NO-REFERENCE HDR IMAGE QUALITY ASSESSMENT METHOD BASED ON TENSOR SPACE

Feifan Guan[1], *Gangyi Jiang*[1,2], *Yang Song*[1], *Mei Yu*[1,2*], *Zongju Peng*[1], *Fen Chen*[1]

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
[2]National Key Lab of Software New Technology, Nanjing University, Nanjing, China

ABSTRACT

The full-reference image quality assessment (IQA) method are limited in practical applications. Here we propose a no-reference quality assessment method for high dynamic range (HDR) images based on tensor space. First, the tensor decomposition is used to generate three feature maps of an HDR image, considering color and structure information of the HDR image. Second, for a given HDR image, the corresponding multi-scale manifold structure features are extracted from the first feature map. For the second and third feature maps of the HDR image, multi-scale contrast features are extracted. Finally, the extracted features are aggregated by support vector regression to obtain the objective quality score of the HDR image. Experimental results show that the proposed method is superior to some representative full and no-reference methods, and even superior to the full-reference HDR IQA method, HDR-VDP-2.2, on the Nantes database. The proposed method has a higher consistency with human visual perception.

***Index Terms*—**No-reference, high dynamic range, image quality assessment, tensor space, feature maps

1. INTRODUCTION

With the rapid development of optical imaging and data-processing technologies, there has been growing interest in high-dynamic-range (HDR) images in recent years. Unlike traditional low-dynamic-range (LDR) scenes, the luminance levels in HDR scenes can range from 10^{10}:1 [1]. HDR achieves a more complete representation of the luminance variations in real scenes that the human eye can see, ranging from direct sunlight to faint starlight in real scenes. Hence, there is a better contrast distribution in HDR images than in LDR images, which leads to a higher degree of detail preservation.

Similar to LDR images, HDR images can be distorted when they are acquired, processed, compressed, and transmitted; these distortions may affect the visual effects of HDR images. Therefore, quality assessment of HDR image/video systems, in terms of quality of experience, is an essential issue. Although subjective evaluation can better reflect human visual experience, it is time consuming and difficult to embed into an actual system. Thus, objective quality assessment tools are needed. Objective quality assessment methods can be classified into three categories: full-reference (FR) methods, which compare the test image with a reference image; reduced-reference (RR) methods, which use part of the information from the reference image; and no-reference (NR) methods, which do not use any information about the reference image. In this study, only FR and NR methods are considered. Over the past several decades, research on LDR image quality assessment (IQA) has made remarkable progress. However, LDR IQA methods are designed for gamma encoded images, typically with luminance values in the range 0.1–100 cd/m2, while HDR images have linear values and are meant to capture a much wider range of luminance. Obviously, LDR IQA methods cannot be directly applied in HDR IQA tasks well [2]. And currently, there is still a lack of an effective NR method of evaluating HDR images, but, in practical applications, NR IQA is the only realistic option for HDR systems. Thus, NR HDR IQA is an urgent problem to solve.

In this paper, a new NR HDR IQA method is proposed based on tensor space, in which the HDR image with color information is represented as a third-order tensor, and the HDR image quality is blindly assessed in tensor space by extracting the multi-scale manifold structure features and visual contrast.

The rest of the paper is organized as follows. In Section 2, we introduce the tensor space of a HDR image, and in Section 3 we describe the proposed NR HDR IQA method. The experimental results are discussed in Section 4, and the conclusions are given in Section 5.

2. TENSOR SPACE

For high-dimensional data, the classical data-processing approach is to transform it into a vector, which will lead to the dimensions of the data samples to be too high and destroy the structure of the data. To solve this problem, a

This work is supported by the National Natural Science Foundation of China (NSFC) (61671258, 61620106012, 61771269, U1301257) and National High-tech R&D Program of China (2015AA015901).

tensor can be used. In this section, we discuss how to use the tensor decomposition to analyze the feature representation of HDR images.

Assuming that a HDR image contains three channels and the size of each channel image is $M \times N$, the HDR image can then be represented by a third-order tensor $\mathcal{I}$ of the size $M \times N \times 3$. The Tucker3 decomposition [3] of the HDR image $\mathcal{I}$ is defined as follows.

$$\mathcal{I} = \boldsymbol{\zeta} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} = \boldsymbol{\xi} \times_3 \mathbf{U}^{(3)}, \qquad (1)$$

where ξ is the core tensor with the same order and dimension as $\mathcal{I}$, $\boldsymbol{\xi} = \boldsymbol{\zeta} \times \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ are orthogonal matrices of sizes $M \times M$, $N \times N$, and 3×3, respectively.

The i-th channel matrix of size $M \times N$ of the core tensor is denoted by $\boldsymbol{\theta}_i$ ($1 \le i \le 3$), and by the formula (1), the t-channel of the HDR image can be expressed as a linear combination of $\boldsymbol{\theta}_i$.

$$\mathbf{I}_t = \mathbf{U}_{t,1}^{(3)} \boldsymbol{\theta}_1 + \mathbf{U}_{t,2}^{(3)} \boldsymbol{\theta}_2 + \mathbf{U}_{t,3}^{(3)} \boldsymbol{\theta}_3 = \sum_{i-1}^{3} \mathbf{U}_{t,i}^{(3)} \boldsymbol{\theta}_i, \qquad (2)$$

From the above equation, it is found that each channel image of the HDR image is a linear combination of $\{\boldsymbol{\theta}_i \,|i=1,2,3\}$, and each row of the orthogonal matrix $\mathbf{U}^{(3)}$ represents the correlation coefficient between each channel image. Here, $\{\boldsymbol{\theta}_i \,|i=1,2,3\}$ represents a set of feature maps. In addition, the energy of the HDR image $\mathcal{I}$ is the sum of the energy of each feature map $\{\boldsymbol{\theta}_i \,|i=1,2,3\}$.

$$\|\mathcal{I}\|^2 = \|\boldsymbol{\xi}\|^2 = \sum_{i=1}^{3} \|\boldsymbol{\theta}_i\|^2, \qquad (3)$$

where $\|\cdot\|$ represents the Frobenius norm.

According to (3) and its corresponding SVD, we have

$$\|\boldsymbol{\theta}_1\|^2 \ge \|\boldsymbol{\theta}_2\|^2 \ge \|\boldsymbol{\theta}_3\|^2, \qquad (4)$$

It can be seen from (3) and (4) that the energy of a HDR image is distributed between its feature maps in descending order from the first to the last feature map. Here, we call $\boldsymbol{\theta}_1$ the first feature map, $\boldsymbol{\theta}_2$ the second feature map and $\boldsymbol{\theta}_3$ the third feature map. Since $\{\boldsymbol{\theta}_i \,|i=1,2,3\}$ are matrices of three channels of the sub-tensor, after we obtain the matrix $\mathbf{U}^{(3)}$, the sub-tensor can be obtained by tensor multiplication.

$$\boldsymbol{\xi} = \mathcal{I} \times_3 \left(\mathbf{U}^{(3)}\right)^T. \qquad (5)$$

Here, ξ is the sub-tensor we must discuss. Specifically, the first feature map contains the main energy of three channels, the feature maps $\{\boldsymbol{\theta}_i \,|i=1,2,3\}$ reflect the change of the HDR image along three channels, and these three feature maps are collectively called tensor space.

3. PROPOSED NR HDR IMAGE QUALITY ASSESSMENT METHOD

Based on the tensor decomposition, three feature maps are generated, and two complementary types of features are separately extracted from the feature maps in tensor space to form a feature set that is predictive of the perceived quality. More concretely, the color HDR image is represented by the third-order tensor, and the three feature maps are obtained by tensor decomposition. The first type is a multi-scale manifold structure feature derived by manifold learning from the first feature map. The second type belongs to a multi-scale perceived detail contrast feature derived from the second and third feature maps. Finally, a SVR technique is adopted to fuse all the above features into their corresponding subjective scores. We elaborate on each component in detail in the following.

3.1. Extraction of multi-scale manifold structure feature in a HDR image

3.1.1. The establishment of the best projection matrix

- Constructing training sample set: For each selected training set, the first feature maps are obtained from the training images by the tensor decomposition. From the first feature maps, 20000 image blocks of size 8×8 are selected randomly, and each image block is transformed into a column vector, thereby obtaining a column vector with a length of 64. All column vectors generated from image blocks form a training sample matrix **X**.
- PCA processing: Studies have shown that the retina and lateral geniculate nucleus (LGN) will whiten the input visual signal [4]. Therefore, in order to simulate the function of the retina and LGN, we use the PCA algorithm to reduce the dimension of the training sample matrix and whiten it. Here, the PCA process is realized by eigenvalue decomposition of the covariance matrix; the training sample matrix after whitening is denoted $\mathbf{X}_w$.
- Best projection matrix: After whitening of the training sample matrix, the best projection matrix **W** is calculated by the orthogonal locality-preserving projection (OLPP) algorithm of manifold learning. The specific process of the algorithm is shown in [5]. Here, **W** can be used to extract the manifold feature of image blocks.

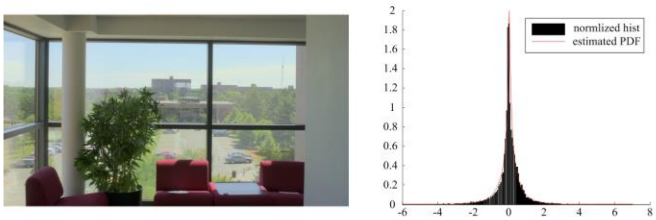
3.1.2 Extraction of manifold structure feature matrix

To estimate the quality of a HDR image, the first feature map of the test image is divided into a plurality of 8×8 non-overlapping image blocks, while each image block subtracts

$$\mathbf{d}_k = \mathbf{W} \times \mathbf{y}_k, \qquad (6)$$

Finally, the manifold structure feature matrix **D** of the test HDR image **Y** is generated as follows.

$$\mathbf{D} = \left[\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_k\right]. \qquad (7)$$



(a) An HDR image (b) Histogram distribution

Fig.1. Statistical histogram and GGD fitting curve of the manifold feature matrix.

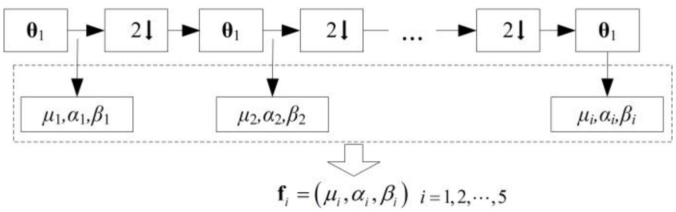


Fig.2. Block diagram of multi-scale manifold structure feature extraction.

3.1.3 Extraction of multi-scale manifold structure features

Here, the histogram of the manifold feature matrix **D** is fitted through the generalized Gaussian distribution (GGD) model to describe the manifold structure feature of a HDR image. Fig. 1 gives a concrete example, i.e., a test HDR image with its corresponding histogram distribution and GGD fitting curve. The density function of the GGD is defined as follows.

$$g(x;\mu,\alpha,\beta)=\frac{\alpha}{2\beta\Gamma(1/\alpha)}\exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right), \tag{8}$$

where $\Gamma(\cdot)$ is a gamma function computed by

$$\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}dt, x>0, \tag{9}$$

where μ is the mean of the intensity of the image blocks, α the shape parameter, and β the scale parameter. As effective quality-aware features, μ, α, and β are used to characterize the manifold feature of the HDR image.

At the same time, due to the fact that a multi-scale operation can better reflect the details of a HDR image, here the first feature map $\boldsymbol{\theta}_1$ is down-sampled four times by 1/2, and five resolution feature maps are obtained. For each feature map of different scales, the same processing is used to obtain a total of five sets of GGD fitting parameters $\mathbf{f}_i=(\mu_i,\alpha_i,\beta_i)$ (i=1,2…,5) as shown in Fig. 2. $\mathbf{f}_i$ represents the manifold structure feature of an ith scale HDR image. Thus, the final multi-scale manifold structure features set of the HDR image is produced as follows.

$$\mathbf{f}_{MS}=[\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3,\mathbf{f}_4,\mathbf{f}_5], \tag{10}$$

where the dimension of $\mathbf{f}_{MS}$ is15(3×5=15).

3.2. Extraction of perceived detail contrast feature in HDR image

Rich detail information and good contrast information can give people a better visual experience. According to the properties of the tensor decomposition, the second and third feature maps contain less structural information than the first feature map, which mainly shows the details and the contrast information of the original HDR image.

However, the standard deviation can reflect the dispersion degree of data, which can be used to measure contrast distortions of an image. Here, as complementary feature, we use the standard deviation as an estimate of the perceived detail contrast. Specifically, the second and third feature maps are first divided into multiple image blocks using an 8×8 non-overlapping sliding window, and then the standard deviation of each image block is calculated using the following equations. Finally, the mean of the standard deviation of all image blocks is taken as the final standard deviation. The standard deviation computational formula is

$$\sigma=\frac{1}{P}\sum_{P}\left(\frac{1}{p-1}\sum_{j=1}^{p}\left(x_j-\bar{x}\right)^2\right)^{1/2}, \tag{11}$$

$$\bar{x}=\frac{1}{P}\sum_{j=1}^{P}x_j, \tag{12}$$

where p represents the number of pixels in each image block, P is the total number of blocks, and $\bar{x}$ is the mean intensity of the pixels.

Similarly, for the five scales corresponding to the feature maps $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, five sets of standard deviations $\mathbf{f}_{i+5}=(\sigma_{\boldsymbol{\theta}2}(i),\sigma_{\boldsymbol{\theta}3}(i))$ (i=1,2…,5) are obtained. Thus, the multi-scale perceived detail contrast feature set is produced as follows.

$$\mathbf{f}_{PC}=[\mathbf{f}_6,\mathbf{f}_7,\mathbf{f}_8,\mathbf{f}_9,\mathbf{f}_{10}]. \tag{13}$$

3.3. Quality assessment

Combining the manifold feature set $\mathbf{f}_{MS}$ in the first feature map and the perceived detail contrast feature set $\mathbf{f}_{PC}$ in the second and third feature maps, the perceived quality feature set **f** of a HDR image in our method is obtained as follows.

$$\mathbf{f}=[\mathbf{f}_{MS},\mathbf{f}_{PC}]. \tag{14}$$

where the dimension of the feature set is 25 (3×5+2×5=25). In this paper, the features extracted in the tensor space are used as the HDR-image-quality-aware features, and the SVR technique is used to establish the quality assessment model to blindly evaluate the HDR image quality.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Databases and training sets

The validity of the algorithm was tested and compared in these two public HDR image databases. One of the databases is provided by Nantes University [6] and the other is based on the JPEG XT standard and is provided by the EPFL Laboratory [7]. Each database contains reference images and a series of distorted images. Each image gives the corresponding MOS values.

Since the best projection matrix is obtained through training, the training images must be selected. In order to eliminate the influence of the training process on the accuracy of the features extracted, the storage type of the manifold training images should be same as that of the test images because the data with different storage methods of HDR images differ greatly. In order to avoid overlapping of training images and test images, and to prove the effectiveness of the method, 10 non-distorted HDR images selected from the DML-HDR image set [8, 9] are used as training image set S1 for quality assessment of HDR images with ".hdr" format in Nantes database; for the EPFL database with images of ".pfm" format, we randomly selected 10 non-distorted HDR images from the EPFL database to form the training image set S2, and the remaining images in EPFL database were used for testing.

4.2. Performance indexes

According to the relevant process of the video quality experts group (VQEG) [10], the output value Q of the model is nonlinearly fitted with five parameter logistical functions to obtain the prediction MOS_p, which is computed by

$$MOS_p=\lambda_1\left(\frac{1}{2}-\frac{1}{1+\exp\left(\lambda_2\left(Q-\lambda_3\right)\right)}\right)+\lambda_4 Q+\lambda_5. \tag{15}$$

where λ_1, λ_2, λ_3, λ_4, and λ_5 are the model adjustment parameters.

The experimental results are given by three commonly used performance evaluation indexes: the Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), and the root-mean-square error (RMSE). A better objective method should have higher PLCC and SROCC values and lower RMSE values.

4.3. Performance comparison

Owing to the lack of NR HDR IQA methods, in this subsection we compare the proposed method to four representative IQA methods including two FR LDR methods (MSE and SSIM [11]), a NR LDR method (DIIVINE [12]), and a FR HDR method (HDR-VDP-2.2 [13]) for HDR images. Because it is unreasonable to evaluate HDR images directly by traditional LDR IQA methods, so we first transform HDR images to a perceptually uniform (PU) space [14], then the LDR methods were computed in the PU space. It is worth mentioning that the HDR-VDP-2.2 method is considered the state-of-the-art FR HDR IQA method.

Table 1. Performance comparison of different assessment methods.

Metho -ds	Indexes	PU-MSE	PU-SSIM	PU-DIIVI-NE	HDR-VDP-2.2	Propos-ed
Nantes	PLCC	0.4471	0.6056	0.2613	0.7329	**0.9269**
	SROCC	0.4197	0.6528	0.2271	0.7047	**0.9153**
	RMSE	0.9019	0.8006	0.9712	0.6485	**0.3671**
EFPL	PLCC	0.8241	0.9178	0.5892	**0.9500**	0.9015
	SROCC	0.8385	0.9191	0.5081	**0.9419**	0.8740
	RMSE	0.6798	0.4750	0.9668	**0.3736**	0.5016

Table1 lists the performance indexes for each evaluation method computed on the two databases. Two sets of the best performance indexes in the table are emboldened. As can be seen from table 1, the performances of the proposed method achieved good results. First, in the Nantes image database, our method is optimal, not just far superior to the traditional FR and NR methods, and even better than the FR HDR IQA method (that is, HDR-VDP-2.2). Second, on the EPFL image database, the proposed method is not the best, but like the NR HDR IQA method, its performance is closest to the representative HDR IQA method (HDR-VDP-2.2). Compared to the typical LDR NR method DIIVINE, the advantages of the proposed method are very obvious. Meanwhile, the performance of the proposed method is relatively stable. In contrast, the evaluation results using other methods, even HDR-VDP-2.2, are not stable. Therefore, compared to other methods, the predictive results of the proposed method are closer to that of the subjective evaluation and the performance is stable.

5. CONCLUSIONS

Based on the tensor space of a HDR image and a combination of the HDR image color feature, manifold structure feature, and perceived detail contrast feature, a new NR HDR IQA method was proposed. First, three feature maps were obtained by tensor decomposition. The three feature maps contain information such as the structure, detail, and color of the original HDR image. The three feature maps were then extracted with different features. In contrast to the traditional image feature extracted in the gray domain, we achieved an effective feature extraction in the tensor space of the HDR image. The experimental results show that the proposed method is highly consistent with human visual perception. On the basis of this research, our future work will be focused on developing more efficient feature extraction methods and concentrating on HDR video quality assessment.

6. REFERENCES

[1] D. Kundu, D. Ghadiyaram, A Bovik, et al, "No-Reference Quality Assessment of Tone-Mapped HDR Pictures," *IEEE Transactions on Image Processing,* vol. 26, no. 6, pp. 2957-2971, Mar. 2017.

[2] P. Hanhart, M. V. Bernardo, M. Pereira, "Benchmarking of objective quality metrics for HDR image quality assessment," *Eurasip Journal on Image & Video Processing*, vol. 2015, no. 1, pp. 1-18, Dec. 2015.

[3] J. Zhang, Y. Han, J. Jiang, "Tucker decomposition-based tensor learning for human action recognition," *Multimedia Systems*, vol. 22, no. 3, pp. 343-353, 2016.

[4] E. P. Simoncelli, B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 24, pp. 1193-1216, Mar. 2001.

[5] C. Deng, X. He, J. Han, "Orthogonal Laplacian faces for face recognition," *IEEE Transactions on Image Processing*, vol. 15, no, 11, pp. 3608−3614, Oct. 2006.

[6] M. Narwaria, P. L. Callet, "Tone mapping based high dynamic range image compression: study of optimization criterion and perceptual quality," *Optical Engineering*, vol. 52, no. 10, pp. 102008, Oct. 2013.

[7] P. Korshunov, P. Hanhart, T. Richter, "Subjective quality assessment database of HDR images compressed with JPEG XT," *Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015, pp. 1-6.

[8] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, "Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content," *International Conference on Multimedia Signal Processing (ICMSP)*, 2014.

[9] A. Banitalebi-Dehkordi, M. Azimi, Y. Dong, "Quality assessment of High Dynamic Range (HDR) video content using existing full-reference metrics," *ISO/IEC JTC1/SC29/WG11*, France, 2014.

[10] Video Quality Experts Group, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," 2003, http://www.vqeg.org/.

[11] Z. Wang, A. C. Bovik, H R Sheikh, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[12] A. K. Moorthy, A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350-3364, Apr. 2011.

[13] M. Narwaria, R. Mantiuk, P. L. Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 010501, Jan. 2015.

[14] T. O. Aydın, R. Mantiuk, H. P. Seidel, "Extending quality metrics to full luminance range images," *Human Vision and Electronic Imaging XIII*, vol. 6806, pp. 68060B, Mar. 2008.