No-Reference Image Quality Assessment for High Dynamic Range Images

Debarati Kundu, Deepti Ghadiyaram, Alan C. Bovik and Brian L. Evans
Wireless Networking and Communications Group
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX USA
Email: debarati@utexas.edu, deepti@cs.utexas.edu, bovik@ece.utexas.edu, bevans@ece.utexas.edu

***Abstract*—Being able to automatically predict digital picture quality, as perceived by human observers, has become important in many applications where humans are the ultimate consumers of displayed visual information. Standard dynamic range (SDR) images provide 8 bits/color/pixel. High dynamic range (HDR) images which are usually created from multiple exposures of the same scene, can provide 16 or 32 bits/color/pixel, but must be tonemapped to SDR for display on standard monitors. Multi-exposure fusion (MEF) techniques bypass HDR creation, by fusing the exposure stack directly to SDR format while aiming for aesthetically pleasing luminance and color distributions. Here we describe a new no-reference image quality assessment (NR IQA) model for HDR pictures that is based on standard measurements of the bandpass and on newly-conceived differential natural scene statistics (NSS) of HDR pictures. We derive an algorithm from the model which we call the Gradient Image Quality Assessment algorithm (G-IQA). NSS models have previously been used to devise NR IQA models that effectively predict the subjective quality of SDR images, but they perform significantly worse on tonemapped HDR content. Towards ameliorating this we make here the following contributions: (1) We design a HDR picture NR IQA model and algorithm using both standard space-domain NSS features as well as novel HDR-specific gradient based features that significantly elevate prediction performance, (2) We validate the proposed models on a large-scale crowdsourced HDR image database, and (3) We demonstrate that the proposed model also perform well on legacy natural SDR images. The software is available at: http://signal.ece.utexas.edu/%7Edebarati/higradeRelease.zip.**

I. Introduction

RECENT years have seen a huge growth in the popularity of High Dynamic Range (HDR) images due to their ability to accurately represent the wide range of variation of illumination in real scenes. Unlike traditional Standard Dynamic Range (SDR) scenes, the luminance levels in HDR scenes can range from 10,000 to 1 [1]. HDR rendering also finds its use in computer graphics where the lighting calculations are performed over a wider dynamic range. This results in a better contrast variation threby leading to a higher degree of detail preservation. However, in order to visualize HDR images on standard displays meant for SDR images, they need to tone-mapped to an SDR image. In addition to tone-mapped SDR images, we also come across images created by multi-exposure fusion that takes a stack of SDR images at varying exposure levels and fuses them to create an output SDR image that is more informative than the input images. This bypasses the intermediate step of creating an HDR irradiance map. In addition, many of the HDR images may be post-processed (color saturation, color temperature, detail enhancement, etc.) based on the scene content and for aesthetic purposes.

For all these applications, different algorithms result in different SDR images, so a natural question is how to evaluate the quality of the images obtained. Subjective testing is important in evaluating the visual quality of images produced by different algorithms. A faster and less expensive alternative is objective quality evaluation. Recently, full-reference image quality assessment (FR-IQA) algorithms were proposed for evaluating tone-mapped SDR images [1] [2] [3] and SDR images created by multi-exposure fusion [4] based on the principles of structural fidelity. However, in evaluating the quality of these images created by HDR processing algorithms, it is hard to come up with a 'reference' image because the input to the algorithms is an exposure stack that may have a varying number of images based on the camera settings used.

In this paper, we propose a scene-statistics based no-reference image quality assessment (NR-IQA) algorithm for evaluating the quality of SDR images obtained by tone-mapping, multi-exposure fusion, and post-processing methods. Many NR metrics rely on machine learning approaches using features expressive of statistical regularities possessed by pristine images, commonly called natural scene statistics (NSS) models [5] [6]. NSS models for good quality natural images hold relatively well irrespective of image content and it is assumed that distortions tend to deviate from these statistical regularities. These techniques have not been used before for NR evaluation of HDR image quality.

In developing the proposed algorithm, we first conduct a large-scale crowdsourced study on our ESPL-LIVE HDR Database [7]. The database contains tone-mapped, multi-exposure fused and post-processed images. Second, we evaluate spatial and gradient domain features and combine those with the highest correlation with the subjective evaluation scores in the database. The proposed NR-IQA algorithm outperforms the state-of-the-art NSS based NR-IQA algorithms that have done remarkably well on non-HDR processed images. For additional validation, we evaluate the correlation performance of the proposed algorithm and other NSS-based NR-IQA algorithms on legacy natural image databases.

D. Kundu and B. L. Evans are with the Embedded Signal Processing Laboratory (ESPL), and D. Ghadiyaram and A. C. Bovik are with the Laboratory for Image and Video Engineering (LIVE) at UT Austin.

II. Proposed algorithm

This section outlines the different perceptually relevant features employed in this NR-IQA algorithm. This is based on the assumption that the pointwise and pairwise log-derivative statistics of the pixels and the pixel gradient magnitudes change with the different types of HDR processing methods and this deviation may be used to predict the quality scores.

A. Computing Log-Derivatives

DErivative Statistics-based QUality Evaluator (DESIQUE) [8] has successfully used log-derivative based features in predicting the quality of natural images with non-HDR processing artifacts. The log-derivative statistics of the images are based on the difference between a particular pixel and its neighbors after converting the pixels to the logarithm domain [8]. Let I(i, j) be the pixel value in the (i, j)-th spatial location of the MxN image I, $i \in \{1, 2, .., M\}$, $j \in \{1, 2, ..., N\}$. The logarithm image is given by

$$J(i,j) = \log[I(i,j) + C] \tag{1}$$

where C is a small positive constant added to avoid numerical instabilities. Considering the different neighboring directions, the following log-derivatives are defined:

$$D1 : \nabla_x J(i,j) = J(i,j+1) - J(i,j), \tag{2}$$

$$D2 : \nabla_y J(i,j) = J(i+1,j) - J(i,j), \tag{3}$$

$$D3 : \nabla_{xy} J(i,j) = J(i+1,j+1) - J(i,j), \tag{4}$$

$$D4 : \nabla_{yx} J(i,j) = J(i+1,j-1) - J(i,j), \tag{5}$$

$$D5 : \nabla_x \nabla_y J(i,j) = J(i-1,j) + J(i+1,j) - J(i,j-1) - J(i,j+1), \tag{6}$$

$$D6 : \nabla_{cx} \nabla_{cy} J(i,j)_1 = J(i,j) + J(i+1,j+1) - J(i,j+1) - J(i+1,j), \tag{7}$$

$$D7 : \nabla_{cx} \nabla_{cy} J(i,j)_2 = J(i-1,j-1) + J(i+1,j+1) - J(i-1,j+1) - J(i+1,j-1). \tag{8}$$

B. Spatial domain scene statistics

For this work, we model the scene statistics of the images in the spatial domain, specifically the MSCN pixels and the σ-field of the image. The pixels of the image are preprocessed by mean subtraction and divisive normalization. Let $M \times N$ be the dimension of the image I, and $I(i,j)$ be the pixel value in the (i,j)-th spatial location, $i \in \{1, 2, .., M\}$, $j \in \{1, 2, ..., N\}$. MSCN pixels are generated by

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1}, \tag{9}$$

where the local mean $\mu(i,j)$ and standard deviation $\sigma(i,j)$ are defined as

$$\mu(i,j) = \sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l} I(i+k, j+l) \tag{10}$$

and

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} w_{k,l} [I(i+k, j+l) - \mu(i,j)]^2} \tag{11}$$

where $w = \{w_{k,l} | k = -K, .., K, l = -L, ..., L\}$ is a symmetric local convolution window centered at the (i,j)-th pixel. K and L determine the size of local patch considered in the calculation of the mean and standard deviation. In [12], the authors considered 7×7 image patches, and a circularly symmetric 2D Gaussian kernel; however, experiments show that the distributions of the MSCN patches are not very sensitive to the window size, or the convolution kernel.

The variance normalized image $(\hat{I})$ tends to be more uniform than the original image, and almost looks like a noise pattern, except at object boundaries. Also, their histograms seem to show a Gaussian like distribution. The standard deviation image σ looks more like the original image, highlighting object boundaries and attenuating textures. The MSCN pixels have been modeled using a Generalized Gaussian Distribution (GGD) and used in image quality assessment [12] [13].

Log-derivatives of the adjacent MSCN coefficients are also modeled by a GGD. The shape(α) and scale(γ) parameters of the GGD fitted to the seven types of log-derivatives have also been used as features in the spatial domain.

We also extract two quantities from the σ-field: mean(Φ_σ) and square inverse of coefficient of variation(Ψ_σ):

$$\Phi_\sigma = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sigma(i,j), \tag{12}$$

$$\Sigma_\sigma(i,j) = \sqrt{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [\sigma(i+k, j+l) - \Phi_\sigma(i,j)]^2}, \tag{13}$$

$$\Psi_\sigma = \left(\frac{\Phi_\sigma}{\Sigma_\sigma} \right)^2. \tag{14}$$

C. Gradient Domain Scene Statistics

The gradient field of the image gives important information about the distribution of edges and variations in local contrast. It has been used in state-of-the-art FR-IQA algorithms [14]. Tone-mapping and multi-exposure fusion algorithms modify local gradients of the multi-exposure stacks, which results in changing contrast of the resultant fused image both locally and globally.

For this algorithm, local gradient is computed by convolving the image with a Sobel filter along the horizontal and vertical directions. The features, as summarized in Table I are also extracted from the gradient magnitude field. The resultant algorithm that combines the spatial domain features with these gradient magnitude features is referred to as Gradient Image Quality Assessment algorithm (**G-IQA**). The features

(a) MOS = 40.47 (b) MOS = 49.23 (c) MOS = 52.80

Fig. 1: Image of the same scene tone-mapped using three different versions. (a) Method 1 (Durand TMO [9]), (b) Method 2 (Fattal TMO [10]), and (c) Method 3 (Reinhard TMO [11]). The caption of each image shows the MOS.

TABLE I: Spearman's Rank Ordered Correlation Coefficient (SROCC) and Pearson's Linear Correlation Coefficient (PLCC) between each feature and DMOS across 50 train-test (4:1) combinations on the ESPL-LIVE HDR Database over a single image scale and considering the L-component. Low correlations between each individual feature and DMOS show that the features complement each other

Domain	Feature Description	SROCC	PLCC
Spatial	Shape and Scale parameters of the GGD fitted to the MSCN coefficients (9) $[f_1 - f_2]$	0.238	0.266
Spatial	Shape and Scale parameters of the GGD fitted to the log-derivative of the seven types of neighbors $[f_3 - f_{16}]$	**0.439**	**0.436**
Spatial	Two parameters extracted from the σ-field (11) $[f_{17} - f_{18}]$	0.369	0.358
Gradient	Shape and Scale parameters of the GGD fitted to the MSCN coefficients of gradient magnitude field $[f_{19} - f_{20}]$	0.250	0.277
Gradient	Shape and Scale parameters of the GGD fitted to the log-derivative of the seven types of neighbors of gradient magnitude field $[f_{21} - f_{34}]$	0.386	0.384
Gradient	Two parameters extracted from the σ-field of gradient magnitude field $[f_{35} - f_{36}]$	0.388	0.392

are computed over two image scales and across the three channels in LAB color space. G-IQA (L) indicates a version of the proposed method using only the luminance channel (L).

III. Results

This section outlines the results of evaluating the performance of state-of-the-art NR-IQA algorithms on the ESPL-LIVE HDR Database. The performance of the proposed algorithm has been evaluated by measuring correlation with subjective scores (after non-linear regression) and the results have also been analyzed to determine statistical significance.

Once the features were extracted, a mapping is obtained from the feature space to the DMOS scores using a regression method, which provides a measure of the perceptual quality. We used a support vector machine regressor (SVR), specifically LibSVM [15] to implement ϵ-SVR with the radial basis function kernel, and γ is by default the inverse of the number of features.

A. Experiments on ESPL-LIVE HDR Database

The proposed algorithms G-IQA and G-IQA(L) has been evaluated on the 1,811 images of the ESPL-LIVE HDR database. The database consists of 747 tone-mapped images (obtained by the algorithms outlined in [9] [10] [11] [16]), 710 images created by multi-exposure fusion (using local and global energy methods and the ones described in [17] [18] [19]), and 354 images obtained by varying different post-processing parameters in Photomatix, which is a software to create HDR content. The original multiple exposure stacks were captured using Canon Rebel T5, Nikon D2x, and Nikon D5300 digital SLR cameras and displayed at a resolution of 960 $\times$ 540 for landscape and 540 $\times$ 304 for portrait orientations. The images have been evaluated by human subjects in the Amazon Mechanical Turk crowdsourcing platform. Overall 327,720 image evaluations have been gathered from 5,462 unique participants. On an average, each image has been evaluated by 110 subjects.

We randomly split the data into disjoint training and testing sets at 4:1 ratio and the split was randomized over 100 trials. Care was taken to ensure that the same source scene does not appear in the training and the testing sets in order to prevent artificial inflation of the results. The Spearman's rank ordered correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC) values between the predicted and the ground truth quality scores for every iteration and the median value of the correlations were reported. We discovered that there is significant room for improvement in using the present NR-IQA metrics to predict HDR artifacts. The results are summarized in Table II.

Table I shows the correlation of each type of feature with the MOS on the ESPL-LIVE HDR database. The low correlations between each individual features and the MOS indicate the need to combine complementary feature in order to predict the quality scores of image inflicted with a wide

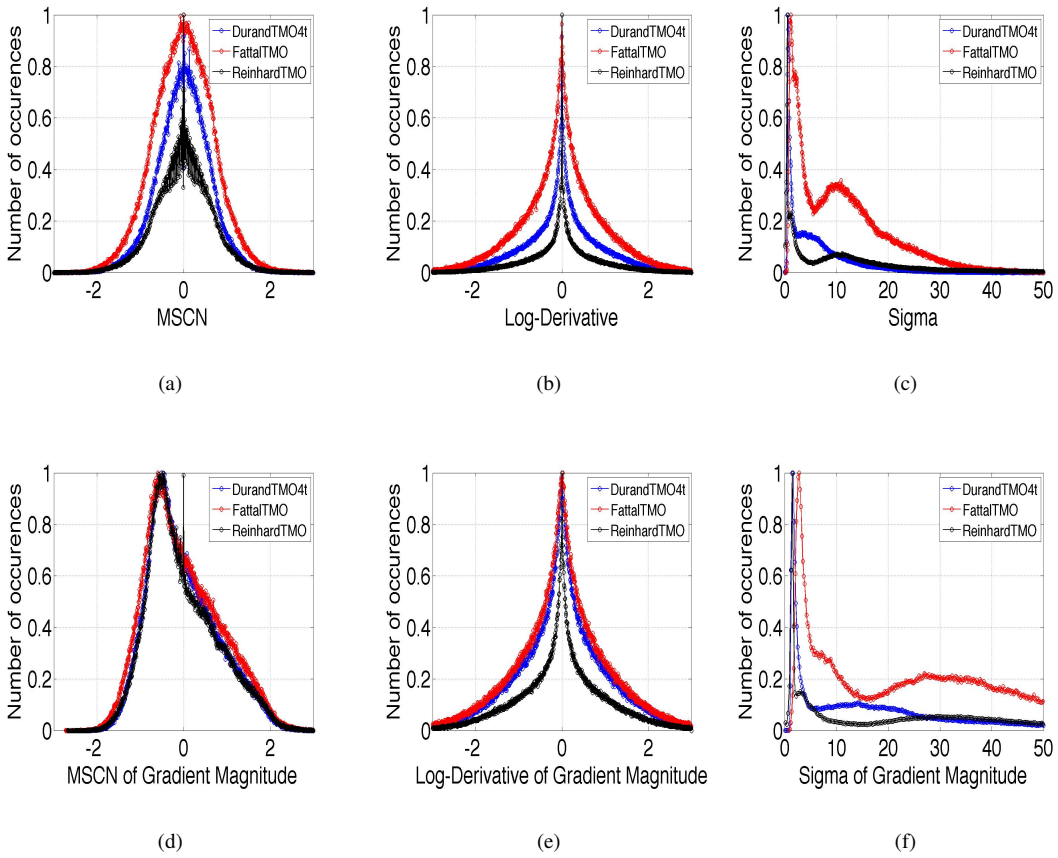

Fig. 2: Histograms of (a) MSCN pixels, (b) Log-derviatives of the MSCN pixels (c) σ-field of the pixels (d) MSCN coefficients of the gradient magnitude field (e) Log-derviatives of the MSCN coefficients of the gradient magnitude field (f) σ-field of the gradient magnitude field. The legends "Method 1", "Method 2", and "Method 3" represents processing by Durand TMO [9], Fattal TMO [10], and Reinhard TMO [11] respectively for the images shown in Figure 1.

range of artifacts. Figure 2 shows three images of the same scene tone-mapped using three different versions. Each of the tonemapping operators give rise to visually distinctly different images. It also shows the corresponding changes in the features extracted from the different domains.

Table III shows the root-mean-squared-error (RMSE), reduced $\tilde{\chi}^2$ statistic between scores predicted by the algorithms and the MOS for various algorithms (after logistic function fitting) and outlier ratio (expressed in percentage). Top performing algorithms show lower RMSE and outlier ratio values.

Fig. 3 shows box plots of the distribution of Spearman's Rank Ordered Correlation Coefficient values for each of the 100 trials of random train-test splits on the ESPL-LIVE HDR Image Database. This enable us to study the robustness of performance of the algorithms with variations of the choice of the training set. The proposed method shows smaller variation in the degree of correlation with human subjective evaluation.

To analyze the variation of SROCC between the scores predicted by the algorithm and the DMOS, the percentage of train/test splits was varied from 90% of the content used for training and the remaining 10% used for testing to 10% for training and 90% for testing. The knee of the curve occurs roughly at 60:40 train:test splits. This shows that the results

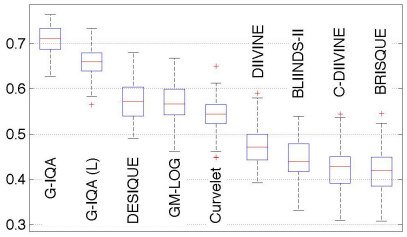

Fig. 3: Box plot of SROCC of learning based NR-IQA algorithms on images in the ESPL-LIVE HDR Image Database for 4:1 train-test splits over 100 trials. For each box, median is the central box, edges of the box represent the 25th and 75th percentiles, the whiskers span the most extreme non-outlier data points, and the outliers are plotted individually.

are not affected by overfitting or underfitting to the training data. Figure 4 shows the results.

B. *Determination of Statistical Significance*

For this purpose, eight representative NR-IQA algorithms were selected. The statistical significance tests were carried out for multiple training-test splits, using different 4:1 train-test splits of the database each time, and similar results

TABLE II: Median Spearman's Rank Ordered Correlaton Coefficient (SROCC) and Pearson's Linear Correlation Coefficient (PLCC) between the algorithm scores for various IQA algorithms and the MOS scores for ESPL-LIVE HDR database. The table has been sorted in the descending order of SROCC of the 'Overall category'. The numbers within parentheses in the "Overall" category show the confidence intervals on correlation values, computed by over 100 trials. Red indicates the proposed method. The bold values indicate the best performing algorithm.

	IQA	Tone Mapping		Multi-Exposure Fusion		Post Processing		Overall	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
1	G-IQA	**0.728**	**0.764**	**0.711**	**0.705**	**0.616**	**0.643**	**0.719** (0.671, 0.766)	**0.718**(0.652, 0.776)
2	G-IQA (L)	0.672	0.702	0.634	0.637	0.551	0.582	0.661 (0.595, 0.732)	0.658(0.590, 0.738)
3	DESIQUE	0.542	0.553	0.572	0.584	0.529	0.563	0.570 (0.481, 0.657)	0.568(0.467, 0.650)
4	GM-LOG	0.549	0.562	0.545	0.541	0.578	0.599	0.556 (0.448, 0.638)	0.557(0.465, 0.639)
5	CurveletQA	0.584	0.623	0.517	0.535	0.481	0.506	0.547 (0.458, 0.610)	0.560(0.447, 0.631)
6	DIIVINE	0.523	0.530	0.453	0.472	0.392	0.447	0.482 (0.326, 0.578)	0.484(0.331, 0.583)
7	BLIINDS-II	0.412	0.442	0.446	0.459	0.486	0.510	0.444 (0.310, 0.519)	0.454(0.326, 0.545)
8	C-DIIVINE	0.453	0.453	0.423	0.460	0.432	0.470	0.434 (0.265, 0.551)	0.444(0.277, 0.538)
9	BRISQUE	0.340	0.370	0.494	0.516	0.468	0.483	0.418 (0.300, 0.500)	0.444(0.313, 0.528)

TABLE III: Root-mean-square error (RMSE), reduced $\tilde{\chi}^2$ statistic between the algorithm scores and the DMOS for various NR-IQA Algorithms (after logistic function fitting) and outlier ratio (expressed in percentage) for each distortion category for ESPL-LIVE HDR database. Red indicates the proposed method. The bold values indicate the best performing algorithm for that category.

	IQA	Tone Mapping			Multi-Exposure Fusion			Post Processing			Overall		
		RMSE	$\tilde{\chi}^2$	OR	RMSE	$\tilde{\chi}^2$	OR	RMSE	$\tilde{\chi}^2$	OR	RMSE	$\tilde{\chi}^2$	OR
1	G-IQA	**6.711**	9.908	**0.000**	**6.884**	21.155	**0.000**	**6.884**	2.376	**0.000**	**7.033**	13.918	**0.275**
2	G-IQA (L)	7.434	8.624	0.662	7.484	5.263	**0.000**	7.308	3.131	**0.000**	7.628	12.558	0.552
3	DESIQUE	8.577	12.079	0.683	7.862	11.588	0.687	7.402	1.851	**0.000**	8.296	19.614	0.829
4	GM-LOG	8.632	5.002	1.170	8.028	15.027	0.702	7.420	0.851	**0.000**	8.357	20.659	0.829
5	CurveletQA	8.177	17.408	0.694	8.054	10.754	0.714	7.922	2.892	**0.000**	8.511	15.253	0.829
6	DIIVINE	8.805	10.025	0.791	8.371	5.663	0.667	7.979	2.659	**0.000**	8.821	12.115	0.829
7	BLIINDS-II	9.330	7.565	0.697	8.517	19.979	0.752	7.818	1.976	**0.000**	8.975	21.948	0.828
8	C-DIIVINE	9.167	15.338	1.356	8.485	8.374	0.671	7.852	1.428	**0.000**	8.983	12.305	0.966
9	BRISQUE	9.535	16.712	1.356	8.227	5.681	0.685	7.894	7.146	**0.000**	9.049	17.259	0.831

TABLE IV: Results of the F-test performed on the residuals between model predictions and MOS scores on ESPL-LIVE HDR database. Each cell in the table is a codeword consisting of 4 symbols that correspond to "Tone Mapping Operators"', "Multi-Exposure Fusion", "Post Processing", and "Overall" distortions. "1"("0") indicates that the performance of the row IQA is superior(inferior) to that of the column IQA. - indicates that the statistical performance of the row IQA is equivalent to that of the column IQA. The matrix is symmetric. Red indicates the proposed methods.

	G-IQA	DESIQUE	BRISQUE	GM-LOG	C-DIIVINE	DIIVINE	BLIINDS-II	CurveletQA
G-IQA	- - - -	1 - - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1
DESIQUE	0 - - 0	- - - -	1 - - 1	- - - -	- - - -	- - - -	- - - -	- - - -
BRISQUE	0 0 - 0	0 - - 0	- - - -	- - - -	- - - -	- - - -	- - - -	0 - - -
GM-LOG	0 0 - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
C-DIIVINE	0 0 - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
DIIVINE	0 0 - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	0 - - -
BLIINDS-II	0 0 - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
CurveletQA	0 0 - 0	- - - -	1 - - -	- - - -	- - - -	1 - - -	- - - -	- - - -

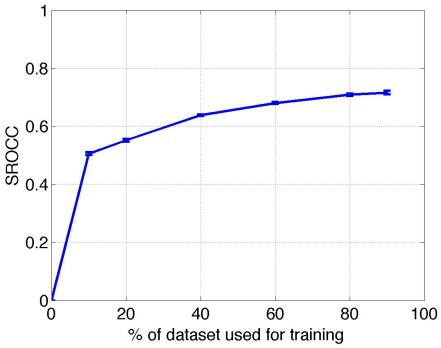

Fig. 4: Mean SROCC between predicted and subjective DMOS scores for G-IQA (and the associated 95% confidence intervals) as a function of the percentage of the content used for training on images in the ESPL-LIVE HDR Image Database over 50 trials.

were obtained. The table outlines the results obtained for one such representative trial. To determine whether the IQA algorithms are significantly different from each other, the F-statistic, as in [20], was used to determine the statistical significance between the variances of the residuals after a non-linear logistic mapping between the two IQA algorithms, at the 95% confidence interval. Table IV shows the results for eight selected IQA algorithms and all distortions. Overall, the proposed algorithm is found to be statistically superior to the other NR-IQA algorithms.

C. Experiments on other databases

In addition to the ESPL-HDR database, for the sake of completeness, the performance of the proposed algorithm has also been tested on the legacy LIVE database [20]. Table V shows the performance of the proposed algorithm on the LIVE database [20]. Similar technique of splitting the data into disjoint training and testing sets at 4:1 ratio, randomized over 100 trials, was followed. The high degrees of correlation with the subjective data shows that the proposed methods can also

TABLE V: Median Spearman's Rank Ordered Correlation Coefficient (SROCC) and Pearson's Linear Correlation Coefficient (PLCC) between algorithm scores and DMOS for various NR-IQA algorithms across 100 train-test (4:1) combinations on the LIVE Database of natural images. Performances of some FR-IQA algorithms (shown in italics) have been included for comparison. The table has been sorted in the descending order of SROCC of the 'Overall category'. The numbers within parentheses in the "Overall" category show the confidence intervals on correlation values, computed by over 100 trials. Bold values indicate the best performing algorithm for that category. Red indicates the proposed method. Italics indicate FR-IQA algorithms.

	IQA	JP2K		JPEG		Gaussian Noise		Blur		Fast Fading		Overall	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
1	GM-LOG	0.882	0.904	0.878	**0.917**	0.978	0.988	0.915	0.925	**0.899**	**0.917**	**0.914** (0.860, 0.941)	**0.917**(0.857, 0.942)
2	G-IQA	0.905	**0.914**	**0.883**	0.915	**0.983**	**0.990**	0.917	0.925	0.836	0.860	0.906 (0.788, 0.952)	0.907(0.786, 0.952)
3	BRISQUE	0.878	0.888	0.852	0.889	0.962	0.975	**0.941**	**0.942**	0.863	0.887	0.902 (0.798, 0.950)	0.900(0.786, 0.949)
4	C-DIIVINE	0.872	0.882	0.839	0.876	0.965	0.974	0.915	0.915	0.891	0.915	0.898 (0.817, 0.944)	0.905(0.816, 0.945)
5	BLIINDS-II	**0.907**	0.912	0.846	0.884	0.939	0.960	0.906	0.918	0.884	0.902	0.897 (0.775, 0.938)	0.900(0.746, 0.946)
6	DESIQUE	0.875	0.893	0.824	0.869	0.975	0.985	0.908	0.925	0.829	0.865	0.878 (0.805, 0.944)	0.884(0.797, 0.938)
7	G-IQA (L)	0.848	0.853	0.839	0.870	0.955	0.960	0.865	0.891	0.788	0.836	0.866 (0.721, 0.934)	0.861(0.710, 0.930)
8	CurveletQA	0.816	0.824	0.827	0.836	0.969	0.979	0.896	0.900	0.826	0.866	0.863 (0.694, 0.916)	0.859(0.493, 0.911)
9	DIIVINE	0.824	0.828	0.759	0.798	0.937	0.950	0.854	0.888	0.759	0.792	0.827 (0.451, 0.924)	0.829(0.452, 0.919)
10	GRNN	0.816	0.822	0.765	0.748	0.916	0.939	0.877	0.896	0.816	0.861	0.776 (0.652, 0.833)	0.784(0.688, 0.854)
11	BIQI	0.668	0.689	0.580	0.612	0.776	0.782	0.744	0.783	0.567	0.578	0.634 (0.173, 0.811)	0.642(0.194, 0.815)
12	*MS-SSIM*	0.963	0.975	0.979	0.979	0.977	0.988	0.954	0.965	0.939	0.949	0.954	0.951
13	*SSIM*	0.939	0.941	0.947	0.946	0.964	0.982	0.905	0.900	0.939	0.951	0.913	0.907
14	*PSNR*	0.865	0.876	0.883	0.903	0.941	0.917	0.752	0.780	0.8736	0.880	0.864	0.859

capture the processing, compression and transmission artifacts arising in SDR images.

IV. Conclusion

In this paper, we describe the different spatial domain features extracted in the proposed scene-statistics based NR-IQA algorithm and show that these features are effective in capturing the visual artifacts arising from the different HDR algorithms. This paper evaluates a new no-reference image quality assessment algorithm using spatial and gradient domain features. Based on a large crowdsourcing study of HDR images with 5,462 unique participants and 327,720 image evaluations, the proposed algorithm is the most effective algorithm at quantifying human perceptual responses of visual artifacts, among the 9 NR-IQAs evaluated against the subjective test results.

V. Acknowledgment

We would like to thank the Texas Advanced Computing Center at The University of Texas at Austin for letting us use the high-performance computing resources in order to generate the HDR-processed images.

References

[1] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. on Image Process.*, vol. 22, no. 2, pp. 657–667, Feb 2013.

[2] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct 2015.

[3] H. Ziaei Nafchi, A. Shahkolaei, R. Farrahi Moghaddam, and M. Cheriet, "FSITM: A feature similarity index for tone-mapped images," *IEEE Signal Process. Lett*, vol. 22, no. 8, pp. 1026–1029, Aug 2015.

[4] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. on Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov 2015.

[5] W. S. Geisler, "Visual Perception and the Statistical Properties of Natural Scenes," *Annual Review of Psych.*, vol. 59, no. 1, pp. 167–192, 2008.

[6] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, 2001.

[7] D. Kundu, "Subjective and objective quality evaluation of synthetic and high dynamic range images," Ph.D. dissertation, Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, May 2016, http://users.ece.utexas.edu/%7Ebevans/students/phd/debarati_kundu/.

[8] Y. Zhang and D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," *J Electronic Imaging*, vol. 22, no. 4, 2013.

[9] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. ACM SIGGRAPH*, 2002, pp. 257–266. [Online]. Available: http://doi.acm.org/10.1145/566570.566574

[10] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, Jul. 2002. [Online]. Available: http://doi.acm.org/10.1145/566654.566573

[11] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002. [Online]. Available: http://doi.acm.org/10.1145/566654.566575

[12] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[13] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer." *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.

[14] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb 2014.

[15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Systems and Technology*, vol. 2, 2011, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. on Vis. and Comp. Graphics*, vol. 3, no. 4, pp. 291–306, Oct. 1997. [Online]. Available: http://dx.doi.org/10.1109/2945.646233

[17] S. Paul, I. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain," *Journal of Circuits, Systems, and Comp.s, submitted*, 2016. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/48782-multi-exposure-and-multi-focus-image-fusion-in-gradient-domain

[18] F. Pece and J. Kautz, "Bitmap movement detection: Hdr for dynamic scenes," *Journal of Virtual Reality and Broadcasting*, vol. 10, no. 2, 2013. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:0009-6-36506

[19] S. Raman and S. Chaudhuri, "Bilateral Filter Based Compositing for Variable Exposure Photography," in *Eurographics - Short Papers*, P. Alliez and M. Magnor, Eds. The Eurographics Association, 2009.

[20] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov 2006.