OBJECTIVE ASSESSMENT OF TONE MAPPING ALGORITHMS

Hojatollah Yeganeh   and   Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: hyeganeh@uwaterloo.ca, zhouwang@ieee.org

ABSTRACT

There has been a growing interest in recent years to develop tone mapping algorithms that can convert high dynamic range (HDR) to low dynamic range (LDR) images, so that they can be visualized on standard displays. With a number of tone mapping algorithms proposed, a natural question is which one gives the best performance. Although subjective assessment methods provide useful references, they are expensive and time-consuming, and are difficult to be embedded into the design stage of tone mapping algorithms for optimization and parameter tuning purposes. This paper focuses on objective assessment of tone mapping operators. Inspired by the success of the structural similarity index method for image quality assessment, we propose a new objective assessment algorithm that creates multi-scale similarity maps between HDR and LDR images. Our experiments show that the proposed method correlates well with subjective rankings of existing tone mapping operators. Furthermore, we demonstrate how the proposed algorithm can be employed in an existing tone mapping algorithm for optimal parameter tuning.

Index Terms— image quality assessment, high dynamic range image, tone mapping, structural similarity

1. INTRODUCTION

High dynamic range (HDR) images allow for a greater dynamic range of luminance levels between its brightest and darkest regions than standard dynamic range (SDR) or low dynamic range (LDR) images. The wider dynamic range enables more accurate representations of the intensity levels in the real scene. A common problem that is often encountered in practice is how to visualize HDR images using standard LDR displays. For this purpose, a number of tone mapping algorithms have been proposed that convert HDR to LDR images, for example [1–4]. With multiple tone mapping operators (TMOs) at hand, some questions naturally arise: Which TMO creates the best quality LDR image? Since the reduction of dynamic range inevitably causes information loss, which LDR image most faithfully maintains the information in the HDR image?

The most straightforward method to assess TMOs is subjective evaluation. In [5], perceptual evaluations were performed for six TMOs with regard to similarity and preference. In [6], seven TMOs were compared using two architectural interior scenes. Fourteen observer were asked to rate basic image attributes as well as the naturalness of the LDR images. In [7], one of the most comprehensive subjective comparisons of TMOs was carried out, where ten subjects were asked to rate the brightness, contrast, details, colors, as well as overall quality of the LDR images generated by fourteen TMOs. These subjective tests create precious references in studying the performance of tone mapping algorithms. However, subjective approaches are limited in several aspects. First, they are often expen-

sive and time-consuming. Second, they are difficult to be incorporated into the design and optimization of tone mapping algorithms. Third, it may not be appropriate to treat the subjective evaluation data as the "golden" standard. For example, some detailed structures in an HDR image may be mapped by a TMO to a flat region in the LDR image, thus are invisible to human subjects. This is apparently not a desired attribute of TMO, but would not be reflected by subjective ratings simply because subjects do not know the existence of such structures.

Little has been done in developing objective methods to assess TMOs. In [8], a dynamic range independent image quality assessment algorithm was proposed based on a visibility model of the human visual system. When comparing a pair of HDR-LDR images, the algorithm produces quality maps, which show good correlations with subjective classifications of image distortion types (blur, sharpening, contrast reversal and no distortion) [8]. Since no spatial pooling is performed, this approach does not provide a single quality score of an entire image, making it impossible to be validated with subjective evaluations of overall image quality.

The purpose of this work is to develop an objective quality measure of a given LDR image using the original HDR image as a reference. Our work is motivated by the success of the structural similarity (SSIM) index [9], which is based on separated comparisons of the luminance, contrast and structure of local image patches. It has been shown to correlate well with perceived image quality when tested with a number of large-scale subject-rated independent databases. Its performance is further improved by a multi-scale version of SSIM [10]. Direct use of SSIM to compare HDR and LDR images, however, is problematic [8], because the luminance and contrast comparisons in SSIM are invalid across different dynamic ranges. In this paper, we propose a novel method that follows the basic principle of SSIM to capture structural information loss, but modifies the contrast comparison component to allow for comparisons across varying dynamic ranges. Our method creates an overall quality score of the entire image as well as multi-scale quality maps that indicate local quality variations across scale and space. This makes it easy to be verified with subjective evaluations. In addition, we demonstrate that the proposed method can be directly employed for the optimization and parameter tuning of tone mapping algorithms.

2. PROPOSED METHOD

2.1. Local Quality Assessment

Our approach follows the philosophy of the SSIM approach, which is based on the assumption that the main purpose of vision is to extract structural information from the visual scene, and thus a measure of structural information loss should provide a good prediction of perceived image distortion. The original local SSIM definition

includes a luminance, a contrast and a structure comparison components. Since the local luminance and contrast between HDR and LDR images are meant to be different, it is not appropriate to compare local luminance and contrast the same way as in the original SSIM. Let x and y be two local image patches extracted from the HDR and LDR images, respectively. We define our local similarity measure as

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

The second term is the same as the structure comparison component in SSIM, where σ_x, σ_y and σ_{xy} are the local standard deviations and cross correlation between the two patches, respectively, and C_2 is a positive constant to avoid instability when the denominator is close to 0. The first term gives a modified local contrast comparison from SSIM. The reasoning behind its construction is as follows. First, the contrast differences between HDR and LDR image patches should not be penalized as along as their contrasts are both significant or both insignificant (this is different from comparing images with the same dynamic range as in SSIM, where any change in contrast is penalized). Second, the measure should penalize the cases that the contrast is significant in one of the image patch, but not the other. A critical issue here is to quantify the significance of local contrast. To do this, we let the local standard deviation pass through a nonlinear mapping function given by

$$\sigma' = \begin{cases} 0, & \sigma < T_1 \\ \frac{1}{2}\left\{1 + \cos\left[\frac{\pi}{T_2 - T_1}(\sigma - T_2)\right]\right\}, & T_1 \le \sigma \le T_2 \\ 1, & T_2 < \sigma \,, \end{cases} \qquad (2)$$

where T_1 and T_2 are two threshold values that divide the ranges of insignificant and significant contrasts, and a raised cosine function is employed to provide a smooth transition between the two types of ranges. Note that when two image patches are both significant (σ greater than T_2) or both insignificant (σ smaller than T_1), the first term of Eq. (1) equals 1, and thus the S_{local} measure is fully determined by the structure comparison component in Eq. (1).

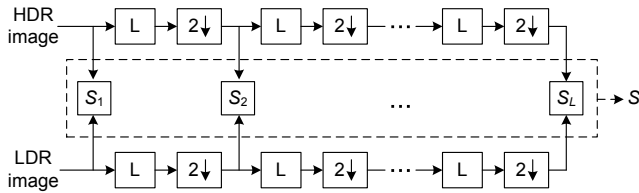## 2.2. Multi-scale Measurement

Fig. 1. Multi-scale diagram of the proposed method.

The local S_{local} measure described above is applied to an entire image using a sliding window approach across the image. This results in a quality map that indicates the quality variation across space. As in [10], we adopt a multi-scale approach, as depicted in Fig. 1, where the images are iteratively low-pass filtered and downsampled. The local quality map is generated at each scale, and the map is then averaged to provide a single score for the scale by

$$S_l = \frac{1}{N_l} \sum_{i=1}^{N_l} S_{\text{local}}(x_i, y_i) \,, \qquad (3)$$

where x_i and y_i are the i-th image patches and N_l is the number of patches in the l-th scale. Examples of the proposed multi-scale computation are shown in Fig. 2. Finally, the single-scale measures are combined to a multi-scale measure of the overall image quality:

$$S = \prod_{l=1}^{L} S_l^{\beta_l} \,, \qquad (4)$$

where L is the total number of scales and β_l is the weight assigned to the l-th scale.

There are several parameters in the implementation of the proposed algorithm. When computing S_{local}, we set $C_1 = 0.01$, $C_2 = 10$, $T1 = 0.5$, and $T_2 = 4$, respectively. To create the quality map at each scale, we employ a Gaussian sliding window of size 11×11 with standard deviation 1.5. When combining the measures across scales, we set $L = 5$ and $\{\beta_l\} = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ (which follows the subjective measures reported in [10]), respectively. The proposed algorithm is computationally efficient. The Matlab implementation of the full computation for comparing a pair of HDR-LDR images takes less than 1 second on a regular PC.

2.3. Comparison with Subjective Evaluations

To verify the proposed algorithm, we compare it with the subjective evaluation results of TMOs reported in [7], which are available online at [11]. In particular, we employed the overall quality ranking test results [7], where LDR images created by 14 TMOs were ranked by 10 naive subjects for overall image quality. Figure 3 shows the scatter plot of the proposed measure S against subject ranking, where rank numbers 1 and 14 correspond to the best and worst quality, respectively. It can be observed that the proposed measure generally agrees quite well with subjective rankings. To the best of our knowledge, no other TMO quality measure creates overall quality scores (e.g., the method in [8] generates quality maps only), and therefore no other method was included for comparison.
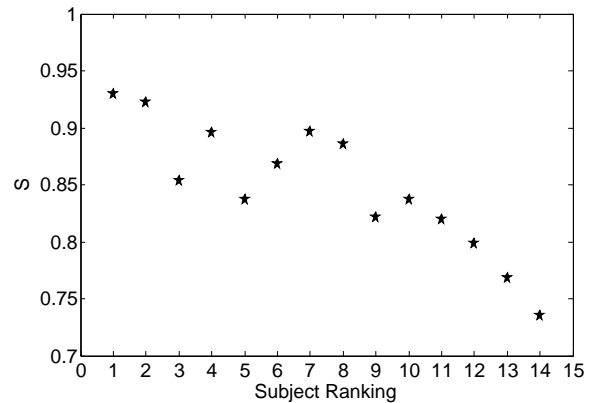

Fig. 3. The proposed quality measure versus subjective ranking for LDR images created by 14 TMOs in [7, 11].

To have a better understanding of the proposed measure across different scales, we compute Spearman's rank correlation coefficient (SRCC) and Kendall's rank correlation coefficient (KRCC) between subjective rankings and both single-scale (for 5 scales) and multi-scale measures in the proposed algorithm. The results are shown in Table 1. It appears that our multi-scale approach improves upon all single-scale measures.
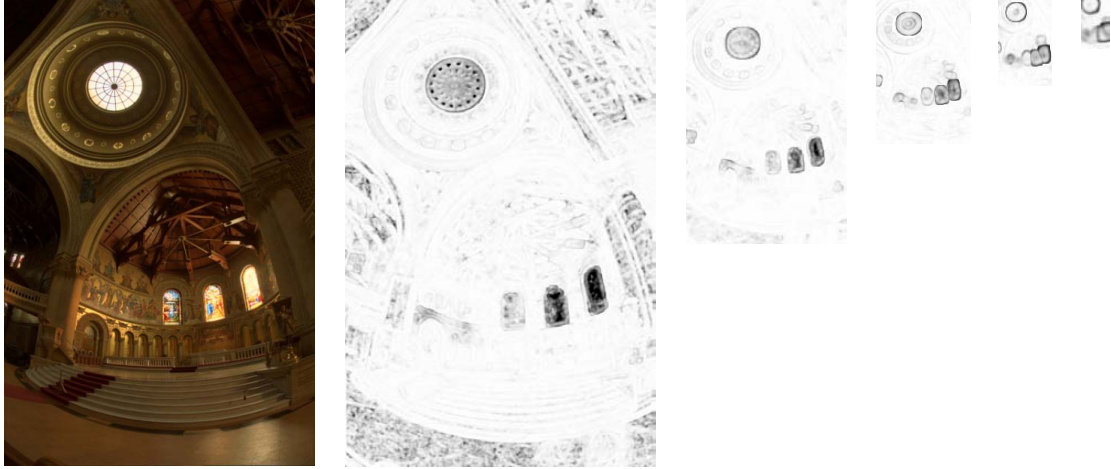
(a) $S = 0.9288$ ($S_1 = 0.9371$; $S_2 = 0.9642$; $S_3 = 0.9524$; $S_4 = 0.9158$; $S_5 = 0.8286$)

(b) $S = 0.7980$ ($S_1 = 0.8419$; $S_2 = 0.8573$; $S_3 = 0.8330$; $S_4 = 0.7795$; $S_5 = 0.6361$)

Fig. 2. LDR images and their quality maps and scores in five scales. The images were created using Adobe Photoshop "Highlight compression" and "Exposure and Gamma" methods (not optimized for quality), respectively. The structural details of the brightest regions are missing in Image (b), but are more visible in Image (a). These are clearly reflected in the quality maps.

Table 1. SRCC and KRCC results between subject rankings and single-scale (for 5 scales) and multi-scale approaches of the proposed method.

	SRCC	KRCC
S_1	0.5165	0.3590
S_2	0.6209	0.4359
S_3	0.8242	0.7077
S_4	0.8060	0.6436
S_5	0.7308	0.5897
S	0.8571	0.7179

3. PARAMETER TUNING IN TMO ALGORITHM

The value of objective image quality measures is beyond evaluating images and algorithms. An interesting extended application is to employ them in the optimal design and parameter tuning of TMO algorithms. To demonstrate this, we use the proposed measure to tune the adaptive logarithmic tone mapping algorithm proposed in [4]. Specifically, the algorithm was define by

$$L_\text{d} = \frac{L_\text{dmax} \cdot 0.01}{\log_{10}(L_\text{wmax} + 1)} \cdot \frac{\log(L_\text{w} + 1)}{\log\left(2 + \left(\left(\frac{L_\text{w}}{L_\text{wmax}}\right)^{\frac{\log(b)}{\log(0.5)}}\right) \cdot 8\right)},$$

(5)

where L_w and L_wmax are world luminance and maximum luminance of the scene, and L_d and L_dmax are display luminance and maximum luminance of display, respectively, and b is a tuning parameter. Figure 4 plots how the proposed measure varies as a function of b from 0.1 to 2.0. From the figure, $b = 0.9$ is picked as the optimal setting. Similar shape of curves has been seen in our other experiments as well, but with different optimal points. Figure 5 shows the LDR
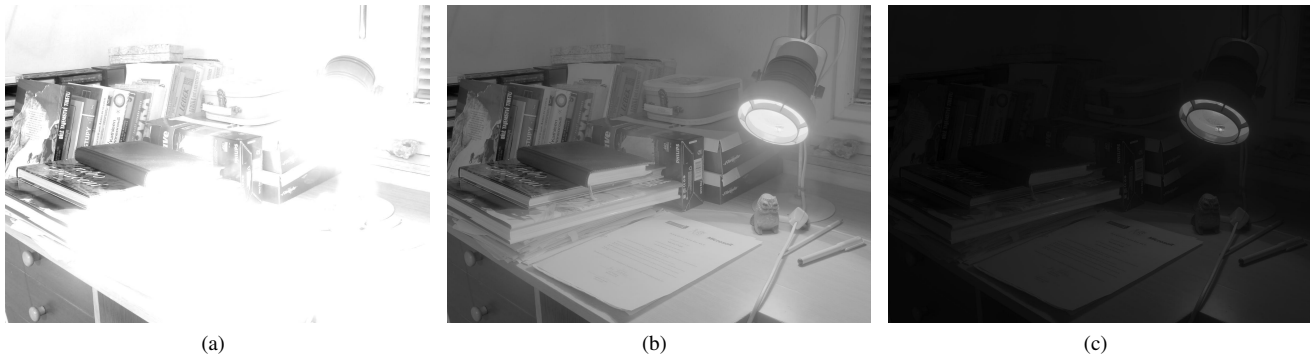
(a)	(b)	(c)

Fig. 5. LDR images generated with different parameter b in Eq. (5). (a) $b = 0.1$, $S = 0.4331$; (b) $b = 0.9$, $S = 0.9515$; (c) $b = 2.0$, $S = 0.8088$.
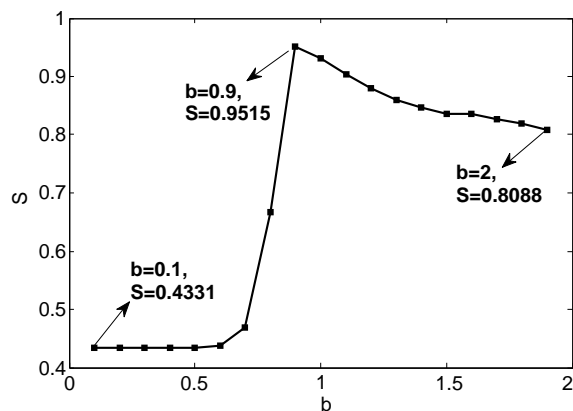

Fig. 4. The proposed quality measure S as a function of parameter b in Eq. (5).

images created using three different b values.

4. CONCLUSION AND FUTURE WORK

We propose an objective method to assess the quality of LDR images created from HDR images by following the design principle of the multi-scale SSIM approach. Our experiments show that the proposed measure agrees well with subjective rankings of overall image quality. We also demonstrate the usefulness of the proposed measure in optimal parameter tuning of tone mapping algorithms. The proposed algorithm is computationally efficient and supplies multi-scale quality maps that indicate local quality variations across both scale and space. These promising results demonstrate the potentials of the proposed approach as one of the initial attempts in objective quality evaluation of tone-mapped images. Future work includes more careful parameter calibrations, and the development of advanced spatial pooling strategies that combine local quality maps to an overall quality score.

5. ACKNOWLEDGMENT

6. REFERENCES

[1] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *in Proc. of 29th annual Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, vol. 21, pp. 267–276, 2002.

[2] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.

[3] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *29th Annual Conf. Computer Graphics & Interactive Techniques*, pp. 249–256, 2002.

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

[5] F. Drago, W. L. Martens, K. Myszkowski, and S. H. P, "Perceptual evaluation of tone mapping operators," *In Proc. Of the SIGGRAPH Conf. Sketches and Applications*, 2003.

[6] A. Yoshida, V. Blanz, K. Myszkowski, and H. Seidel, "Perceptual evaluation of tone mapping operators with real-world scenes," *Human Vision and Electronic Imagin X, SPIE*, vol. 5666, pp. 192–203, 2005.

[7] M. Cadik, M. Wimmer, L. Neumann, and A. Artusi, "Image attributes and quality for evaluation of tone mapping operators," *in Proc. Of Pacific Graphics*, pp. 35–44, 2006.

[8] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. P. Seidel, "Dynamic range independent image quality assessment," *International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, 2008.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, pp. 600–612, 2004.

[10] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *in Proc. of 37th Asilomar Conf. Signals, Systems and Computers*, 2003.

[11] M. Cadik *et al.*, "Evaluation of tone mapping operators," http://www.cgg.cvut.cz/members/cadikm/tmo.