Dense Scale Invariant Feature Transform-Based Quality Assessment for Tone Mapping Image

Liyan Zhang and Hualin Ma*
Zhejiang Business Technology Institute, Ningbo 315012, China
*Corresponding author

***Abstract*—High dynamic range (HDR) images give the human visual system (HVS) a better visual experience due to their wide luminance range. However, traditional display devices can't capture such a large luminance dynamic range, which needs to be remedied by tone mapping (TM) operation. In addition, TM images usually lose a lot of detail information in low luminance and high luminance areas, So a TM image quality evaluation method based on dense scale invariant feature transform (DSIFT) is proposed in this work. First, the DFIFT descriptors of the HDR image and the TM image are extracted respectively, and the local similarity is calculated to represent the detail loss of the TM image. Then, the local quality map is refined by using the Gauss exposure curve because HVS is more sensitive to the detail loss in low-dark and high-light areas of the TM images. Finally, considering the color distortion characteristics of TM image, an objective perception quality evaluation model is established. Experimental results on a public database demonstrate that the proposed method is in good agreement with human visual perception.**

Keywords—high dynamic range; tone mapping; quality evaluation

I. INTRODUCTION

Real world scenes usually have a very large luminance dynamic range, and HDR images can better reflect the real feelings of HVS[1]. However, due to the limitation of the dynamic range in conventional display devices, it is impossible to represent all the luminance ranges in the real world. Therefore, it is necessary to compress the luminance range of HDR images by tone mapping technology. Compared with low dynamic range (LDR) images, TM images usually have richer detail information especially in high-light and low-dark areas[2]. In addition, the subjective quality of TM images varies greatly with different TM operators, mainly in contrast, color and detail. Therefore, the quality evaluation of TM images is one of the important problems to be solved in the imaging fields[3].

At present, most of the full-reference quality evaluation methods for ordinary LDR images mainly include RFSIM[4], FSIM[5], GMSD[6], MDSI[7], RVSIM[8], etc. Although the sensitive features for common distorted images are extracted such as gradient, phase consistency and contrast, the distortion characteristics of TM images are not taken into account, so the above methods are not directly applicable to TM image quality evaluation. TM images usually have richer detail information in low and high luminance areas, but the image quality will also be significantly different mainly reflected in color distortion and detail loss because of the performance difference of TM operators. Furthermore, different from the traditional full-reference image quality evaluation methods, the luminance dynamic range of the reference image and the distorted image is quite different and can't be directly compared for TM image quality assessment. Recently, according to the characteristics of TM image, relevant research teams have proposed some objective evaluation models for TM image and confirmed that these models are highly consistent with human visual perception. The most classical TMQI model is proposed by Wang et al.[9]. It uses MS-SSIM[10] which removes the luminance component to obtain the structural similarity between the original image and the reference image, and adds the natural statistical features of TM image on this basis. The TMQI model achieves good results but doesn't consider the color distortion of TM image. Hossein et al.[11] propose the FSITM model to address this deficiency. The phase similarity between HDR image and TM image is obtained mainly from the local phase information of the images. However, this method separates the RGB channels of the images and can't measure the color distortion of TM image accurately.

In view of the characteristics of TM image such as the detail loss in high-light and low-dark areas, this work proposes an objective quality evaluation model for TM image based on dense scale invariant feature transformation. Firstly, the DSIFT descriptors of HDR image and TM image are extracted, and the local similarity is calculated to represent the phenomenon of detail loss in TM image. Then, the local quality map is refined by using the Gauss exposure curve because HVS is more sensitive to the detail loss in low-dark and high-light areas of TM image. Finally, color distortion is measured by converting the color space of TM image, and the feature aggregation is performed by using support vector machine (SVM). Compared with the existing state-of-the-art quality assessment methods, the proposed method is in good agreement with human visual perception.

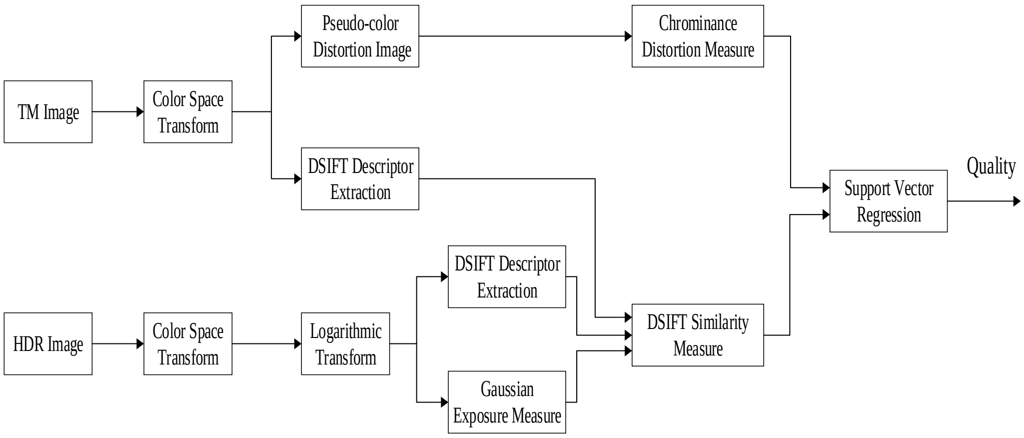

FIGURE I. DIAGRAM OF THE PROPOSED METHOD

II. DIAGRAM OF THE PROPOSED METHOD QUALITY EVALUATION FOR TONE MAPPING IMAGES

Considering that the TM image is prone to the detail loss and color distortion, the HDR image and TM image are first transformed from RGB color space to YCbCr color space and the corresponding luminance components are extracted. Since the luminance range of the HDR image is higher than that of the TM image, it is necessary to convert the luminance value of the HDR image to an appropriate perceptual value by using a nonlinear logarithmic transformation. Then, the DSIFT descriptors of the HDR image and the TM image are extracted respectively and the local similarity is calculated to represent the phenomenon of detail loss in TM image. Since the HVS is more sensitive to the detail loss in the low-dark and high-light areas of the TM image, the Gaussian exposure curve is utilized to refine the local quality map. Finally, color distortion is measured by calculating the chrominance similarity between the TM image and the pseudo-distorted image whose visual effect is similar to the grayscale image, and the feature aggregation is performed by using SVM. Figure I shows the diagram of the proposed method.

A. Global Color Distortion Metric

Because HVS is much more sensitive to luminance than to color, most of the existing TM image quality evaluation models only consider the luminance information and ignore the important chrominance information. The images obtained by different TM operators usually have different degrees of color distortion. Fig. 2 gives an example of the different TM images. It can be seen that Fig. 2(a) has bright colors, while Fig. 2(b) and (c) appear very dim, which results in poor visual perception. Therefore, the quality evaluation based on the colorful TM image will be more reliable.

In order to obtain the chrominance information of TM image, the RGB color space is transformed into YCbCr color space firstly which is more in line with human visual characteristics. The definition is as follows:

$$\begin{bmatrix} Y_{\mathrm{T}} \\ Cb_{\mathrm{T}} \\ Cr_{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R_{\mathrm{T}} \\ G_{\mathrm{T}} \\ B_{\mathrm{T}} \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \tag{1}$$

Where Y_{T} is the luminance information of TM image, Cb_{T} and Cr_{T} are the chrominance information. The inverse process of Eq. (1) is expressed as follows:

$$\begin{aligned} R_{\mathrm{T}} &= Y_{\mathrm{T}} + 1.402(Cr_{\mathrm{T}} - 128) \\ G_{\mathrm{T}} &= Y_{\mathrm{T}} - 0.344(Cb_{\mathrm{T}} - 128) - 0.714(Cr_{\mathrm{T}} - 128) \\ B_{\mathrm{T}} &= Y_{\mathrm{T}} + 1.772(Cb_{\mathrm{T}} - 128) \end{aligned} \tag{2}$$

From Eq. (2), we can learn that the visual effect of the image is closer to the gray image when the Cb_{T} and Cr_{T} component is closer to 128. According to this characteristic, the L1 norm between the chrominance component and 128 is used to approximate the global color distortion of TM image and it can be formulated as:

$$\begin{aligned} C_{\mathrm{Cb}}^{\mathrm{g}} &= \frac{1}{N}(\| Cb_{\mathrm{T}} - 128 \|_1) \\ C_{\mathrm{Cr}}^{\mathrm{g}} &= \frac{1}{N}(\| Cr_{\mathrm{T}} - 128 \|_1) \end{aligned} \tag{3}$$

Where N is the total pixel numbers of chrominance component, $\|\cdot\|_1$ is the L1 norm operator, $C_{\mathrm{Cb}}^{\mathrm{g}}$ and $C_{\mathrm{Cr}}^{\mathrm{g}}$ are global color distortion metric. When the value of Cb_{T} and Cr_{T} are further away from 128, the value of $C_{\mathrm{Cb}}^{\mathrm{g}}$ and $C_{\mathrm{Cr}}^{\mathrm{g}}$ are greater, the color of TM image is richer.

B. Local Detail Loss Metric

HDR images give HVS a better visual experience because of their wide luminance range especially in high luminance and low luminance areas. However, due to the performance differences of TM operators, TM images usually lose a lot of detail information in high and low luminance areas, as shown in Figure II. The indoor scene in Figure II(a) retains more details, but the details of outdoor scene are almost invisible causing the poor visual perception. while the outdoor scene in Figure II(b) and (c) retains more detail information. Therefore, TM image quality evaluation needs to pay attention to the detail loss in high-light and low-dark areas specially.

Recently, Liu et al.[12] have used DSIFT descriptors for image registration. Unlike traditional SIFT descriptors, DSIFT needn't to detect interest points and is directly used to represent local gradient information in eight directions of each pixel in an image. It is more accurate than using gradient magnitude directly, so the DSIFT descriptors are used to detect the detail loss areas of TM image in this work.

(a) (b) (c)

FIGURE II. THREE TM IMAGES BY DIFFERENT TM ALGORITHMS: (A) TM IMAGE 1. (B) TM IMAGE 2. (C) TM IMAGE 3

Similar to the TM image, the RGB color space is converted to YCbCr color space to obtain the luminance component Y_{H} for HDR image. However, it is infeasible to compare their absolute luminance components directly because the luminance dynamic range of HDR image is much larger than that of TM image. Therefore, the logarithmic transformation is used to convert the luminance value of HDR image into the appropriate relative luminance value $Y_{\mathrm{H}}^{\mathrm{r}}$ firstly. Then extract the DSIFT features from Y_{T} and $Y_{\mathrm{H}}^{\mathrm{r}}$, respectively.

$$\begin{aligned} D_{1:M}^{\mathrm{T}} &= \mathrm{DSIFT}(Y_{\mathrm{T}}) \\ D_{1:M}^{\mathrm{H}} &= \mathrm{DSIFT}(Y_{\mathrm{H}}) \end{aligned} \tag{4}$$

Where $D_{1:M}^{\mathrm{T}}$ and $D_{1:M}^{\mathrm{H}}$ are DSIFT feature descriptors of TM image and HDR image respectively, $\mathrm{DSIFT}(\cdot)$ is the operator for calculating DSIFT features, M is the feature dimension and is set to 32 in this work.

Then calculate the DSIFT feature similarity between $D_{1:M}^{\mathrm{T}}$ and $D_{1:M}^{\mathrm{H}}$ for measure local detail loss of TM images.

$$D_{1:M}^{\mathrm{SIM}} = \frac{2D_{1:M}^{\mathrm{T}} D_{1:M}^{\mathrm{H}} + C}{{D_{1:M}^{\mathrm{T}}}^2 + {D_{1:M}^{\mathrm{H}}}^2 + C} \tag{5}$$

Where $D_{1:M}^{\mathrm{SIM}}$ is the DSIFT feature similarity map, C is a small positive constant and is set to 0.0001.

Because HVS is more sensitive to the detail loss in high-light and low-dark areas of HDR images, it is necessary to consider the distortion distribution weight in different areas of HDR images. In this work, we simply use the Gauss exposure curve to simulate the distortion distribution weight, which is defined as follows:

$$E_{\mathrm{H}} = \exp\left(\frac{(Y_{\mathrm{H}}^{r} - m)^2}{2\sigma_{\mathrm{H}}^2}\right), \quad m = \mathrm{Median}(Y_{\mathrm{H}}^{r}) \tag{6}$$

Where E_{H} is the distortion sensitivity weight, $\mathrm{Median}(\cdot)$ is the median operator, σ_{H} controlls the flatness of Gauss curve and is set to 0.2. When the pixel value of HDR image is closer to m, i.e. medium exposure area, the weight of distortion allocation is smaller.

Finally, the original DSIFT feature similarity maps are weighted by distortion sensitivity function to obtain the refined local detail loss quality maps $D_{1:M}^{\mathrm{f}}$, which is given as:

$$D_{1:M}^{\mathrm{f}} = D_{1:M}^{\mathrm{SIM}} E_{\mathrm{H}} \tag{7}$$

Figure III(a-c) are the refined local detail loss quality maps corresponding to the TM images in Figure II(a-c) respectively. The color depth of the pixel in the quality map indicates the quality of the TM image. The darker the color, the worse the quality of the region. For example, the outdoor scene in Figure II(a) is invisible and the corresponding quality value in Figure III(a) is very dark. So the proposed DSIFT feature similarity can reflect the local detail loss of TM images well.

(a) (b) (c)

FIGURE III. THE LOCAL DETAIL LOSS QUALITY MAPS CORRESPONDING TO THE TM IMAGES IN FIGURE II

C. Feature Aggregation

SVM can map samples from low-dimensional space to high-dimensional space through non-linear mapping and use linear algorithm to analyze the non-linearity of samples. In addition, it can find the optimal classification hyperplane in the feature space based on the principle of structural risk minimization and has a strong ability to deal with non-linearity of samples, so it is used for feature aggregation in this work.

In this paper, a total of 34-dimensional features are extracted including global color distortion metric $C_{\mathrm{Cb}}^{\mathrm{g}}$ and $C_{\mathrm{Cr}}^{\mathrm{g}}$, local detail loss metric $D_{1:M}^{\mathrm{f}}$. Then the 34-dimensional features and the subjective scores are input into SVM for training and testing in order to obtain the TM image quality evaluation model.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Database and Evaluation Criteria

In order to verify the validity of the proposed model, the consistency experiment between objective quality evaluation and subjective perception for TM image was carried out on TMID database[13]. The database consists of 15 HDR images and 120 TM images generated by 8 different TM operators. Each TM image corresponds to a mean opinion score (MOS). The value of MOS ranges from 1 to 8, of which 1 represents the best quality and 8 represents the worst quality.

According to the reference standard of image quality evaluation given by VQEG, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SROCC) and Kendall Rank-order Correlation

Coefficient (KROCC) are selected to evaluate the performance of the method. PLCC reflects the correlation between subjective evaluation and objective evaluation, and the closer the absolute value is to 1, the higher the prediction accuracy is. SROCC reflects the consistency between subjective evaluation and objective evaluation, and the closer the absolute value is to 1, the better the monotony is; KROCC reflects the correlation degree of classification variables, and the closer the absolute value is to 1, the better the quality prediction model is.

In addition, when using SVM for feature aggregation, the input is a 34-dimensional feature set generated by the objective model and the ideal output is the MOS value. In the training process, 80% TM images and the corresponding HDR images in the database are randomly selected as the training set, and the remaining 20% images are used as the test set. At the same time, in order to verify the accuracy of the proposed model, the training and testing process is repeated 1000 times, and the median value is taken as the final performance index.

B. Performance Comparison

In order to evaluate the performance of the proposed TM image quality evaluation model, we compare it with the five full-reference quality evaluation models for the ordinary images, i.e. RFSIM[4], FSIM[5], GMSD[6], MDSI[7], RVSIM[8] and the two full-reference quality evaluation models for the TM images, i.e. TMQI[9], FSITM[11]. The results are shown in Table I.

TABLE I. PERFORMANCE COMPARISON OF PLCC, SROCC AND KROCC

Model	**PLCC**	**SROCC**	**KROCC**
RFSIM[4]	0.4456	0.3957	0.2727
FSIM[5]	0.4885	0.3578	0.2420
GMSD[6]	0.5167	0.4167	0.2808
MDSI[7]	0.5643	0.4325	0.3143
RVSIM[8]	0.6187	0.4839	0.3452
TMQI[9]	0.7715	0.7407	0.5585
FSITM[11]	0.7496	0.7028	0.5160
Proposed	0.8045	0.7754	0.5912

It can be seen that the five full-reference quality evaluation models for ordinary LDR images can’t predict the quality of TM images accurately, because the luminance dynamic range of HDR images and TM images is inconsistent and the similarity calculation can’t be carried out directly, so the traditional LDR image quality evaluation methods are not suitable for the quality evaluation of TM images. Considering the difference of luminance dynamic range between reference image and test image, TMQI model directly uses MS-SSIM without luminance component to obtain structural similarity between HDR image and TM image, and the natural statistical feature of TM image is added on this basis. Compared with five objective quality evaluation models for LDR images, the performance of TMQI model has been greatly improved, but the effect is not ideal because the color distortion of TM images is not considered. FSITM model is mainly based on FSIM model. Considering the inconsistency of dynamic range between reference image and test image, the luminance value of HDR image is first converted, and then the phase consistency similarity between HDR image and TM image is calculated. The results show that the performance of FSITM model is greatly improved compared with FSIM model, but its effect is not good compared with TMQI model.

IV. CONCLUSION

TM images are more consistent with the perception characteristics of HVS than ordinary LDR images, especially in the high-light and low-dark areas of TM image. Aiming at the problems of detail loss and color distortion in TM image, an objective quality evaluation model for TM image based on dense scale invariant feature transformation is proposed. Considering the inconsistency of luminance dynamic range between HDR image and TM image, the sensitivity of human eyes to distortion in different luminance regions and the importance of chrominance, the proposed model is in good agreement with human eye perception. However, there is no so-called reference image in most practical applications and the full-reference quality assessment model is subject to many constraints. Therefore, in the future work, we will establish a no-reference quality assessment model for TM image. In addition, it can also consider automatic feature extraction through training a deep network to avoid the defects of manual feature extraction in traditional methods.

ACKNOWLEDGMENT

This project was supported by Education Department of Zhejiang province(Y201840219).

REFERENCES

[1] A. Artusi, T. Richter, T. Ebrahimi and R. K. Mantiuk, "High Dynamic Range Imaging Technology," IEEE Signal Processing Magazine, vol.34, pp.165-172, 2017.

[2] K. Gu, S. Wang, G. Zhai, et al, "Blind Quality Assessment of Tone-Mapped Images via Analysis of Information, Naturalness and Structure," IEEE Transactions on Multimedia, vol.18, pp.432-443, 2016.

[3] G. Yue, C. Hou, K. Gu, S. Mao and W. Zhang, "Biologically Inspired Blind Quality Assessment of Tone-Mapped Images," IEEE Transactions on Industrial Electronics, vol.65, pp.2525-2535, 2018.

[4] L. Zhang, L. Zhang and X. Mou, "RFSIM: A Feature Based Image Quality Assessment Metric Using Riesz Transforms," IEEE International Conference on Image Processing, 2010.

[5] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," IEEE Transactions on Image Processing, vol.20, pp.2378-2386, 2011.

[6] W. Xue, L. Zhang, X. Mou and A. C. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," IEEE Transactions on Image Processing, vol.23, pp.684-695, 2014.

[7] H. Z. Nafchi, A. Shahkolaei, R. Hedjam and M. Cheriet, "Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator," IEEE Access, 2017, pp.5579-5590.

[8] G. Yang, D. Li, F. Lu, Y. Liao and W. Yang, "RVSIM: a feature similarity method for full-reference image quality assessment," EURASIP Journal on Image and Video Processing, 2018, pp.1-15.

[9] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," IEEE Transactions on Image Processing, vol.22, pp.657-667, 2013.

[10] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," IEEE Asilomar Conference on Signals, 2003.

[11] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam and M. Cheriet, "FSITM: A Feature Similarity Index For Tone-Mapped Images," IEEE Signal Processing Letters, vol.22, pp.1026-1029, 2015.

[12] C. Liu, J. Yuen and A. Torralba, "SIFT flow: dense correspondence across scenes and its applications," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.33, pp.978-994, 2011.

[13] H. Yeganeh, "Objective quality assessment of tone-mapped images," https://ece.uwaterloo.ca/~z70wang/research/tmqi/TMQI_SubejctiveTest.rar, 2013.