

B. *Natural Scene Statistics*

For this work, we model the scene statistics of tone mapped images in the spatial domain, MSCN pixels and the σ-field of the image. The pixels of the image are preprocessed by mean subtraction and divisive normalization. Let $M \times N$ be the dimension of the image I, and $I(i,j)$ be the pixel value in the (i,j)-th spatial location, $i \in \{1,2,..,M\}$, $j \in \{1,2,...,N\}$. MSCN pixels are generated by

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j)+1} \tag{4}$$

where the local mean $\mu(i,j)$ and standard deviation $\sigma(i,j)$ are defined as:

$$\mu(i,j) = \sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l} I(i+k, j+l) \tag{5}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l}[I(i+k,j+l) - \mu(i,j)]^2} \tag{6}$$

$w = \{w_{k,l} | k = -K,..,K, l = -L,...,L\}$ is a symmetric local convolution window centered at the (i,j)-th pixel. K and L determine the size of local patch considered in the calculation of the mean and standard deviation. In [13], the authors considered 7×7 image patches, and a circularly symmetric 2D Gaussian kernel; however, experiments show that the distribution of the MSCN patches are not very sensitive to the window size, or the convolution kernel.

The variance normalized image $(\hat{I})$ tends to be more uniform than the original image, and almost looks like a noise pattern, except at object boundaries. Also, their histograms seem to show a Gaussian like distribution. The standard deviation image σ looks more like the original image, highlighting object boundaries and attenuating textures. The MSCN pixels have been modeled using an Asymmetric Generalized Gaussian Distribution and used in image quality assessment [13] [14].

As a measure of the image naturalness, we consider the scale parameter of the distribution of the MSCN pixels (β) and standard deviation of the σ-field, obtained from (6). Let ϕ be the variance of the σ field. The modified TMQI index is given by:

$$Q = aS^\gamma + \frac{1}{2}(1-a)\beta^{\delta_1} + \frac{1}{2}(1-a)\phi^{\delta_2} \tag{7}$$

TABLE I: SROCC, PLCC, and KCC between the algorithm scores for various IQA algorithms and the DMOS scores for all images of the TMQI database [1] along with runtime (in seconds). Proposed algorithms are shown in Red. Results for the italicized algorithms have been taken from the original paper. '-' indicates that the corresponding results were not present.

IQA	SROCC	PLCC	KCC	Runtime
TMQI-NSS-σ	**0.8810**	**0.9439**	**0.7857**	0.3212
TMQI-NSS-Entropy	**0.8810**	0.9438	0.7143	1.2759
SHDR-TMQI	**0.8810**	0.9346	0.7143	0.8010
FSITM-TMQI [3]	0.8571	0.9230	**0.7857**	0.9428
STMQI [7]	0.8503	0.9382	0.7638	1.5385
TMQI-SDSP [9]	*0.8408*	-	-	-
TMQI-II [2]	0.8333	0.8790	0.7143	**0.2002**
FSITM [3]	0.8333	0.8948	0.7143	0.4741
TMQI-SR [9]	*0.8170*	-	-	-
TMQI [1]	0.8095	0.9082	0.6429	0.5206

IV. Experimental Results

This section outlines the performance of the proposed algorithms on two HDR datasets. The first one ("TMQI Database") [1] contains 15 reference natural HDR images and 8 tone-mapped SDR images for each of them, generated using different algorithms. The SDR images were ranked according to the quality from 1 (best) to 8 (worst) by 209 subjects. The second one is a tone mapping based HDR compression dataset ("HDR-JPEG Database") [19] comprising of 10 different still images and 14 distorted versions obtained by JPEG compression of the original one with 7 different bitrates and using

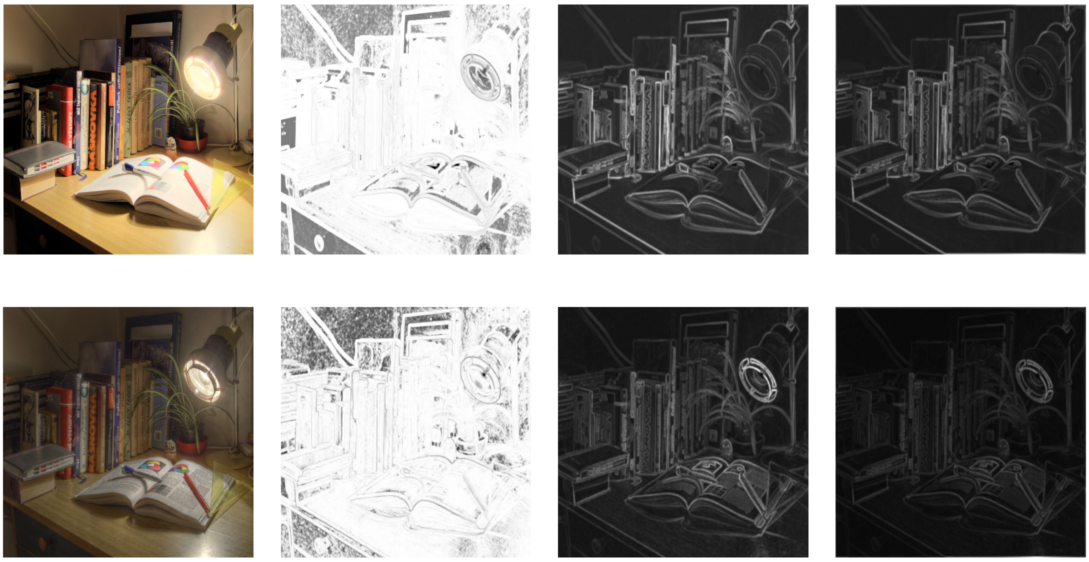


Fig. 2: Results from the TMQI database for outputs of two different tone-mapping operators: Tone-mapped SDR image (column 1), the corresponding local structural similarity map (column 2), the fidelity maps obtained by the product of the structural similarity map and the σ-map (column 3) and that obtained by the product of the structural similarity map and the local entropy map (column 4). Brighter gray level means higher similarity. Results are shown only for the coarsest scale.

two different optimization criteria (Mean Squared Error and the Structural Similarity Index Metric [15]). This database contains both natural and synthetic scenes.

The performance of TMQI, FSITM, TMQI-II, and STMQI FR-IQA algorithms have been evaluated using the MATLAB source codes provided by the authors. TMQI-NSS-σ uses the TMQI index in conjunction with the MSCN based natural scene statistics model and the σ-map as the local pooling strategy. SHDR-TMQI and TMQI-NSS-Entropy employ a similar scheme but use the saliency detection method proposed in [11] and local entropy respectively for pooling the structural fidelity score. These pooling based IQA algorithms employing the MSCN based naturalness measure outperform the state-of-the-art FR-IQA algorithms both for tone-mapping artifacts (Table I)[1] as well as for multiply distorted HDR images having both tone-mapping and JPEG compression artifacts (Table II).

For the different variations of the TMQI algorithm, the relative weightage of the structural similarity term with respect to the naturalness term has been kept constant ($a = 0.8012$). Five levels have been considered for all the IQA algorithms except for SHDR-TMQI, where two levels have been considered in order to ensure that the size of the image do not fall below 128×128; an implementation restriction imposed by the authors of Itti's saliency measure [12]. The source code of our proposed algorithm can be downloaded from [20].

[1]Correlation measures are computed between predicted ranks of tone-mapped images and ground-truth rankings for each source image. Result shows median of correlations computed.

TABLE II: SROCC, PLCC, and KCC between the algorithm scores for various IQA algorithms and the DMOS scores for all images of the HDR-JPEG database [19] along with runtime (in seconds). Proposed algorithms are shown in Red.

FR-IQA algorithms	SROCC	PLCC	KCC	Runtime
SHDR-TMQI	**0.8510**	0.8533	0.6700	3.0003
TMQI-NSS-σ	0.8485	0.8520	0.6659	1.6470
TMQI-NSS-Entropy	0.8454	**0.8645**	**0.6719**	6.7424
TMQI [1]	0.7947	0.8057	0.6127	3.4394
FSITM-TMQI [3]	0.6300	0.6584	0.4762	8.3486
TMQI-II [2]	0.5096	0.5137	0.3642	**1.3424**
FSITM [3]	0.4720	0.5167	0.3422	5.2617
STMQI [7]	0.3464	0.3244	0.2449	11.9965

Spearman Rank-order Correlation Coefficient (SROCC), Pearson's Linear Correlation Coefficient (PLCC) and Kendall's Correlation Coefficient (KCC) have been used to evaluate the performance of FR-IQA algorithms. Execution time (in seconds) for each algorithm (on a Linux desktop having 12 GB RAM, Intel Xeon CPU, 3.33 GHz clock) has also been evaluated. Results for the TMQI and HDR-JPEG Databases are summarized in Tables I and II respectively.

V. Conclusion

We show that simple perceptual pooling techniques that take into account the local contrast improve the performance of the TMQI algorithm for quality evaluation and propose a different NSS model to better qualify the image naturalness. In addition to tone-mapping artifacts, the proposed methods show good correlation with human observers for JPEG compressed tone-mapped images.

REFERENCES

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

[2] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct 2015.

[3] H. Ziaei Nafchi, A. Shahkolaei, R. Farrahi Moghaddam, and M. Cheriet, "FSITM: A feature similarity index for tone-mapped images," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1026–1029, Aug 2015.

[4] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. IEEE Int. Conf. on Image Process.*, Sept 2012, pp. 1477–1480.

[5] 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.

[6] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. IEEE Int. Conf. on Image Process.*, Sept 2012, pp. 1473–1476.

[7] H. R. Nasrinpour and N. D. Bruce, "Saliency weighted quality assessment of tone-mapped images," in *Proc. IEEE Int. Conf. on Image Process.*, Sept 2015, pp. 4947–4951.

[8] N. D. Bruce and J. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, no. 950, June 2007.

[9] X. Liu, L. Zhang, H. Li, and J. Lu, "Integrating visual saliency information into objective quality assessment of tone-mapped images," in *Proc. Springer Intl. Conf. on Intelligent Computing Theory*, 2014, pp. 376–386.

[10] L. Zhang, Z. Gu, and H. Li, "Sdsp: A novel saliency detection method by combining simple priors," in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 171–175.

[11] J. Petit, R. Brémond, and J.-P. Tarel, "Saliency maps of high dynamic range images," in *Proc. ACM Symp. Appl. Perception in Graphics and Visualization*, 2009, pp. 134–134.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998. [Online]. Available: http://dx.doi.org/10.1109/34.730558

[13] 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, Dec 2012.

[14] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, March 2013.

[15] 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.

[16] P. G. J. Baren, *Contrast sensitivity of the human eye an its effects on image quality*. Bellingham, Washington: SPIE Press, 1999.

[17] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, 2002.

[18] P. Reinagel and A. M. Zador, "Natural scene statistics at the center of gaze," *Network: Computation in Neural Systems*, vol. 10, pp. 1–10, 1999.

[19] M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Ppion, "Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality," *Optical Engineering*, vol. 52, no. 10, Oct 2013.

[20] D. Kundu and B. L. Evans, "Full-reference high dynamic range image quality assessment, software release 1.0," January 2016, http://signal.ece.utexas.edu/%7Ebevans/HDRImaging/.