Blind Quality Index for Tone-mapped HDR Images and Multi-Exposure Fused Images

Jingli Fang and Mingxing Jiang

1 School of Animation and Digital Art, Anhui Sanlian University, Hefei, Anhui, 230601, China
2 School of Information Engineering, Anhui Institute of International Business, Hefei, Anhui, 231131, China
* Corresponding author’s e-mail: jmx0551@gmail.com (Mingxing Jiang)

Abstract. Tone mapping operators (TMOs) and multi-exposure fusion algorithms (MEFs) are employed to visualize the informative contents of high dynamic range (HDR) images on the standard dynamic range (SDR) devices, but also lead to visual quality degradation. Thus, it is urgent to evaluate the perceptual quality of tone-mapped images (TMIs) and multi-exposure fusion images (MEFIs). This paper proposes a blind quality index for TMIs/MEFIs by capturing exposure, colorfulness and structure characteristic features. The connection between feature space and associated subjective ratings is established via a regression model. The proposed method is tested on the public database and compared to the state-of-the-art blind image quality assessment (IQA) methods. The experimental results show that the proposed method achieves superior performance.

1. Introduction

Compared with standard dynamic range (SDR) display, high dynamic range (HDR) imaging technology has a wider color gamut and high dynamic range [1]. However, HDR imaging and display devices are still rare. To capture the HDR imaging experience on the SDR device, tone mapping operators (TMOs) and multi-exposure fusion algorithms (MEFs) are employed to reproduce scenes with HDR. The former is convert HDR images to SDR images by the tone mapping operators (TMOs) [1]. The latter generate an SDR image by using sequential stack of different exposed SDR images [2]. However, different tone-mapping (TM) and MEFs affect the TMI/MEFI on multiple attributes (e.g., structure, naturalness, color, etc.), and quality degradation is caused by a mixture of artifacts. Therefore, it is necessary to exploit more effective image quality assessment (IQA) methods to monitor and gauge the quality of tone-mapped images (TMIs) and multi-exposure fusion images (MEFIs).

As shown in the Figure 1, the difference of TMIs/MEFIs in exposure and color distortion is obvious after the tone-mapping or multi-exposure fusion. The subjective score (MOS) indicates better image quality.

![Figure 1. Six images from ESPL-LIVE HDR database [6]. Higher MOS indicates better image quality.](image-url)
We can see that Fig. 1 (a2) and Fig. 1 (b1) preserve more details and color information of indoor or outdoor scene than others. Focusing on the details and contrast preservation, existing IQA methods for TMIs/MEFIs mainly exploit new natural scene statistics (NSS) features in spatial [3] and transform domains [4] [5] for evaluating the image naturalness. Other researchers incorporate image structure [5] and texture [6] to enrich the structural fidelity measure. For example, Gu et al. [3] proposed a blind tone-mapped quality index (BTMQI) by analysing the information entropy, statistical naturalness, and detail preservation of the TMI/MEFI. Kundu et al. [5] designed new space-domain NSS features and HDR-specific gradient-based features to train an effective IQA model (i.e., HIGRADE), which verify the superiority on the ESPL-LIVE HDR database [7]. Yue et al. [6] combined the structural, textural and color information to simulate these responses of opponent cells. Although promising performances have been achieved, most methods focus on TMI/MEFI attributes (e.g., structure, naturalness or colorfulness) for synthetical distortions. However, they seldom consider local specific artifacts difference (i.e., abnormal exposure and halo effect) related with the TMOs/MEFs. In this paper, we propose a blind image quality assessment (IQA) method to evaluate TMI/MEFI quality which makes use of exposure and colorfulness attributes.

2. Proposed method
The framework of the proposed method is depicted in Fig. 2. As a common pipeline, it consists of feature extraction module and quality prediction module. First, three groups of quality-sensitive features (i.e., exposure, colorfulness, and structure) are extracted from each image in the image set. Then, a feature vector combined of these features are aggregated into a quality score via the support vector regression (SVR) which reveals the connection between feature space and human mean opinion score (MOS). The quality score of the test image is obtained directly through the quality score prediction model.

2.1. Exposure and Colorfulness features
For a TMI/MEFI, abnormal exposure is a very obvious distortion that appears in the TMI/MEFI. For a TMI/MEFI, over-exposure in general determines the degree of distortions [8]. Traditional methods usually use a simple threshold (a certain percentage of maximum luminance) to detect over-exposure. If the pixel value is greater than or equal to the threshold, the pixel is considered over-exposed. However, this simple threshold method does not consider the gradual transition from the overexposed area to the adjacent area. To accurately detect over-exposure area, the over-exposure map [9] (M>0.5) is adopted. The experimental validation leads to selection of the LAB-based method [9] as a more effective over-exposure detection metric. Here, L represents L* channel and C = (a b)c represents a* and b* channels of the TMI/MEFI. The over-exposed map M is generated via the joint calculation of L and C values. The pixel is defined as overexposed point, when L is larger or ||C|| is smaller. Formally, Mi is expressed as

$$M_i = 0.5 \left( \tanh \left( \delta (L_i - L_0) + (C_i - \|C_i\|_2) \right) + 1 \right).$$

(1)
where $L_U$ and $C_U$ represent the boundary value of the overexposure area. We set $\delta = 0.01$, $L_U = 70$ and $C_U = 50$ in this study. Fig. 3 depicts the over-exposure map of Figs. 1(a1)-(a3) from ESPL-LIVE database. Since over-exposure information is vital for the quality perception, over-exposure ratio (the number of overexposed pixels divided by the total pixels) and over-exposure entropy are first extracted to measure the impact of over-exposure on image quality.

For areas other than overexposure (marked as $\Omega$), color distortion usually presents the status of the subject. For the $\Omega$, the feature extraction should emphasize the measurement of color cast. The colorful index is expressed as

$$C_{ab} = \log\left(\frac{\delta_a}{\mu_a}\right) \log\left(\frac{\delta_b}{\mu_b}\right),$$

where $\delta_a, \delta_b, \mu_a$ and $\mu_b$ are define as the variance and mean of A and B channels of the $\Omega$.

2.2. Structure features

With massive edges in TMIs/MEFIs, it is very important to analyze the microstructure of TMI/MEFI. Texture is the spatial distribution and dependence between gray levels in the local area. To precisely describe the microstructure of each TMI/MEFI, we employ CS-LBP [10] and GLCM [11] to extract textural information of TMIs/MEFIs. Comparing the LBP with 256 different binary patterns, the CS-LBP produces 16 different binary patterns of encoding local area by threshold the difference of gray value of four pairs center-symmetric pixels as a binary number. Next, we convert the TMI into a GLCM map, and extract three features that reflect edge characteristics: energy, contrast, correlation, and uniformity.

In practice, HVS is not sensitive to the distortion of high-frequency signals, and more sensitive to the distortion of low-frequency signals. After the TM/MEF, edge information loss is inevitable, and halo artifacts [12] appear. To measure the halo effect, we utilize the Canny edge detector (with the default settings) to obtain edge map as well as the gradient map. Fig. 4 shows three edge maps of Figs. 1(b1)-(b3) computed via the Canny edge detector. By observation, Fig. 4(a1) shows more clear structures than Fig. 4(a2) and Fig. 4(a3). Considering that the halo effect makes the edges blurred and scattered to the surrounding area, surrounding 8x8 image patch for all edge point are calculated to measure the edge dispersion. To be special, the mean, standard deviation, peak value, skewness of 8x8 patch for each edge point are used as the statistical characteristics to quantify the halo effect.
3. Experimental results

3.1. Database and Evaluation methodology

The ESPL-LIVE HDR database consists of 1811 images which are divided into 3 types, i.e., tone mapping (TM), multi-exposure fusion (MEF) and post processing (PP). Generally speaking, two important indicators (i.e., prediction accuracy and monotonicity) are used to indicate the performance of an objective IQA method. In our experiments, three criteria are used for this purpose. The first and second indices are the Pearson Linear Correlation Coefficient (PLCC) and Root Mean-Squared Error (RMSE) are used to measure the prediction accuracy and consistency, and Spearman Rank Correlation Coefficient (SRCC) is selected to estimate the prediction monotonicity. The larger PLCC and SRCC values indicate better performance, while the RMSE is the opposite.

3.2. Performance Evaluation Results

To ensure the results are not biased to specific train-test splits, we follow other training-based methods by randomly repeating 1000 times. For the evaluation on the ESPL-LIVE HDR database, we randomly split the database into two subsets, 80% for training, 20% for testing, respectively. Table 1 shows the experimental results, which are denoted in forms of median value. From the table, some meaningful phenomenon can be intuitively observed. First of all, the proposed method performs better than those recently developed blind IQA metrics, including BTMQI, HIGRADE-2, Yue’ method and BLIQUE-TMI. Next, the proposed method and HIGRADE-2 represent the best performance on blind IQA methods of TMIs/MEFIs. In terms of overall performance, our proposed method is more suitable for measuring TMI/MEFI quality.

Table 1. Performance comparison on the ESPL-LIVE HDR Database.

| Method     | NRSL | BTMQI | HIGRADE-2 | Yue’ method | BLIQUE-TMI | Proposed |
|------------|------|-------|-----------|-------------|------------|----------|
| PLCC       | 0.5516 | 0.6175 | 0.7304    | 0.7019      | 0.7010     | 0.7400   |
| SRCC       | 0.5415 | 0.6152 | 0.7295    | 0.6915      | 0.6960     | 0.7309   |
| RMSE       | 8.3291 | 7.8674 | 6.8354    | 7.1327      | -          | 6.7786   |

![Figure 5. Scatter plots of correlation between objective score and MOS using different categories from the ESPL-LIVE HDR database: (a) overall map, (b) TM map, (c) MEF map and (d) PP map.](image)

3.3. Performance Comparison on Individual Distortions

In addition, the scatter plots of MOS values versus objective scores predicted by the proposed method for each distortion type (e.g., RamanTMO, ReinhardTMO and WardHistAdjTMO) on the ESPL-LIVE HDR Database are shown in Fig. 5. The linear correlation between MOS values and predicted scores demonstrates the great monotonicity and accuracy of the proposed method. Fig. 5(a) describes the overall correlation distribution, and Fig. 5(b)-(d) describe the TM, MEF and PP categorie respectively. We use different colored points represent 11 distortion types (three categories as mentioned above) in each figure. By observing, four feature maps prediction scores via the proposed correlate well with the MOSs, especially TM categorie. However, PP categorie performs poorly. These results can be further used as reference basis for parameter adjustment of real-time monitoring systems.
4. Conclusion
This paper proposes an effective blind quality assessment algorithm for the TMIs and MEFIs. The vital importance is to extract effective features from multiple image attributes by analyzing characteristics of TMIs/MEFIs. Specifically, three aspects (exposure, colorfulness and structure) are measured to characterize the global and local information of the TMI/MEFI perceived by HVS. The image exposure ratio and entropy, colourful index are calculated to represent the distortion caused by exposure. Next, CSLPB and GLCM are used to measure image structure distortion. All the extracted global and local features constitute a feature vector, and a prediction model is built via SVR to learn a quality predictor from feature space to quality space. Experimental results have proved that the proposed method has better quality evaluation capabilities of TMIs/MEFIs than other blind IQA ones in evaluating TMIs/MEFIs.

Acknowledgments
The authors would like to thank the editors and the anonymous reviewers for their constructive comments and suggestions, which greatly helped in improving this paper. This work was supported in part by the School Fund Project: Research on the Artistic Effect of Combination of Real Scene Shooting and Animation Technology (Grant No. zsy21003).

References
[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, (2010) High dynamic range imaging: Acquisition, display, and image-based lighting. Morgan Kaufmann.
[2] R. Hassen, Z. Wang, M. M. Salama (2015) Objective quality assessment for multiexposure multifocus image fusion, IEEE Trans. Image Process., 24: 2712-2724.
[3] K. Gu, S. Wang, G. Zhai, and S. Ma (2016) Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure, IEEE Trans. Multimed., 18: 432–443.
[4] Q. Jiang, F. Shao, W. Lin, G. Jiang (2019) BLIQUE-TMI: blind quality evaluator for tone-mapped images based on local and global feature analyses, IEEE Trans. Circuits Syst. Video Technol., 29:323-335.
[5] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, (2017) No-reference quality assessment of tone-mapped HDR pictures, IEEE Trans. Image Process., 26: 2957-2971.
[6] G. Yue, C. Hou, K. Gu, S. Mao, W. Zhang (2018) Biologically Inspired Blind Quality Assessment of Tone-Mapped Images, IEEE Trans. Ind. Electron., 66: 2525-2536.
[7] D. Kundu, D. Ghadiyaram, A. C. Bovik, B. L. Evans (2017) Large-scale crowdsourced study for tone-mapped hdr pictures, IEEE Trans. Image Process., 26: 4725-4740.
[8] B. Hoefflinger (2007) High-dynamic-range (HDR) vision. Springer.
[9] D. Guo, Y. Cheng, S. Zhuo, and T. Sim (2010) Correcting over-exposure in photographs, in: IEEE CVPR, pp. 515-521.
[10] M. Heikkila, M. Pietikainen, C. Schmid (2009) Description of interest “ areas with local binary patterns,” Pattern Recognit., 42: 425-436.
[11] R. Haralick, K. Shanmugam, I. Dinstein (1973) Textural features for image classification,” IEEE Trans. Syst., Man, Cybern., Syst., SMC-3: 610–621.
[12] Y. Li, L. Sharan, E. H. Adelson (2005) Compressing and companding high dynamic range images with subband architectures, ACM Trans. Graph., 24: 836-844.