SALIENCY WEIGHTED QUALITY ASSESSMENT OF TONE-MAPPED IMAGES

Hamid Reza Nasrinpour *

Department of Electrical & Computer Engineering
University of Manitoba
Winnipeg, MB, Canada
hamid.nasrinpour@umanitoba.ca

Neil D. B. Bruce

Department of Computer Science
University of Manitoba
Winnipeg, MB, Canada
bruce@cs.umanitoba.ca

ABSTRACT

Different Tone-Mapping operators (TMOs) produce different Low Dynamic Range (LDR) images based on a single High Dynamic Range (HDR) image. The Tone-Mapped image Quality Index (TMQI) algorithm provides a quantitative means of assessing the quality of resultant LDR images. In this paper we test the hypothesis that TMQI predictions of human image quality can be further aligned with human judgement of image quality in considering visual attention, or regions that humans are predicted to fixate within a scene. We propose a modified version of the TMQI algorithm, a Saliency weighted Tone-Mapped Quality Index (STMQI) which demonstrates higher correlation with subjective ranking scores than the standard TMQI metric.

Index Terms— Image quality, Tone-mapping, High dynamic range, perception, saliency

1. INTRODUCTION

High dynamic range (HDR) images permit accurate representation of natural scenes with a vast range of intensity values. HDR images, benefiting from a very large dynamic range, can capture small changes in luminance or scene irradiance [1]. However, typical display devices cannot capture the complete dynamic range of HDR images [2]. As a result, Tone Mapping Operators (TMOs) are necessary to convert HDR images to LDR images. In comparing existing TMOs, mostly human subjective evaluation has been used so far [2]. Such experiments can be found in [3], [4] and [5].

Inherently, TMOs shrink the dynamic range, which yields information loss and therefore assessment of the quality of the tone-mapped result is of importance. Subjective evaluations are of significant value in this respect, but are time consuming and require new studies as additional algorithms are proposed. For this reason, quantifiable measures that agree with subjective human scores are of value as a means of benchmarking algorithms [2]. Apart from the structural information within an image, high perceptual quality and natural-looking contrast are also desired. A Tone-Mapped image Quality Index (TMQI) has been recently developed by Yeganeh and Wang [2] which integrates these two metrics into a single quality assessment score for an entire LDR image being compared to its corresponding HDR image. One application of such an objective quality assessment measure could be providing a fitness function for parameter optimization algorithms for TMOs, or affording a sense of the quality of the result relative to alternative algorithms.

One limitation in the TMQI is arguably the equal contribution of all regions of a scene to the final quality score. Given that human observers may tend to focus on specific regions of a scene, it is reasonable to assume that the relative importance of different regions in discerning image quality should also consider predicted gaze locations where higher quality may be more important. We therefore test the role of saliency in boosting the performance of the TMQI technique. Accordingly, we propose a modified version of TMQI [2] (STMQI), incorporating saliency based on Attention based on Information Maximization (AIM) [6], to address this issue. This choice is premised on the strong ties of this algorithm to both information theory, and agreement with patterns observed in human visual psychophysics experimentation [7].

The remainder of the paper is organized as follows. In section 2, an overview of the AIM saliency algorithm is provided based on [7], [8], and [9]. Section 3 explains the standard TMQI and proposed modified STMQI approach. We then present results obtained in section 4. Finally, Section 5 discusses conclusions from this study and fruitful directions for future work.

2. VISUAL SALIENCY

Images tend to have regions that draw visual attention when examined by a human observer. There are many algorithms for predicting regions likely to be fixated by human observers which might be used to determine the relative importance of regions of an image. As the central hypothesis of the current work lies in examining whether predictions of human judge-

*The authors gratefully acknowledge financial support from the NSERC Discovery Grants program and the University of Manitoba for GETS funding.

ments of image quality may be augmented by saliency, we focus on one particular algorithm (AIM) [9, 8] that performs well across a range of benchmarks studies, and also has strong ties to behavior observed in visual psychophysics studies [7]. This corresponds to regions in a scene that contain the most informative patches within the image [9] in a Shannon sense [10]. In the AIM algorithm patch likelihoods are determined by density estimation on basis coefficients that characterize each patch of an image derived via Independent Component Analysis (ICA). This representation is reminiscent of local Gabor-jets and colour-opponent cells observed in early visual areas of the primate brain.

For experimental results in this paper, a set of 25 ICA basis functions corresponding to a window size of 21×21 have been used. As the patches are independent, the product of likelihood of all the coefficients corresponding to a local image patch yields the joint likelihood. At the final step, the saliency map is obtained by transforming the joint likelihood into Shannon's measure of self-information by $-log(p(x))$, where $p(x)$ is overall likelihood [7]. Some examples of AIM saliency maps are illustrated in Fig. 1.

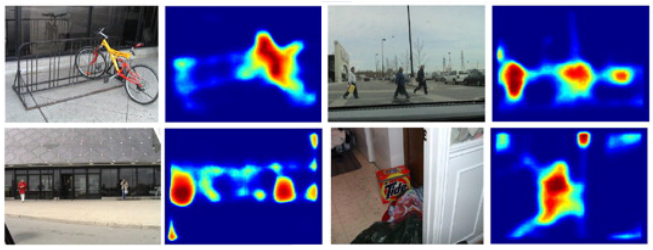

Fig. 1. Examples of (contrast equalized) saliency maps from the AIM algorithm. [7].

3. STANDARD AND PROPOSED MODIFIED TMQI

The standard TMQI algorithm [2] is built upon two design principles in the Image Quality Assessment (IQA) literature: Natural Scene Statistics (NSS) [11] and multi-scale Structural SIMilarity (SSIM) approach [12]. The NSS component measures the extent to which the resultant LDR image is natural-looking in its appearance. SSIM compares the structure and signal strength between LDR and reference HDR image. Either module is applied only on the Y component of the image, after converting LDR image from RGB to Yxy colour space. The combination of the NSS and SSIM measures are then combined, resulting in a Tone Mapped image Quality Index (TMQI) as $Q = a \times S^{\alpha} + (1-a) \times N^{\beta}$ where S and N are SSIM and NSS scores, respectively, and the other parameters are to adjust relative weight and sensitivity of the two components. We have used the same parameters used in [2], $a = 0.8012$, $\alpha = 0.3046$ and $\beta = 0.7088$ such that the direct benefits of saliency within the modified metric are easily discernible.

3.1. Statistical Naturalness

Statistical naturalness model in the standard TMQI is build based on brightness and contrast [2]. Yeganeh and Wang [2] selected approximately 3,000 8bits/pixel grayscale images from [13, 14]. Histograms of the means and standard deviations of these images were fit by a Gaussian (P_m) and Beta (P_d) probability density function respectively. The parameters of the fit model were estimated by regression and can be found in [2]. For a given Y channel of an LDR image, the mean and standard deviation are calculated with their respective likelihoods determined by the two fit probability density functions (P_m and P_d). These two quantities reflect the global intensity and contrast of an image [2]. As brightness and contrast are independent [15], their joint probability density function is their product. Therefore the NSS measure is the product of the two previously calculated quantities divided by the normalization factor of $\max(P_m, P_d)$. This normalization factor guarantees that the NSS score N, lies between 0 and 1. The NSS component in the proposed STMQI is identical to that in the standard TMQI given that there is no spatial component.

3.2. Structural Fidelity

Preserving the structural details of an HDR image within the tone-mapped LDR image is a key property for a tone-mapping operator. However, all information in HDR images cannot be transferred into the LDR version, due to compression of the intensity range [2]. In this sense, structural fidelity is of paramount importance for tone-mapped images.

In the standard TMQI, structural fidelity between HDR and LDR image is measured by a multi-scale variant of SSIM approach. Assuming x and y are two corresponding local patches from HDR and LDR images, respectively; local structure fidelity measure is defined as

$$S_{local}(x,y) = \frac{2\sigma'_x\sigma'_y + C_1}{{\sigma'_x}^2 + {\sigma'_y}^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2} \qquad (1)$$

where σ_x, σ_y and σ_{xy} are local standard deviations and cross correlation between the two patches extracted from the HDR and LDR image, respectively; σ'_x and σ'_y are mapped versions of σ_x, σ_y; C_1 and C_2 are called stabilizing constants [2]. Compared to the original SSIM algorithm [12], the luminance component is dropped; the structure component (i.e. the second term of Equation 1) is the same; and the contrast component has changed. TMOs change local intensity and contrast of the image; so direct comparison of local intensities between HDR and LDR images is meaningless [2]. As such, any differences in signal strength between the HDR and LDR image should not be penalized as is the case in the original SSIM. Accordingly, in this variant of SSIM, the algorithm

penalizes only those cases where the signal strength is above a visibility threshold in one of the image patches, but below the visibility threshold in the other image patch. Thus a nonlinear map is used to calculate σ' based on σ in such a way that "significant signal strength is mapped to 1 and insignificant signal strength to 0, with a smooth transition in-between." [2]. The exact calculations and formula for this nonlinear mapping function can be found in [2].

Local structure fidelity measure, S_{local}, is determined locally through a sliding window. A Gaussian window of size 11×11 with standard deviation of 1.5, suggested by the original SSIM algorithm [12], was employed to this end. This operation produces a map which represents the variation of structural fidelity throughout the image [2].

In the standard TMQI algorithm, this map is then pooled by simple averaging to produce a single score; whereas in the proposed STMQI the pooling strategy is based on the information conveyed by each local patch. Therefore the sliding window measure of structural integrity is applied to the HDR and LDR images with an additional weighting factor w_a corresponding to the information content of the patch as determined by AIM. The overall quality as determined by structural integrity, therefore allocates more weight to the important and informative parts of an image. Raw AIM saliency scores are scaled to lie within $[0, 1]$ with the saliency weighted mean SSIM over the entire image providing a single score of structural integrity.

Visible details and structure of an image are different at each scale level. As a result, for both TMQI and STMQI, images are iteratively low-pass filtered and downsampled to form a Laplacian image pyramid structure [16]. Moreover, the AIM saliency map is also iteratively low-pass filtered and downsampled to accompany the images at each scale. At each level, l, the structural fidelity score, S_l, is calculated, and the overall SSIM score combines all these scores together using the method in [17] as $S = \prod_{l=1}^{L} S_l^{\beta_l}$ where $L = 5$ is the total number of scales and $\{\beta_l\} = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ is the weight for each scale as reported in [17]. All the other parameters used in the proposed modified TMQI are identical to the parameters used in the original TMQI.

4. SIMULATION STUDIES

The performance of the proposed STMQI was evaluated against the standard TMQI using the dataset provided by [2]. The dataset contains a set of 15 images of indoor and outdoor scenes. Each set includes 8 LDR images generated from a single HDR image. Using default parameters, three built-in Adobe Photoshop TMOs of "Exposure and Gamma," "Equalize Histogram," and "Local Adaptation" were employed to generate three LDR images within each set. The other five images were created by five different TMOs developed by Reinhard *et al.* [18], Drago *et al.* [19], Durand & Dorsey [20], Mantiuk *et al.* [21] and Pattanaik *et al.* [22]. One set of example images is shown in Fig. 2. Twenty subjects were asked to rank the 8 images within each set from 1 to 8. The mean ranking score for each image is reported in the dataset [2]. Therefore a rank-order correlation between our ranking results and subjective rankings can be used as an evaluation metric.

Fig. 2. Sample LDR images generated by different TMOs based on a single HDR image from the dataset in [2].

The two following evaluation metrics, as employed in the original TMQI [2], are applied:

1. Spearman's rank-order correlation coefficient (SRCC) as

$$SRCC = 1 - \frac{\sum_{i=1}^{N} d_i^2}{N\left(N^2-1\right)} \tag{2}$$

where d_i is the difference between i-th image's ranking in the subjective and objective methods.

2. Kendall's rank-order correlation coefficient (KRCC) as

$$KRCC = 1 - \frac{N_c - N_d}{N\left(N^2 - 1\right)} \quad (3)$$

where N_c and N_d are the number of pairs in the correct (concordant) and incorrect (discordant) orders, respectively.

The results are given in Table 1 and Table 2. Table 1 reports SRCC metric for the standard TMQI algorithm and four other proposed STMQIs, each of which incorporates the AIM saliency map in a different manner. Table 2 has the same structure but uses the KRCC evaluation metric in scoring. The details of experimentation on these four additional STMQI metrics follow.

Table 1. SRCC Metric for the Standard TMQI and its Variations

Image Set	Original TMQI	LDR Saliency	Pooled Saliency	HDR Saliency	HDR^5 Saliency
1	0.9048	0.8810	0.8810	0.8810	0.9524
2	0.7857	0.7619	0.7857	0.7857	0.7857
3	0.8095	0.8095	0.8333	0.8333	0.8333
4	0.8810	0.8810	0.9048	0.8810	0.8810
5	0.7381	0.7381	0.7381	0.6667	0.7857
6	0.9762	0.9762	0.9762	0.9762	0.9524
7	0.6905	0.6429	0.6429	0.7857	0.7857
8	0.7143	0.7143	0.7143	0.7143	0.7381
9	0.6905	0.6905	0.6905	0.6905	0.8571
10	0.9286	0.9286	0.9286	0.9048	0.9762
11	0.8810	0.8333	0.8810	0.8810	0.8810
12	0.7143	0.7143	0.7143	0.7143	0.7143
13	0.7143	0.7143	0.7143	0.8333	0.8810
14	0.7381	0.7381	0.7381	0.8333	0.9286
15	0.9524	0.9762	0.9762	0.9762	0.9762
Mean	**0.8080**	**0.8000**	**0.8080**	**0.8238**	**0.8619**

Table 2. KRCC Metric for the Standard TMQI and its Variations

Image Set	Original TMQI	LDR Saliency	Pooled Saliency	HDR Saliency	HDR^5 Saliency
1	0.7857	0.7143	0.7143	0.7143	0.8571
2	0.6429	0.5714	0.6429	0.6429	0.6429
3	0.6429	0.6429	0.7143	0.7143	0.7143
4	0.7143	0.7143	0.7857	0.7143	0.7143
5	0.6429	0.6429	0.6429	0.5000	0.5000
6	0.9286	0.9286	0.9286	0.9286	0.8571
7	0.5714	0.5000	0.5000	0.6429	0.6429
8	0.5714	0.5714	0.5714	0.5714	0.6429
9	0.5714	0.5714	0.5714	0.5714	0.7857
10	0.8571	0.8571	0.8571	0.7857	0.9286
11	0.7143	0.6429	0.7143	0.7143	0.7143
12	0.5714	0.5714	0.5714	0.5714	0.5714
13	0.5714	0.5714	0.5714	0.7143	0.7857
14	0.6429	0.6429	0.6429	0.7143	0.8571
15	0.8571	0.9286	0.9286	0.9286	0.9286
Mean	**0.6857**	**0.6714**	**0.6905**	**0.6952**	**0.7429**

In the LDR Saliency, the w_a values are calculated based on the LDR image being compared to the reference HDR image. On first inspection, it may be surprising that average performance is worse than the standard TMQI. As the reader may have noticed in Fig. 2, LDR images may have quite different appearance (and salient regions). A human observer ranking the images sees a variety of tone-mapped versions, and as such the true salient regions of the image might be different form the AIM saliency map for a particular LDR given that information is already discarded. For this reason, w_a values based on AIM for a single LDR image fails to improve upon the standard TMQI. As an alternative test, one might consider averaging computed saliency across the 8 different LDR tone-mapped images to derive the weighting w_a. The Pooled Saliency column reflects this test and the correlation score for this algorithm is only on par with the standard TMQI. Moreover, this strategy cannot be applied in practice, as in practical applications usually one LDR image is provided to be evaluated. The most natural approach is to compute the saliency and associated weighting with complete information from the HDR image (HDR Saliency). As can be seen in Tables 1 and 2, this yields an STMQI metric that outperforms the standard TMQI on average.

In applying a saliency driven weighting of structural integrity, the relative contrast (or contrast of the weighting map) is an important factor, so a natural extension to consider is of the form w_a^γ to adjust the contrast of the weighting. Therefore the final approach, namely, HDR^5 Saliency employs a power law transformation with $\gamma = 5$ to the multi-scale saliency weighting maps. This yields a higher emphasis on regions expected to draw attention from human observers in the weighting matrix for the STMQI algorithm. A γ value of 5 was found to be optimal in testing, and the average performance of the STMQI with the non-linearity is much better than all alternatives. The significant positive value as an optimal parameter further suggests that spatial weighting in line with expected viewing behaviour is a very important factor in deriving high-quality performance metrics.

5. CONCLUSION

The TMQI model allows for quantitative assessment of the quality of tone-mapped images using their accompanying HDR images as references. In this paper, we have demonstrated that inclusion of prediction of salient regions of an image in tone-mapped image quality assessment, results in an improved tone-mapping quality metric. The proposed STMQI shows strong correlation with subjective evaluations of image quality, and achieves higher Spearman and Kendall rank-order correlation coefficients than that of the standard TMQI metric. Future development of this work is directed at assessing alternative saliency measures and towards more sophisticated measures of *naturalness* based on both global and local image statistics towards stronger accordance with human decisions.

6. REFERENCES

[1] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[2] Hojatollah Yeganeh and Zhou Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, 2013.

[3] Jiangtao Kuang, Hiroshi Yamaguchi, Changmeng Liu, Garrett M. Johnson, and Mark D. Fairchild, "Evaluating hdr rendering algorithms," *ACM Trans. Appl. Percept.*, vol. 4, no. 2, July 2007.

[4] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. P. Seidel, "Dynamic range independent image quality assessment," in *ACM SIGGRAPH 2008 papers*. ACM, 2008, p. 69.

[5] Akiko Yoshida, Volker Blanz, Karol Myszkowski, and Hans peter Seidel, "Perceptual evaluation of tone mapping operators with real-world scenes," in *Human Vision & Electronic Imaging X, SPIE*, 2005.

[6] Neil D. B. Bruce and John K. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, no. 9, 2007.

[7] Neil Bruce and John Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 1–24, 2009.

[8] Neil D. B. Bruce and John K. Tsotsos, "Saliency based on information maximization," in *Advances in Neural Information Processing Systems 18*. 2006, pp. 155–162, MIT Press.

[9] Neil D. B. Bruce, "Features that draw visual attention: an information theoretic perspective," *Neurocomputing*, vol. 65-66, pp. 125–133, 2005.

[10] Claude E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[11] Zhou Wang and A.C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, 2011.

[12] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions On Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[13] "Computer vision test images, http://www.cs.cmu.edu/ cil/v-images.html," 2005.

[14] Gerald Schaefer and Michal Stich, "Ucid - an uncompressed colour image database," in *In Storage and Retrieval Methods and Applications for Multimedia 2004, volume 5307 of Proceedings of SPIE*, 2004, pp. 472–480.

[15] Valerio Mante, Robert Frazor, Vincent Bonin, Wilson Geisler, and Matteo Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nat Neurosci*, vol. 8, no. 12, pp. 1690–1697, Dec. 2005.

[16] P.J. Burt and E.H. Adelson, "The laplacian pyramid as a compact image code," *Communications, IEEE Transactions on*, vol. 31, no. 4, pp. 532–540, 1983.

[17] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik, "Multi-scale structural similarity for image quality assessment," in *in Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, (Asilomar*, 2003, pp. 1398–1402.

[18] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, July 2002.

[19] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, pp. 419–426, 2003.

[20] Frédo Durand and Julie Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, July 2002.

[21] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Percept.*, vol. 3, no. 3, pp. 286–308, July 2006.

[22] Sumanta N. Pattanaik, Jack Tumblin, Hector Yee, and Donald P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 2000, SIGGRAPH '00, pp. 47–54, ACM Press/Addison-Wesley Publishing Co.