https://doi.org/10.2352/issn.2169-2629.2019.27.45

No-Reference Image Quality Metric for Tone-Mapped Images

Muhammad Usman Khan, Imran Mehmood, Ming Ronnier Luo, Muhammad Farhan Mughal; State Key Laboratory of Modern Optical Instrumentation, Zhejiang University; Hangzhou, China

Abstract

Tone-mapping operators transform high dynamic range (HDR) images into displayable low dynamic range (LDR) images. Image quality evaluation of these LDR images is not possible by comparison with their corresponding high dynamic range images. Hence, a no-reference image quality metric for tone-mapped LDR images is proposed based on the fitting to the present psychophysical results including different visual image quality attributes. Ten images, including HDR natural scenes, were tone-mapped using six TMOs. They were used in the assessment and visual attributes were determined to predict the quality of these images. The visual attributes (brightness and Naturalness) were modeled using parameters derived from CAM16-UCS. Results showed that the quality prediction of the model had a reasonable degree of accuracy.

Introduction

Image quality metric is a mathematical tool to estimate the quality of an image. There are two ways to determine image quality, subjective scaling and objective metrics. The former involves visual assessments of images carried out by human observers based on some predefined judgement criteria. While the subjective assessment is the most reliable way of judging the image quality, it is quite time-consuming, hence not preferred for large image database. Objective metrics, on the other hand, are a fast way of determining image quality but these metrics use only limited number of parameters, i.e., their scope is limited to predicting quality based on limited number of isolated perceptual characteristics of the human visual system. Moreover, objective metrics are less reliable when predicting quality for a variety of test images; hence, such metrics are usually application-specific.

Based on the types of application, different terms are used for objective metrics, but the aim of all these metrics is the same, i.e., to find a measure of deterioration from a reference image. For example, image quality metrics aimed at predicting overall image quality, image difference metrics predict the perceived differences between a reference image and its reproduced degraded image without estimating the effect of these differences on overall image quality, image fidelity metrics predict visibility of image reproduction errors, image similarity metrics measure how test image matches to its reference image.

Objective metrics can be classified into three types [1]: full reference, reduced reference and no-reference metrics. Full reference metrics, as the name suggests, compare parameters from a reference image with those from a test image such as MSE [2], ΔE^*_{ab} [3], S-CIELAB [4], SSIM [5], TMQI [6], FSITM [7] etc. Reduced reference metrics use limited reference parameters from test images. No information is available in case of no-reference metrics; instead, these metrics either work by detecting predefined application specific distortions or they are based on human visual perception.

Full reference metrics can be further classified into three types namely pixel-based, structure-based and perception-based metrics [1]. As the name suggests, pixel-based metrics applied on pixel level. MSE [2], ΔE^*_{ab} [3], and S-CIELAB [4] are a few examples of such metrics.

Structure-based metrics usually based on human visual system to predict the difference between image structures. Structure similarity (SSIM) index [5] compares luminance, contrast and structures of test and reference images to estimate the similarity between two images. While it looks a promising method, SSIM is sensitive to image distortions where visually best-looking images can be ranked lower than the ones which have higher visible distortions [5]. Tone-mapped image quality index (TMQI) [6] is another structure-based metric which combines naturalness with the multi-scale extended approach of SSIM [5] to model different scenarios affecting image quality. Advantage of TMQI over other metrics is that it can be used to test the quality of tone-mapped images by directly comparing them with their corresponding HDR reference images. Feature Similarity Index for Tone-Mapped (FSITM) Images [7] can also compare the tone-mapped images with HDR images and estimate the quality. Locally weighted mean phase angle maps calculated at different scales and orientations for both test and reference images and then compared to get a measure of test image quality.

No-reference image quality metrics follow two trends for the estimation of image quality; 1) to determine predefined distortion ([8] determine jpeg distortions, [9] is an evaluation algorithm for blurring artifacts while [10] is a blind image quality index (BIQI)) or 2) to directly calculate the image quality by considering human visual system characteristics of determining visual image quality attributes such as naturalness, colorfulness, brightness, contrast etc. Such models require the psychophysical data to determine the relevant visual attributes critical in determining image quality.

Choi *et al.* in [11] carried out a psychophysical study and developed an image quality model which was based on contrast, colorfulness and naturalness for large size display. Gong *et al.* [12] conducted the psychophysical analysis using smartphone and tablet displays under different lighting environments to develop a comprehensive model which can predict image quality for different environments. This model [12] considers naturalness and clearness as attributes for determining the quality of natural scenes and colorfulness and clearness for determining the quality of non-natural scene images. Both of these models predict the image quality with better accuracies, but Choi *et al.* [11] model is a full-reference metric while Gong *et al.* [12] model need natural image regions from the reference image to calculate naturalness.

This study is aimed first to conduct an experiment for describing image quality and then to use the data to develop a no-reference image quality model.

Psychophysical Experiment

Ten HDR images from RIT HDR image database [13] tone-mapped using six TMOs (Drago's adaptive logarithmic mapping [14], Reinhard and Devlin's photoreceptor-based operator [15], Reinhard's photographic tone reproduction operators (Local and Global) [16], Schlick's quantization function [17] and Ward's histogram adjustment operator [18]) were used in the experiment. Figure 1 shows ten HDR images tone mapped using Reinhard's local operator. Multiple tone-mapped images

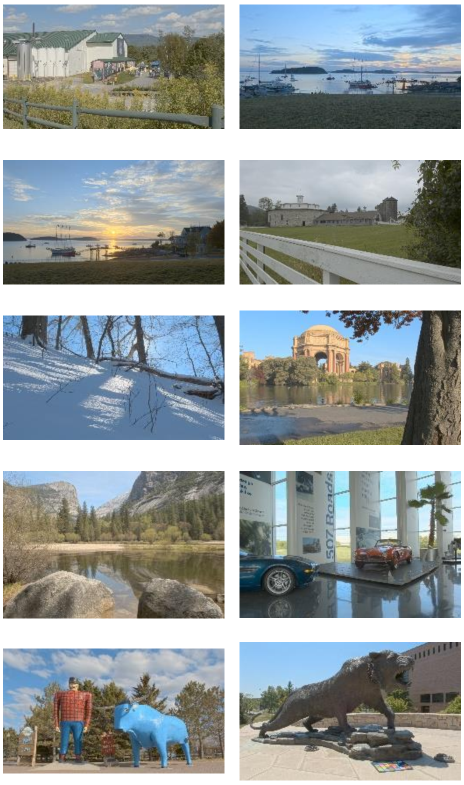

Figure 1. *RIT HDR database [13] images tone-mapped with Reinhard Local [16] tone-mapping operator*

were generated for each TMO using different values of controlling parameters (if there is any variation parameter). Visually best-looking images were selected by experts based on visual assessment of these images where controlling parameters of a specific TMO may be different for each scene.

The test images were displayed in pairs on a calibrated NEC PW272 display with maximum luminance level set to 287 cd/m^2. Twenty observers (12 males and 8 females), aged between 21-33, participated in the experiment. All the observers passed the Ishihara test for color deficiency and they were also trained before conducting the experiment. Example experimental window is shown in Figure 2 where observers evaluated the image pairs on the basis of eight evaluation criteria namely luminance contrast (C_L), color contrast (C_C), brightness (Q), colorfulness (M), shadow details (D_S), highlight details (D_H), naturalness (N) and preference (P). Table 1 lists the attributes together with their description. Pair comparison method was used for evaluation. Observers were asked to decide the quality of a pair of images, e.g., which image is more preferred.

Ten percent of the images were assessed twice to be used to test the intra-observer repeatability. In total, 26,400 image pairs were evaluated (((15 pairs x 10 images) + 150images x 10% repeatability) x 8 attributes x 20 observers) to test the performance of six TMOs and to get the image attributes relevant in determining the image quality.

Results

Firstly, inter and intra-observer variability were calculated using the coefficient of variation. For intra-observer variability, best observer gave 0.07 while worst observer gave 0.33 CV value. For inter-observer variability, best CV value was 0.16 while the the worst was 0.21. This shows that the data was reliable.

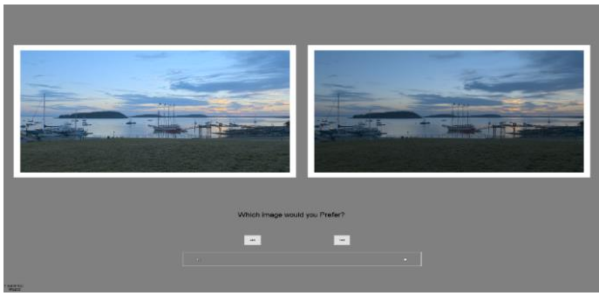

Figure 2. *Experimental window*

Table 1: Definition of image appearance attributes

Attributes	Definition
Luminance Contrast	variation of luminance between different object which makes the image object distinguishable
Color Contrast	variation of colors between different object surfaces which make them distinguishable
Brightness	Visual sensation of an image area which appears to be radiating more or less light
Colorfulness	Visual sensation of an image area which appears to be more or less chromatic
Shadow Details	Visibility of image contents in low light areas
Highlight Details	Visibility of image contents in bright areas
Naturalness	Naturalness is how close the objects in images are to reality.
Preference	Preference is selection of an image by observer based on higher overall image quality

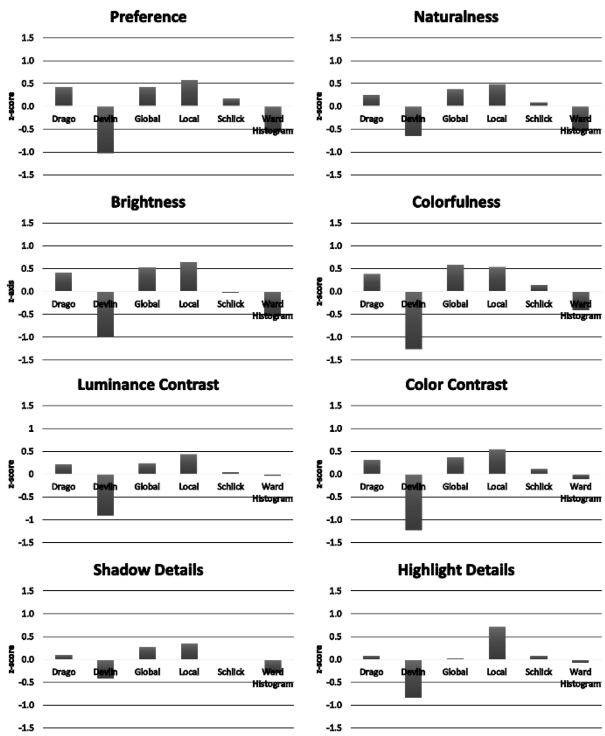

Figure 3. *Average z-scores of six TMOs for all evaluated criteria*

Subjects raw data for each criterion from the psychophysical judgements were transformed to z-scores (higher the better) and plotted in Figure 3. Results suggested that the Reinhard Local operator [16] performed the best in tone reproduction of HDR images in all evaluated criteria except colorfulness (M). This was followed by Reinhard Global operator [16] which performed best in colorfulness, fourth in highlight details and ranks behind Local operator in all other evaluation criteria. Drago's operator [14] ranked second in highlight details and third in rest of all criteria. Schlick's operator [17] ranked third in highlight details while fourth in all other criteria. Ward's [18] and Reinhard-Devlin [15] operators ranked consistently fifth and sixth in all evaluation criteria, respectively.

Preference (P), which is an essential attribute evaluated is correlated directly with overall image quality where the most preferred images would have higher quality than those which were less preferred. When human observers look at images, their decision of preference is highly influenced by a number of visual image appearance attributes like naturalness, colorfulness, contrast, details, brightness, among others. Given the availability of psychophysical data, the relationship between image preference and other attributes can be developed to find out the most influential attributes dictating the image preference.

A model (P) given in equation (1) was developed using a linear regression analysis of psychophysical data to predict preference (P) with an R^2 value of 0.943 and STRESS value is 4.23.

$$P = -0.011C_L + 0.255C_C + 0.233Q + 0.098M \ldots$$
$$-0.021D_S - 0.033D_H + 0.582N - 0.076 \quad (1)$$

The high R^2 between preference (P) and naturalness (N) and brightness (Q) suggests that these are the two most significant attributes contributing in image preference and using only these attributes to determine preference gives R^2 value of 0.93.

Objective Model Development

Since the preference of images is directly related to overall image visual quality, the term "image quality" will be used instead of preference in the objective image quality model. By using the most influential attributes of brightness and naturalness from equation (1), an image quality model was proposed.

$$IQ = -0.014\overline{Q} + 1.313N - 0.177 \quad (2)$$

The constants were determined using the linear regression analysis of psychophysical preference and calculated brightness and naturalness data. Note that the IQ model developed here is based on CAM16-UCS [19] due to its property of high uniformity. $\overline{Q}$ can be directly obtained by taking the mean of brightness Q over entire image pixels.

Naturalness in images is a measure of closeness of image objects (sky, persons, trees, grass, etc.) to their appearance in the real world. It can be modeled using a combination of different image appearance attributes since there is no direct way of measuring naturalness of images. Naturalness model of Choi *et al*. [11] was adopted with slight modifications in the calculation of N in equation (2). Choi *et al*.'s naturalness model [11] was based on image appearance attributes of sharpness, colorfulness and reproduction of shadow details. In the present image quality model, image sharpness was replaced with Luminance contrast (C_L) and changes were also made to the models of colorfulness and shadow details which are given in equations (4)-(7).

$$N = 0.927C_L - 0.012M + 0.965D_S - 0.658 \quad (3)$$

where the constants were determined by the optimization of the naturalness visual data. The model gives R^2 value of 0.68.

Luminance contrast (C_L) in equation (3) was calculated by following the idea of Calabria and Fairchild in [20]. They calculated the standard deviation of image lightness channel, denoted by κ_L, to model the lightness contrast. We calculated standard deviations on lightness J and luminance Y channels of the images and denoted them as κ_{J_w} and κ_{Y_w} respectively, where w denotes the images sub-regions of sizes 5×5, 9×9 and 13×13. All the calculated data were then used to fit the luminance contrast data from our psychophysical experiment. The resulting model to fit the luminance contrast visual data gives R^2 value of 0.32.

$$C_L = 0.79\kappa_{J_5} - 0.080\kappa_{J_9} - 0.513\kappa_{J_{13}} \ldots$$
$$-0.332\kappa_{Y_5} + 0.249\kappa_{Y_{13}} + 0.689 \quad (4)$$

Colorfulness (M) in equation (3) was modeled using the Gong *et al.*'s [12] to fit the colourfulness visual data having an R^2 value of 0.11.

$$M = \frac{a_G}{1 + \exp[-b_G(G_{sRGB} - 1)]} \times \frac{\overline{M}}{m_M} \quad (5)$$

where $\overline{M}$ is the mean of CAM16-UCS colorfulness for the whole image and the constants were determined to be $a_G = 2.1548$, $b_G = 1.2482$ and $m_M = 30.5103$. G_{sRGB} is the gamut area ratio of display to sRGB gamut.

Shadow details (D_S) in equation (3) was modeled using the following equation which gives R^2 value of 0.22.

$$D_S = 0.22D_{S_{13}} - 0.394D_{S_9} + 0.215D_{S_5} - 0.331 \quad (6)$$

where the above D_{S_w} is calculated using the clearness model in [12] which first calculates ΔJ_w, the mean square deviation of edge pixels of the lightness channel from their neighboring pixels in a region of size 5x5. These edge pixels were detected using the Sobel operator [21]. The difference in our approach is that the mean square deviation was calculated for multiple window sizes of 5×5, 9×9 and 13×13, and the edge pixels were detected for lightness values of J≤42 only. J=42 corresponds to the 10% of maximum display luminance level, which was adopted to only target the pixels in shadow regions. For each detected edge pixel, the mean square deviation was calculated with its neighboring pixels using the following equation.

$$\Delta J_w(k, r) = \left(\frac{\sum_{i=-N}^{N}\sum_{j=-N}^{N}[J(k-i, r-j) - J(k, r)]^2}{(2N+1)^2}\right)^{\frac{1}{2}} \quad (7)$$

Note that D_{S_w} is the mean of ΔJ_w over all detected edge pixels. w corresponds to window sizes of 5×5, 9×9 and 13×13 for N=2, 4 and 6 respectively in equation (7).

Model Performance

Merit of the model was expressed using Standardized Residual Sum of Squares (STRESS) [22] and Spearman's Rank Order Correlation Coefficient (SRCC). STRESS is a measure of disagreement between observed and calculated data and it ranges between 0-100% where STRESS value of 0 means that observed and calculated data are in perfect agreement with each other. STRESS value for our model was 11.11 which is considered as a very good agreement with the perceived data. This fact is also shown in Figure 4 which plots the predicted image quality scores vs perceived scores.

Spearman's Rank Order Correlation Coefficient (SRCC) determines the correlation between the ranking of objective and subjective data. It is defined as

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2 - 1)} \quad (8)$$

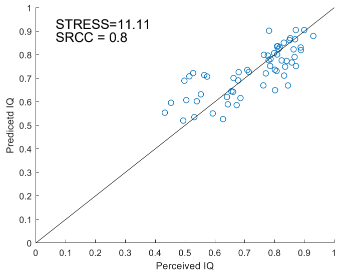

Figure 4. Prediction accuracy of the proposed IQ model

where d_i is the difference between the ranks of i^{th} objective and subjective evaluations. SRCC values of our proposed model was 0.8 which suggest that the model predicted the image quality with reasonable accuracy. The R^2 value of the proposed model was 0.609.

The Gong *et al*.'s IQ model was also tested. Its performance is identical to that of the present model, i.e. STRESS of 11.05 and SRCC of 0.804. Gong *et al*.'s IQ model includes two parts; one for natural scene images (which considers image naturalness and clearness) and the other for non-natural scene images (which considers image colorfulness and clearness). As we used natural scene images in our database, we selected the former model. The naturalness part in the original model relies on finding the color difference between natural object regions (sky, trees, skin, etc.) from reference and test images. The measure of these color differences was then used to get a measure of image naturalness where shifting from reference colors result in lower naturalness values. We modified their IQ model by replacing their naturalness model by equation (3). Modified Gong IQ model also performed equally well as our proposed IQ model.

Conclusion

A no-reference image quality metric was developed. This metric is particularly helpful in situations where reference images are not available or no prior information is available about the test images. Psychophysical analysis of tone-mapped images was carried out, and its data was then used to develop the proposed model which considers brightness and naturalness of an image as main attributes which are critical in determining perceived image quality. CAM16-UCS was used to obtain image appearance Attributes which were then used to determine naturalness and its underlying models and brightness. Results are presented to prove the effectiveness of our proposed IQ metric.

References

[1] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward and K. Myszkowski, High Dynamic Range Imaging: Second Edition (Morgan Kaufmann, Burlington, MA, 2010) pg. 438-439.

[2] Z. Wang and A.C. Bovik, "Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures," IEEE Signal Processing Magazine, vol. 26, no. 1, 98-117 (2009).

[3] CIE, "Colorimetry", Technical Report 15 (2004).

[4] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color-image reproduction," J. of the Society for Information Display, 5(1): 61-63 (1997).

[5] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing, 13(4): 600–612 (2004).

[6] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," in IEEE Transactions on Image Processing, 22(2), 657-667 (2013).

[7] H. Z. Nafchi, A. Shahkolaei and R.F. Moghammadam, "FSITM: A Feature Similarity Index for Tone-Mapped Images", in IEEE Signal Processing Letters, 22(8), 1026-1029 (2015).

[8] Z. Wang, H. R. Sheikh and A. C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, Proc. IEEE Int. Conf. on Image Process., pg. I-477–I-480. (2002).

[9] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," Signal Process. Image Commun., 19, 147–161 (2004).

[10] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," Signal Process. Lett. 17, 513–516 (2010).

[11] S. Y. Choi, M. R. Luo, M. R. Pointer and P. A. Rhodes, Predicting perceived colorfulness, contrast, naturalness and quality for color images reproduced on a large display. 16th Color Imaging Conference, IS&T and SID, pg. 158-164. (2008).

[12] R. Gong, H. Xu, M. R. Luo and H. Li, "Comprehensive model for predicting perceptual image quality of smart mobile devices," Appl. Opt. 54, 85-95 (2015).

[13] M.D. Fairchild, The HDR photographic survey, 15th Color Imaging Conference, IS&T and SID, pg. 233-238. (2007).

[14] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive Logarithmic Mapping for Displaying High Contrast Scenes," Computer Graphics Forum, 22(3): 419–426 (2003).

[15] E. Reinhard and K. Devlin, "Dynamic Range Reduction Inspired by Photoreceptor Physiology," IEEE Transactions on Visualization and Computer Graphics, 11(1):13–24 (2005).

[16] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," ACM Transactions on Graphics, 21(3): 267– 276 (2002).

[17] C. Schlick, "Quantization Techniques for Visualization of High Dynamic Range Pictures," In: G. Sakas, S. Müller and P. Shirley (eds), Photorealistic Rendering Techniques. Focus on Computer Graphics (Tutorials and Perspectives in Computer Graphics). Springer, Berlin, Heidelberg, 1995) pg. 7-20.

[18] G. Ward, H. Rushmeier and C. Piatko, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," IEEE Transactions on Visualization and Computer Graphics, 3(4): 291–06 (1997).

[19] C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill and M. Pointer, "Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS," Color Research & Application, 42(6), 703-718 (2017).

[20] A. J. Calabria, and M. D. Fairchild, "Perceived image contrast and observer preference II. Empirical modeling of perceived image contrast and observer preference data," Journal of Imaging Science and Technology, 47(6), 494-508 (2003).

[21] R. Lukac and K. N. Plataniotis, Color Image Processing: Methods and Applications (CRC Press, 2006).

[22] P. A. García, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," J. Opt. Soc. Am. A 24, 1823–1829 (2007).

Author Biography

Muhammad Usman received his M.Sc. (2011) and M.Phil. (2014) degrees both from Department of Electronics, Quaid-i-Azam University, Islamabad, Pakistan. He is currently doing his Ph.D. in Optical Science and Engineering from Zhejiang University Hangzhou, China. His research interest includes image quality estimation, HDR imaging and tone mapping, color imaging etc.