No-Reference Quality Assessment of Tone-Mapped HDR Pictures

Debarati Kundu, Deepti Ghadiyaram, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, and Brian L. Evans, *Fellow, IEEE*

***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 b/color/pixel. High dynamic range (HDR) images, which are usually created from multiple exposures of the same scene, can provide 16 or 32 b/color/pixel, but must be tonemapped to SDR for display on standard monitors. Multi-exposure fusion 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 HDR IMAGE GRADient-based Evaluator. 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. Toward ameliorating this we make here the following contributions: 1) we design HDR picture NR IQA models and algorithms 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 models also perform well on legacy natural SDR images. The software is available at: http://live.ece.utexas.edu/research/Quality/higradeRelease.zip.**

***Index Terms*— Image quality assessment, high dynamic range, natural scene statistics, no-reference.**

I. Introduction

RECENT years have seen a huge growth in the acquisition, processing and transmission of digital pictures and videos. While most pictures are still represented as Standard Dynamic Range (SDR) images of 8 bits/color/pixel acquired using a fixed exposure, there is a growing practice of acquiring/creating and displaying high dynamic range (HDR) images and other types of pictures created by multiple exposure fusion. These kinds of images allow for more pleasing representation and better use of the available luminance and color ranges in real scenes, which can range from direct sunlight to faint starlight [1]. HDR content can be captured by smart phones and digital SLR cameras, streamed by video-on-demand services, and displayed by modern monitors.

HDR images which are commonly represented by 16 or 32 bits/color/pixel, are typically obtained by blending a stack of SDR images at varying exposure levels, thereby allowing a range of intensity levels on the order of 10,000 to 1. HDR rendering also finds use in computer graphics, where lighting calculations are performed over a wider dynamic range. This results in a better contrast distribution, thereby leading to a higher degree of detail preservation. However, in order to visualize these images on standard SDR display devices, they must be tonemapped to SDR. In addition to tone-mapped SDR images, images are also often created by multi-exposure fusion, where a stack of SDR images taken at varying exposure levels are fused to create an SDR image that is more visually informative than the input images. This bypasses the intermediate step of creating an HDR irradiance map. HDR images may also be post-processed (color saturation, color temperature, detail enhancement, etc.) for aesthetic purposes.

Since different fusion/tone-mapping algorithms result in different SDR images, a natural question is how to evaluate the quality of the images obtained. Subjective testing is important to evaluate the visual quality of images produced by different algorithms. Recently we created the ESPL-LIVE HDR Image Quality Database comprising of more than 1,800 HDR-processed images [2] obtained by tone-mapping and multi-exposure fusion with or without post-processing. We conducted extensive subjective experiments on more than 5,000 observers using Amazon's online crowdsourcing platform, Mechanical Turk.

A highly desirable goal is to design objective quality prediction models that automate the process of IQA, and that can be classified into full-reference (FR) and no-reference (NR) categories. FR-IQA algorithms have been proposed that evaluate tone-mapped SDR images [1], [3], [4] and SDR images created by multi-exposure fusion [5] based on the principles of structural similarity. However, when evaluating the quality of images created by HDR processing algorithms, it is hard to assume a 'reference' image, since the input to the algorithms is an exposure stack that may have a varying number of images of unknown qualities based on the camera settings used. In these applications, NR IQA is the only realistic option.

Manuscript received July 3, 2016; revised November 13, 2016 and February 14, 2017; accepted March 7, 2017. Date of publication March 22, 2017; date of current version April 28, 2017. This work was supported in part by the Special Research Grant, Vice President for Research, The University of Texas at Austin, and in part by the National Science Foundation Grant under Award 1116656. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dacheng Tao.

D. Kundu was with Embedded Signal Processing Laboratory, The University of Texas at Austin, Austin, TX 78712 USA. She is now with Qualcomm Research, Bangalore 560066, India (e-mail: debarati@utexas.edu).

D. Ghadiyaram and A. C. Bovik are with the Laboratory for Image and Video Engineering, The University of Texas at Austin, Austin, TX 78712, USA (e-mail: deepti@cs.utexas.edu; bovik@ece.utexas.edu).

B. L. Evans is with the Embedded Signal Processing Laboratory, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bevans@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2685941

1057-7149 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The most successful NR IQA algorithms for SDR images deploy Natural Scene Statistics (NSS) models as essential building blocks [6]. NSS models are based on the observation that pristine real-world optical images obey certain statistical principles ('naturalness') [7], [8] that are violated by the presence of distortions ('unnaturalness'). NR IQA algorithms that extract NSS features usually train a kernel function to map the features to ground-truth human subjective scores using a supervised learning framework. These techniques have shown remarkable correlations with ground truth human subjective scores on SDR images containing distortions such as blur, additive noise, and compression on legacy IQA databases like LIVE IQA Database [9], TID2013 [10], and so on.

To the best of our knowledge, NSS models have not yet been used by other authors to create IQA models for NR evaluation of tonemapped HDR image quality. Towards filling this gap, we propose two NSS based NR IQA models that deliver good predictive performance of the perceptual quality of tonemapped HDR images. Specifically, we make the following contributions:

1) We develop new natural scene-statistics based NR IQA models that embody the fundamental NSS modeling steps of extracting pointwise and pairwise bandpass log-derivative statistics from image pixels (following a local divisive-energy normalization step). We significantly elevate performance by also extracting gradient image features, including a gradient structure tensor NSS feature. The use of gradient-domain NSS is motivated by the fact that most HDR processing tone-mapping and multi-exposure fusion algorithms modify the local gradients of the input images to achieve aesthetically appealing contrast distributions.
2) We evaluate the performance of the proposed algorithms against other leading NR-IQA models on the new ESPL-LIVE HDR Image Quality Database using measures of correlation, outlier ratio, and root-mean-square error. We also conduct statistical significance tests to validate the conclusions reached. We find that the proposed models significantly outperform the compared state-of-the-art NR IQA algorithms.
3) We also tested performance of the proposed models/algorithms on legacy SDR image databases and found that they perform extremely well; hence including HDR distortion-specific sensitivity does not reduce their performance on SDR content.

The remainder of the paper is organized as follows. Section II outlines related work on NR IQA. Details of the proposed NR IQA algorithms, including the HDR-specific features extracted are described in Section III. Section IV evaluates the performance of the new algorithms on the ESPL-LIVE HDR Image Quality Database and as well as on legacy natural image databases and discusses the results. Section V concludes the paper.

II. Related Work

NR IQA models may be conveniently divided into two categories. The first category aims to evaluate the quality of an image by assuming the presence of a specific type of distortion. For example, JNBM [11], CPBDM [12], and LPCM [13] aim to perform perceptual blur estimation, while algorithms like NJQA [14] and JPEG-NR [15] respectively quantify the presence of noise and JPEG compression artifacts. In [16] Fang *et al.* proposes an NR IQA algorithm that conducts perceptual quality evaluation on contrast-modified images.

By contrast, another category of NR IQA models that are based on NSS generally make no assumptions with respect to feature design regarding the type of distortions present in the images. These models are often learned, and have proved versatile for evaluating the quality of images afflicted by a wide range of distortions or processing artifacts. As was systematically observed in [17], the local mean-subtracted-contrast-normalized (MSCN) coefficients of good quality natural images tend to follow a Gaussian-like distribution [17]. The distribution of MSCN coefficients and products of adjacent pairs of them were successfully employed in the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [18] and the Natural Image Quality Evaluator (NIQE) [19]. The Derivative Statistics-based QUality Evaluator (DESIQUE) [20] supplements BRISQUE by using log-derivative distributions of MSCN pixels. Similarly, a normalized Gaussian-smoothed gradient magnitude (GM) map and a normalized Laplacian of Gaussian (LOG) map are used in the NR IQA GM-LOG model [21]. The gradient map also finds its application in the calculation of local binary pattern used in [22]. Neurons employed in early stages of the visual pathway capture information over multiple orientations and scales, motivating multiscale processing in many NSS-based NR IQA models: log-Gabor decomposition (DESIQUE [20]), Daubechies 9/7 wavelets (BIQI [23]), steerable pyramid wavelets (DIIVINE [24], C-DIIVINE [25]), DCT (BLIINDS-II [26]), and curvelets (CurveletQA [27]). In addition, features like phase congruency (GRNN [28]) and expected image entropy upon a set of predefined directions (Anisotropy [29]) have also been used in NR-IQA. Models like COdebook Representation for No-Reference Image Assessment (CORNIA) [30] uses a supervised learning technique to learn a dictionary of different distortions from the raw image patches instead of using a fixed set of features. Mittal *et al.* [31] applies a "topic model" to the visual words extracted from the pristine and distorted images. The Bag-of-Words model has also been used in [32]. In addition, blind IQA models that use sophisticated deep learning models have also been proposed in [33] (DIQI), [34] (DLIQA-R), [35] (CNN), and [36].

Most of these NR IQA algorithms have been trained and evaluated on the artificially distorted images contained in legacy SDR image databases. Although these algorithms deliver excellent correlations against human ground truth subjective scores on these type of images, they perform poorly on the images produced by HDR processing algorithms, as we will show in the following sections.

III. Proposed Algorithm

Here we describe the perceptually relevant features that are employed in our proposed NR IQA models. The features include established NSS quality descriptors as well as new features designed to handle processing artifacts arising in HDR

processed images. These are based on the observation that NSS features extracted from image gradients are modified by HDR processing artifacts in characteristic ways. These deviations may be used to improve predictions of human subjective responses. In the following subsections, $I(i, j)$ at the (i, j)-th coordinate could be any color channel of an image, represented in any common luminance or color space such as LAB, LMS, or YUV, as specifically studied later in Section IV-E.

A. Log-Derivatives

A recent state-of-the-art NR IQA algorithm called DESIQUE [20] uses log-derivative (log gradient) features to predict the quality of natural images afflicted by (non-HDR) processing artifacts. The log-derivative statistics are computed after logarithmically converting the image luminances (or other color channel values in some color space) [20]. We also deploy these features in our models. Let $M \times N$ be the dimension of the image I, and $I(i, j)$ be the value at the (i, j)-th spatial location, $i \in \{1, 2, \ldots, M\}$, $j \in \{1, 2, \ldots, N\}$. The logarithmic image is then:

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

where C is a small constant added to avoid numerical instabilities. The following log-gradient values are computed as in [20]:

$$D1 : \nabla_x J(i, j) = J(i, j + 1) - J(i, j), \tag{2}$$
$$D2 : \nabla_y J(i, j) = J(i + 1, j) - J(i, j), \tag{3}$$
$$D3 : \nabla_{xy} J(i, j) = J(i + 1, j + 1) - J(i, j), \tag{4}$$
$$D4 : \nabla_{yx} J(i, j) = J(i + 1, j - 1) - J(i, j), \tag{5}$$
$$D5 : \nabla_x \nabla_y J(i, j) = J(i - 1, j) + J(i + 1, j) - J(i, j - 1) - J(i, j + 1), \tag{6}$$
$$D6 : \nabla_{cx} \nabla_{cy} J(i, j)_1 = J(i, j) + J(i + 1, j + 1) - J(i, j + 1) - J(i + 1, j), \tag{7}$$
$$D7 : \nabla_{cx} \nabla_{cy} J(i, j)_2 = J(i - 1, j - 1) + J(i + 1, j + 1) - J(i - 1, j + 1) - J(i + 1, j - 1). \tag{8}$$

B. Spatial Domain Scene Statistics

Similar to the BRISQUE [18] and NIQE [19] models, we also model the scene statistics of images processed by mean subtraction and divisive normalization operators (MSCN). Let $M \times N$ be the dimensions of an image I, and $I(i, j)$ be the luminance value in the (i, j)-th spatial location, $i \in \{1, 2, \ldots, M\}$, $j \in \{1, 2, \ldots, N\}$. Then MSCN values 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 $w = \{w_{k,l} | k = -K, \ldots, K, l = -L, \ldots, L\}$ is a symmetric local Gaussian convolution window centered at the (i, j)-th pixel. K and L determine the size of the local patches used in the calculation of the mean and standard deviation. In [18], the authors used 7×7 image patches, and a circularly symmetric 2D Gaussian kernel; however, our experiments indicate that the distributions of the MSCN patches are not very sensitive to the window size.

Generally, as in [18], [19], and [24], we may view the distribution of (9) as a tool for studying the loss of naturalness of an image, due to distortion or processing. It is effective for measuring tone-mapping effects. The normalized image ($\hat{I}$) of a good quality, unprocessed pristine image will have an empirical probability distribution or a histogram that is Gaussian like. The standard deviation map σ highlights object boundaries and other details of the image. The MSCN values of both original and distorted images are commonly modeled using a Generalized Gaussian Distribution (GGD) when used for image quality assessment [18], [19].

The zero mean Generalized Gaussian Distribution (GGD) used to model MSCN coefficients $\hat{I}(i, j)$ is given by:

$$f(x; \alpha, \gamma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} exp\left[-\left(\frac{|x|}{\beta}\right)^{\alpha}\right] \tag{12}$$

where $\beta = \gamma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$ and $\Gamma(x) = \int_0^{\infty} t^{(x-1)} e^{-t}, x > 0$ is the gamma function. The variables α and γ are shape and scale parameters, respectively, and are valuable quality-aware features that capture deviations of the image statistics corresponding to HDR processing artifacts.

Local adaptation methods [40], [41] that tone-map the HDR irradiance maps to SDR images work on the principle of dividing each pixel by its average luminance, approximating the"dodge-and-burn" technique employed by photographers [42]. This corresponds to subtracting a low-pass filtered version from the original image in the logarithm domain. This further motivates the use of NSS features extracted from the log-derivatives of the MSCN coefficients to evaluate the quality of HDR-processed images. Log-derivatives of the MSCN coefficients modified by(2) -(8) are also well modeled by as GGD. The shape (α) and scale (γ) parameters of GGD fits to the seven types of log-derivatives are useful quality aware features.

The σ-field has previously been shown to provide effective quality-aware NSS features [43]. We extract two quantities from the σ-field: the mean value and the square of the reciprocal of the coefficient of variation. These quantities are respectively defined as:

$$\Phi_{\sigma} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sigma(i, j), \tag{13}$$

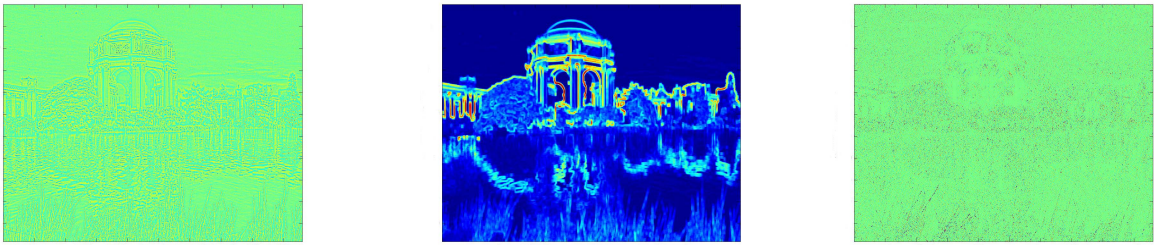
$$\Sigma_{\sigma}(i, j) = \sqrt{\sum_{k=0}^{M-1} \sum_{j=0}^{N-1} [\sigma(i + k, j + l) - \Phi_{\sigma}]^2}, \tag{14}$$

and

$$\Psi_{\sigma} = \left(\frac{\Phi_{\sigma}}{\Sigma_{\sigma}}\right)^2. \tag{15}$$

(a) (b) (c)

Fig. 1. Image of the same scene processed using several different HDR/tone-mapping algorithms obtained from the ESPL-LIVE HDR Image Quality Database (details are available in IV-A). (a) Raman TMO [37], (b) Ward TMO [38], and (c) post-processing by 'Surreal' effect in Photomatix [39]. Each of the processing algorithms give rise to visually distinctive images. The caption of each image shows the associated Mean Opinion Score (MOS).



(a) (b) (c)

Fig. 2. (a) MSCN coefficients (b) Local standard deviation and (c) Log-derivative of the MSCN coefficients map of the luminances of the images in Fig. 1(c).

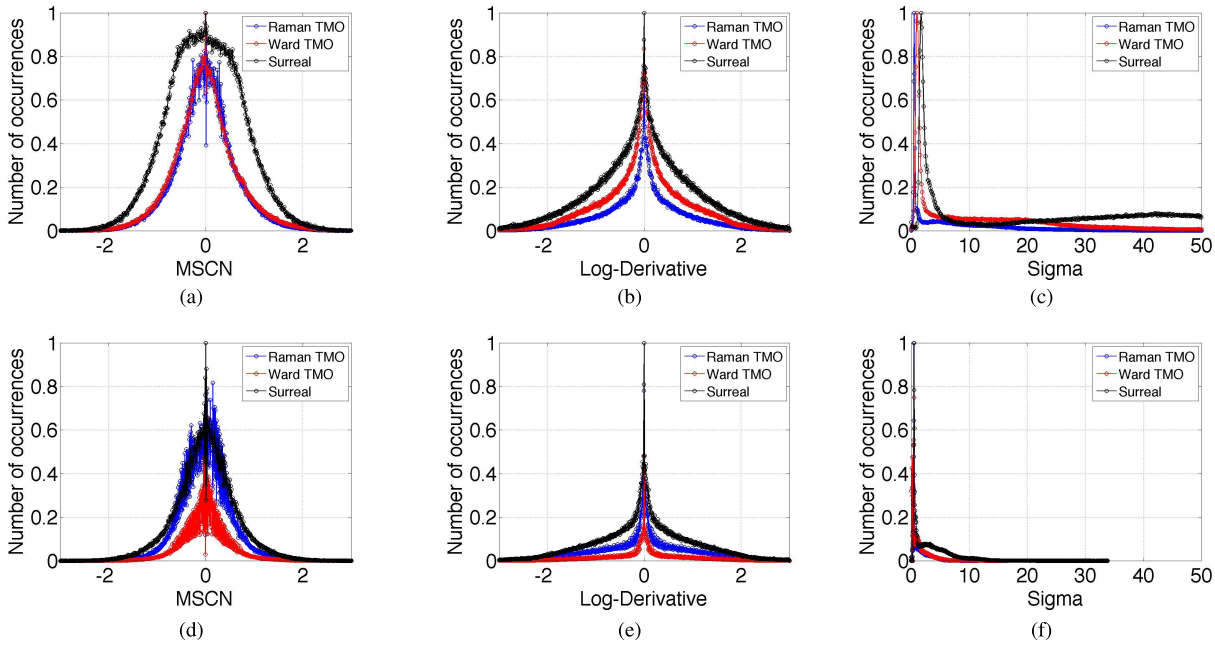


Fig. 3. Histograms of (a) MSCN pixels (b) Log-derivatives of MSCN pixels (c) σ-field of the L-channel of the pixels. (d)-(f) represent the corresponding histograms for the A-channel in the LAB color space. The three plots represent processing by Raman TMO [37], Ward TMO [38], and post-processing by 'Surreal' effect in Photomatix [39] respectively for the images shown in Figure 1.

Figure 1 shows three images of the same scene processed using three different HDR/tone-mapping algorithms. Figure 3 shows the histograms of the MSCN coefficients, the log-derivatives of the MSCN coefficients, and the σ-field of the images in Fig. 1. The spatial domain features extracted at each scale and in each color channel are summarized in Table I. The features are extracted in the three color channels of the LAB color space. In order to separately study the impact of the spatial domain features, we also created and tested a model we refer to as HIGRADE-S, which uses only the described spatial domain features to predict quality.

C. Gradient Domain Scene Statistics

Image gradients contain important information regarding the distribution of edges and variations of local contrast. They are used widely in IQA since they capture changes of local structures when images are distorted by post-acquisition processing techniques. Several state-of-the-art FR-IQA [4], [44], [45] and NR IQA [21] algorithms employ gradient information. Further, many HDR processing algorithms, such as tone-mapping and multi-exposure fusion algorithms, modify the gradients of images in the multi-exposure stacks, which results in modified contrast levels in the resultant fused images.

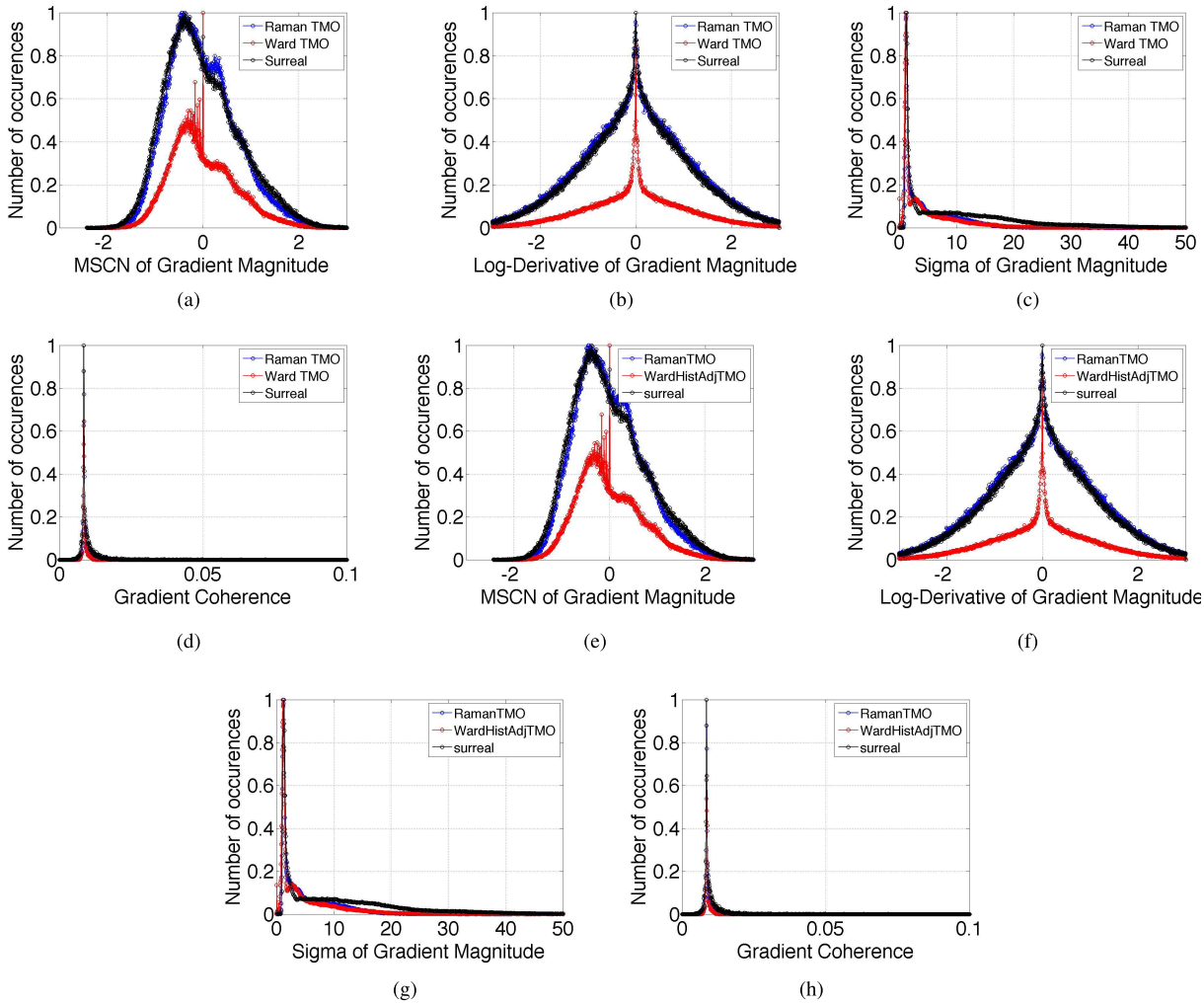


Fig. 4. Histograms of (a) MSCN coefficients (b) Log-derviatives of the MSCN coefficients (c) σ-field (d) Gradient structure tensor coherence of gradient field of the L-channel. (e)-(h) are the corresponding histograms for the A-channel. The legends represent processing by Raman TMO [37], Ward TMO [38], and post-processing by the 'Surreal' effect in Photomatix [39] on the images shown in Figure 1.

TABLE I
SPATIAL DOMAIN FEATURES EXTRACTED FOR EVERY COLOR CHANNEL USED IN THE PROPOSED IQA MODELS

Feature ID	Feature Description
$f_1 - f_2$	Shape and Scale parameters of GGD fit to the MSCN coefficients (9)
$f_3 - f_{16}$	Shape and Scale parameters of GGD fit to the log-derivative of the seven types of neighbors (III-A)
$f_{17} - f_{18}$	Two parameters extracted from the σ-field (11)

While gradient magnitude features are widely used in IQA, gradient orientation information has only been recently studied as an IQA feature [46]. We utilize both gradient magnitude and gradient orientation information extracted from the gradient structure tensor [47], as explained in the following.

1) Gradient Magnitude Features: The local gradient is computed by convolving the image with a simple Sobel operator. The horizontal (H_x) and vertical (H_y) components of the Sobel operator are given by:

$$H_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{16}$$

$$H_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{17}$$

The estimated gradient magnitude M of the image $I(i, j)$ is then given by:

$$M(i, j) = \sqrt{(I * H_x)^2(i, j) + (I * H_y)^2(i, j)}, \tag{18}$$

where $*$ denotes the convolution operator.

The same spatial domain features summarized in Table I are also extracted from the gradient magnitude field. An algorithm that combines the spatial domain features with the gradient magnitude features will be referred to as the HDR Image GRADient based Evaluator-1 (**HIGRADE-1**). Since many of the HDR processing algorithms modify the local gradients across multiple scales [48], we compute all of the image features over two image scales.

2) Gradient Structure Tensor Features: The structure tensor is a well known concept in signal and image processing. For example, in [49], Saad *et al.* used features extracted from a motion structure tensor to conduct objective quality evaluation of natural videos. The gradient structure tensor [47] captures the predominant gradient directions and degree of coherence

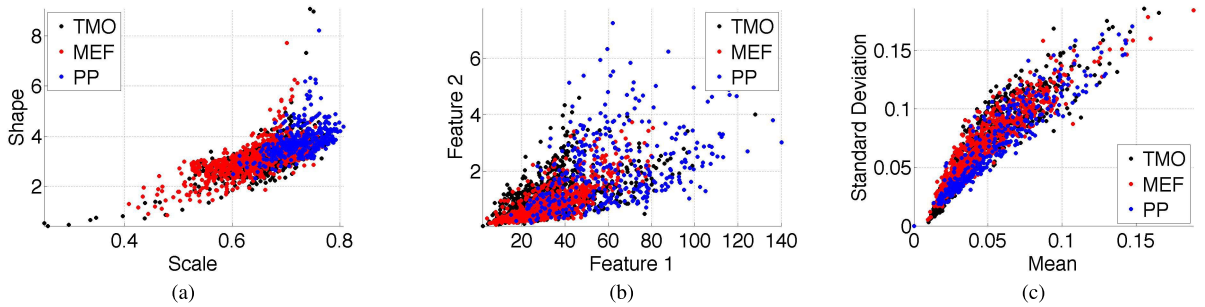


Fig. 5. 2-D scatter plot between (a) shape and scale parameters obtained by fitting GGD to the empirical distributions of MSCN coefficients of gradient magnitude of the image luminances (b) The features Φ_σ and Ψ_σ extracted from the local σ-map of gradient magnitude of the image luminances (c) Mean and standard deviation of the gradient structure tensor coherence for the images in the ESPL-LIVE HDR Image Quality Database. The legends 'TMO', 'MEF', and 'PP' show images processed by tone-mapping, multi-exposure fusion and post-processing algorithms respectively.

of the gradient vector field of an image. It has been used for image interpolation [50], anisotropic filtering [51], and motion detection [52]. It is particularly relevant for analyzing the quality of HDR/tonemapped images which have had gradient manipulations executed on them that can unnaturally modify the spatial distribution of the local image gradient. The 2D structure tensor is given by:

$$J = \begin{bmatrix} f(G_x) & f(G_x.G_y) \\ f(G_x.G_y) & f(G_y), \end{bmatrix} \tag{19}$$

where

$$f(V) = \sum_{l,k} w[i,j]V(i-l, j-k)^2 \tag{20}$$

and where $G_x(i,j)$ and $G_y(i,j)$ are smoothed horizontal and vertical second derivatives at coordinate (i,j), respectively, and w is a window of dimension of $P \times P$ over which the localized structure tensor is computed. The quantities $G_x(i,j)$ and $G_y(i,j)$ are implemented as directional difference-of-Gaussian filters. The relative discrepancy between the two eigenvalues indicates the degree of anisotropy of the local gradient. The coherence measure is defined by:

$$C = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 \tag{21}$$

where λ_1 and λ_2 are the two eigenvalues of the gradient structure tensor. The coherence measure is computed over $P \times P$ non-overlapping blocks of the image. The sample mean, standard deviation, skewness, and kurtosis values of the coherence measure are also used as features. An algorithm that combines the spatial domain features of Table I with the gradient structure tensor features just described is referred to as the HDR Image GRADient based Evaluator-2 (**HIGRADE-2**). The tensor features are also calculated over all three LAB chromatic channels.

To study the effect of using all of the gradient features, we defined a model using features $f_1 - f_{40}$ which we call HIGRADE-3. In order to motivate the gradient based features used in proposed HIGRADE algorithms, we plotted the variations of the different parameters extracted from the image gradients on images produced by different HDR processing algorithms. The scatter plot between the shape and scale parameters obtained by fitting the GGD to the empirical distributions of the MSCN coefficients of the gradient magnitudes of the luminances (of the images in the ESPL-LIVE HDR Image Quality Database) is shown in Fig. 5(a), while Fig. 5(b) shows the scatter plot between the features Φ_σ and Ψ_σ extracted from the corresponding local σ-map. The images obtained by the TMO and MEF algorithms show overlap of the gradient features, but the post-processed images show a difference in the features extracted. The variations of the mean and standard deviation of the local gradient structure tensor coherence computed across all the image blocks is shown in Fig. 5(c).

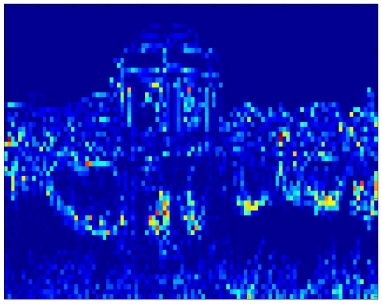


Fig. 6. Gradient structure tensor coherence of MSCN coefficients map of L-channel of Fig. 1(c).

Overall, 216 features constitute the HIGRADE-1 model while 120 features define the HIGRADE-2 model. HIGRADE-3 and HIGRADE-S use 228 and 108 features respectively. In the next section, we study the performances of these models on the HDR and legacy SDR image quality databases.

IV. Results

We conducted an experiment where we evaluated the performance of several state-of-the-art NR IQA algorithms on the recently developed ESPL-LIVE HDR Database as well as on two legacy SDR image quality databases. The performances of the proposed algorithm models were evaluated by measuring the correlations between the predicted and the subjective scores (after non-linear regression), as well as the outlier ratios and root-mean-square-errors. The results were also analyzed to determine the statistical significance of the algorithm comparisons.

A. ESPL-LIVE HDR Image Quality Database

The ESPL-LIVE HDR Image Quality Database was recently created by crowdsourcing a large number of subjective eval-

TABLE II

SPEARMAN'S RANK ORDERED CORRELATION COEFFICIENT BETWEEN EACH FEATURE AND DMOS ACROSS 50 TRAIN-TEST (4:1) COMBINATIONS ON THE ESPL-LIVE HDR DATABASE FOR VARIOUS PROCESSING ALGORITHMS. TO ISOLATE THE EFFECT OF EACH SCALE AND COLOR CHANNEL, THIS TABLE SUMMARIZES THE RESULTS FOR THE COARSEST IMAGE SCALE AND USES THE L-COMPONENT ONLY. 'TMO', 'MEF', 'PP' REPRESENTS TONE-MAPPING, MULTI-EXPOSURE FUSION, AND POST-PROCESSING ALGORITHMS RESPECTIVELY. LOW CORRELATIONS BETWEEN EACH INDIVIDUAL FEATURE AND DMOS SHOW THAT THE FEATURES COMPLEMENT EACH OTHER. FEATURES $f_1 - f_{36}$ ARE USED IN HIGRADE-1 AND FEATURES $f_1 - f_{18}$, $f_{37} - f_{40}$ ARE USED IN HIGRADE-2 RESPECTIVELY. IN ORDER TO STUDY THE EFFECT OF USING ALL THE GRADIENT FEATURES, WE USED FEATURES $f_1 - f_{40}$ IN HIGRADE-3. FURTHER, IN ORDER TO STUDY THE IMPACT OF THE SPATIAL DOMAIN FEATURES SEPARATELY, HIGRADE-S WAS PROPOSED USING FEATURES $f_1 - f_{18}$

Domain	Feature Description	TMO	MEF	PP	Overall
Spatial	Shape and Scale parameters of the GGD fitted to the MSCN coefficients (9) $[f_1 - f_2]$	0.207	0.241	0.423	0.240
Spatial	Shape and Scale parameters of the GGD fitted to the log-derivative of the seven types of neighbors $[f_3 - f_{16}]$	0.333	0.489	0.490	0.424
Spatial	Two parameters extracted from the σ-field (11) $[f_{17} - f_{18}]$	0.543	0.300	0.418	0.404
Gradient	Shape and Scale parameters of the GGD fitted to the MSCN coefficients of gradient magnitude field $[f_{19} - f_{20}]$	0.217	0.202	0.409	0.247
Gradient	Shape and Scale parameters of the GGD fitted to the log-derivative of the seven types of neighbors of gradient magnitude field $[f_{21} - f_{34}]$	0.248	0.385	0.566	0.374
Gradient	Two parameters extracted from the σ-field of gradient magnitude field $[f_{35} - f_{36}]$	0.615	0.280	0.383	0.420
Gradient	Mean, standard deviation, skewness, and kurtosis of gradient structure tensor (Section III-C2) $[f_{37} - f_{40}]$	0.588	0.263	0.405	0.439

uations on a sizeable corpus of images processed using a variety of HDR processing algorithms. More details about the types of processing algorithms, the subjective testing framework, and the method of analyzing the raw scores may be found in [2]. The database contains 1,811 HDR-processed images created from 605 high quality source HDR scenes. The original multiple exposure stacks were captured using a Canon Rebel T5, a Nikon D2x, and a Nikon D5300 digital SLR cameras and displayed at a resolution of 960 × 540 (landscape) and 540 × 304 (portrait). Figure 7 shows some sample images from the new database.

Although legacy subjective IQA databases usually categorize images into distortion types (such as blur, JPEG compression, and color saturation), this database makes no such attempt. It is infeasible to superimpose artificial classification schemes onto realistic HDR images, since, depending on the scene and the type of processing algorithm considered, the image could be impaired by a complex interplay of multiple luminance, structural or chromatic artifacts that are difficult to categorize. Here we briefly describe the different methods used to create the images.

1) Images Generated by Tone Mapping Operators (TMO): The process of generating well-exposed SDR scenes involves estimating the scene radiance map (typically represented as a 32 bit floating point HDR map), followed by tone-mapping it to the displayable gamut of SDR displays (8 bit/color/pixel). Tone-mapping algorithms compute either a spatially varying transfer function or shrink image gradients to fit within the available dynamic range [53]. In our database, 747 tonemapped images were created by using four representative TMOs proposed by Larson *et al.* [38], Fattal*et al.* [48], Durand and Dorsey [54] and Reinhard *et al.* [42].

2) Images Generated by Multi-Exposure Fusion (MEF): The bracketed stack of images, after being downsampled to the display resolution, was first registered using a SIFT based image alignment method [55], then the aligned images were cropped so that every pixel was visible in every image of the stack, thus avoiding "black border" artifacts. The exposure

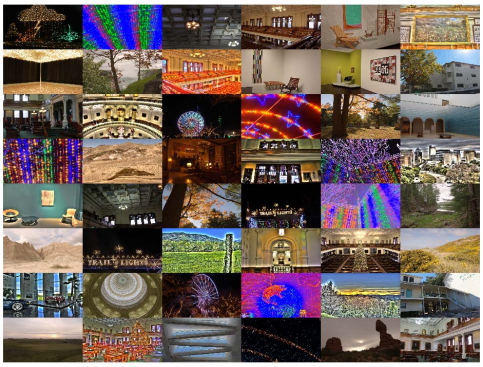


Fig. 7. Sample images from the ESPL-LIVE HDR Image Quality Database. The images include pictures taken both during day and night under diverse illumination conditions. Scenes containing both natural and man-made objects are included.

images were then blended using a MEF algorithm, which can broadly be expressed as [5]:

$$Y(i) = \sum_{k=1}^{K} W_k(i) X_k(i) \tag{22}$$

where K is the number of bracketed images, Y is the fused output image, and $X_k(i)$ and $W_k(i)$ indicate luminance or color either in the spatial domain or coefficients in a transform domain, and the weight at the i-th pixel in the k-th exposure image, respectively. W_k is a relative spatial weight on images captured at the different exposure levels, based on the perceptual information content. Different MEF algorithms differ in the ways that the weights are captured, but they all have an end goal of maintaining details both in underexposed and overexposed regions. These methods bypass the intermediate step of creating an HDR irradiance map, by instead creating an SDR image that can be directly displayed on standard displays. In the ESPL-LIVE HDR Image Database, 710 images were created via multi-exposure fusion (using local and global energy methods and the ones described in [37], [56], and [57]).

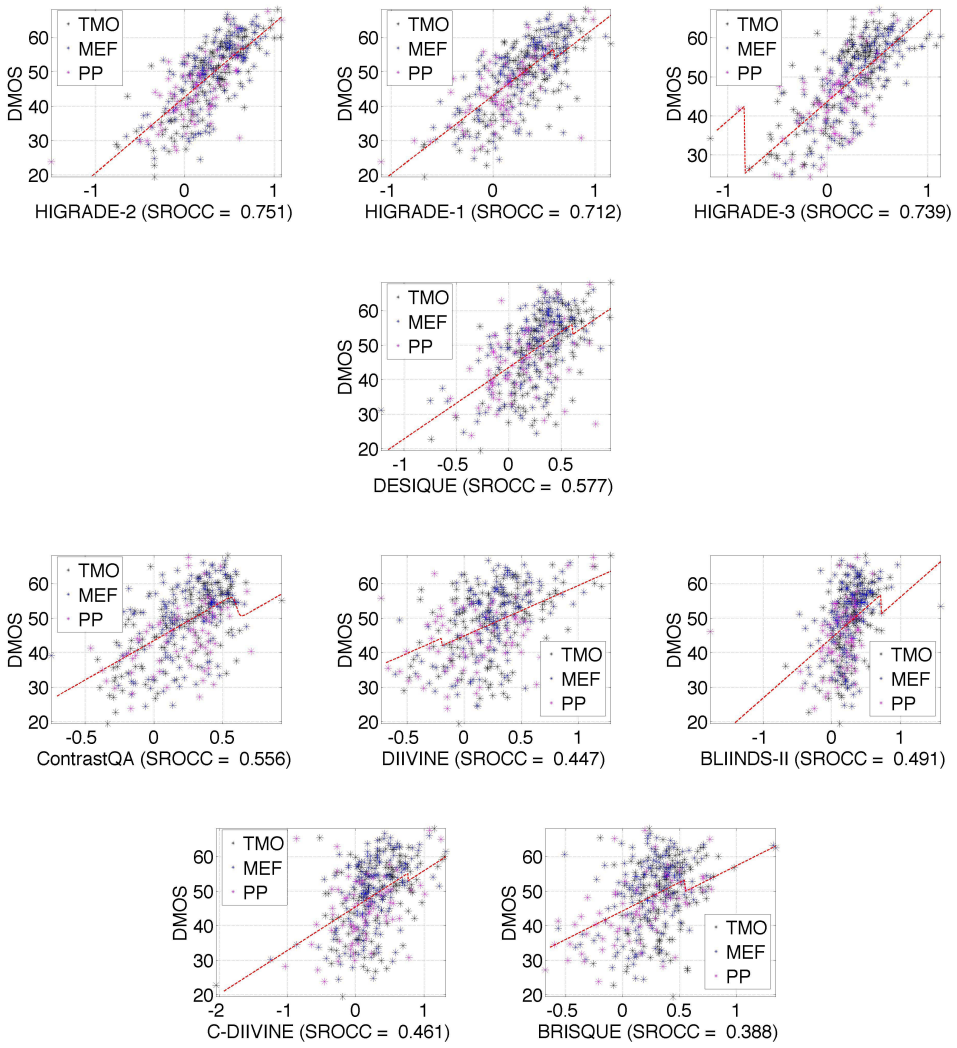


Fig. 8. Scatter plots between the predicted scores and the MOS scores on the ESPL-LIVE HDR Database for a few NR IQA algorithms. The red line indicates the logistic regression fit. The abbreviations "TMO", "MEF", and "PP" indicate Tone-Mapping, Multi-Exposure Fusion, and Post-Processing algorithms respectively.

3) Post Processed Images: Many HDR images created by professional and amateur photographers are post-processed in order to convey different 'feels' about a scene. This can drastically alter the final look of the image. We also included post-processed HDR images in the database for subjective evaluation, since these types of effects are not represented in any existing HDR quality database. In our implementation, we first created an irradiance map using Photomatix and tonemapped it using their default tone-mapping algorithm, followed by post-processing using two commonly used effects: "Surreal" and "Grunge", using different parameter settings on color saturation, color temperature and detail contrast preservation. A total of 354 images in the database were created using this method.

We obtained subjective evaluations from 5,462 observers using Amazon's online crowdsourcing platform, Mechanical Turk. A total of 327,720 human subjective ratings were gathered, and each image was rated by an average by 110 subjects. Although the subjects viewed the images under widely varying conditions, a high degree of consistency was observed among the subjects with respect to their judgments of perceptual quality. For each image, the ratings were divided into two disjoint equal sized subsets and the MOS values computed using each of them. This procedure was repeated over 25 random splits. The median Pearson Linear Correlation Coefficient (PLCC) between the MOS of the two sets was found to be **0.9721**.[1] In a separate control experiment, we also collected 'gold standard' MOS values on 5 images via a subjective test conducted in a traditional laboratory setting. The PLCC between these scores and the crowdsourced MOS on the same images was found to be **0.9465**.

The high values of these correlations indicate exceptionally good consistency between the scores obtained via crowdsourcing. The database is publicly available at http://signal.ece.

[1] All of the correlation values between IQA algorithm scores and/or human ground truth values were computed following non-linear logistic regression as outlined in [9]

TABLE III

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 ON THE ESPL-LIVE HDR DATABASE ALONG WITH ALGORITHM COMPUTATION TIME (ON A MACINTOSH LAPTOP HAVING 8 GB RAM, 2.9 GHz CLOCK, INTEL CORE i7 CPU). THE TABLE WAS SORTED IN DESCENDING ORDER OF SROCC OF THE 'OVERALL CATEGORY'. THE BOLD VALUES INDICATE THE BEST PERFORMING ALGORITHM. THE PROPOSED HIGRADE ALGORITHMS ARE SHOWN IN RED. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	IQA	Tone Mapping		Multi-Exposure Fusion		Post Processing		Overall		Time
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	(Secds.)
1	HIGRADE-2	**0.760**	**0.794**	**0.723**	**0.709**	0.582	0.606	**0.730** (0.675, 0.789)	**0.728**(0.667, 0.784)	3.043
2	HIGRADE-1	0.728	0.764	0.711	0.705	0.616	0.643	0.719 (0.671, 0.766)	0.718(0.652, 0.776)	12.613
3	HIGRADE-3	0.713	0.753	0.707	0.699	**0.627**	**0.663**	0.714 (0.650, 0.789)	0.719(0.645, 0.791)	42.083
4	HIGRADE-S	0.718	0.743	0.698	0.684	0.582	0.619	0.700 (0.617, 0.751)	0.697(0.608, 0.766)	1.098
5	HIGRADE-2 (L)	0.723	0.750	0.687	0.676	0.516	0.560	0.690 (0.602, 0.740)	0.692(0.592, 0.740)	1.029
6	HIGRADE-3 (L)	0.678	0.713	0.642	0.643	0.580	0.606	0.668 (0.600, 0.748)	0.667(0.582, 0.741)	12.400
7	HIGRADE-1 (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)	0.426
8	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)	0.503
9	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)	20.498
10	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)	35.662
11	ContrastQA	0.683	0.717	0.432	0.433	0.401	0.413	0.506 (0.405, 0.631)	0.521(0.402, 0.632)	189.303
12	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)	8.759
13	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)	0.797
14	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)	5.473
15	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)	**0.341**

TABLE IV

ROOT-MEAN-SQUARE ERROR (RMSE), REDUCED $\tilde{\chi}^2$ STATISTIC BETWEEN THE ALGORITHM SCORES AND THE MOS FOR VARIOUS NR-IQA ALGORITHMS (AFTER LOGISTIC FUNCTION FITTING) AND OUTLIER RATIO (EXPRESSED IN PERCENTAGE) FOR EACH DISTORTION CATEGORY ON THE ESPL-LIVE HDR DATABASE. THE BOLD VALUES INDICATE THE BEST PERFORMING ALGORITHM FOR THAT CATEGORY. THE PROPOSED HIGRADE ALGORITHMS ARE SHOWN IN RED. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	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	HIGRADE-2	**6.416**	4.207	**0.000**	6.830	21.907	**0.000**	7.148	3.458	**0.000**	**6.992**	11.197	0.276
2	HIGRADE-1	6.711	9.908	**0.000**	6.884	21.155	**0.000**	6.884	2.376	**0.000**	7.033	13.918	**0.275**
3	HIGRADE-3	6.932	3.895	0.621	**6.791**	9.144	**0.000**	**6.798**	13.531	**0.000**	7.078	11.930	0.277
4	HIGRADE-S	6.916	4.715	**0.000**	6.970	15.356	0.627	7.186	11.933	**0.000**	7.338	13.062	0.552
5	HIGRADE-2 (L)	6.945	3.869	**0.000**	7.128	5.008	0.669	7.422	7.442	**0.000**	7.336	12.920	0.551
6	HIGRADE-3 (L)	7.285	11.940	0.637	7.305	10.488	**0.000**	7.180	2.866	**0.000**	7.548	15.929	0.552
7	HIGRADE-1 (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
8	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
9	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
10	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
11	ContrastQA	7.248	8.681	**0.000**	8.562	21.640	0.755	8.183	6.501	**0.000**	8.556	33.433	0.829
12	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
13	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
14	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
15	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

utexas.edu/~debarati/ESPL_LIVE_HDR_Database/index.html.

B. Regressions Conducted

Once the features described in Section III were extracted, a mapping was obtained from the feature space to the DMOS scores using a regression method, which provides predictions of perceptual quality. We used a support vector regressor (SVR), implemented using the LibSVM [58] package to implement an ϵ-SVR using the radial basis function kernel, where the kernel parameter was by default the inverse of the number of features.

C. Experiments on the ESPL-LIVE HDR Database

We randomly split the data into disjoint training and testing sets at a 4:1 ratio and the split was randomized over 100 trials. Care was taken to ensure that the same source scene did not appear in both the training and the testing sets 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 were computed at every iteration. The median value across iterations of the correlations was reported. Fig. 8 shows the scatter plots between the predicted scores and the MOS scores on the ESPL-LIVE HDR Database for a few NR IQA algorithms.

We discovered that there is significant room for improvement among contemporary top-performing NR IQA models with regards to their HDR quality prediction power. Further, as compared to HIGRADE-S, the HIGRADE models that utilize gradient features delivered higher correlations against human judgments than otherwise (HIGRADE-1, HIGRADE-2,

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 LEGACY LIVE IQA DATABASE. PERFORMANCES OF SOME FR-IQA ALGORITHMS (SHOWN IN ITALICS) HAVE BEEN INCLUDED FOR COMPARISON. '−' INDICATES THAT THE ORIGINAL PAPER DID NOT REPORT THESE VALUES. THE TABLE HAS BEEN SORTED IN THE DESCENDING ORDER OF SROCC OF THE 'OVERALL CATEGORY' FOR NR IQA ALGORITHMS. 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 MODELS. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	IQA	JP2K		JPEG		Gaussian Noise		Blur		Fast Fading		Overall	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
1	Tang's IQA	0.950	-	0.970	-	0.980	-	0.940	-	**0.925**	-	**0.965**(-,-)	-(-,-)
2	CNN	**0.952**	**0.953**	**0.977**	**0.981**	0.978	0.984	**0.962**	**0.953**	0.908	**0.933**	0.956(-,-)	**0.953**(-,-)
3	DLIQA-R	0.928	0.947	0.912	0.940	0.968	0.955	0.946	0.944	0.861	0.890	0.927(-,-)	0.930(-,-)
4	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)
5	HIGRADE-1	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)
6	HIGRADE-2	0.904	0.910	0.867	0.902	0.982	0.990	0.920	0.930	0.841	0.863	0.904 (0.810, 0.943)	0.903(0.819, 0.943)
7	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)
8	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)
9	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)
10	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)
11	HIGRADE-1 (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)
12	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)
13	HIGRADE-2 (L)	0.822	0.843	0.818	0.855	0.941	0.956	0.897	0.907	0.729	0.737	0.837 (0.600, 0.920)	0.840(0.621, 0.911)
14	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)
15	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)
16	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)
17	*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
18	*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
19	*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

and HIGRADE-3). However, using all of the gradient features, as in HIGRADE-3, did not produce an improvement in performance relative to HIGRADE-1 or HIGRADE-2. This may be explained by the redundancy between their feature sets. The results are summarized in Table III.

The algorithm compute times were also measured on a 960×540 image using a Macintosh laptop having 8 GB RAM, 2.9 GHz clock, and an Intel Core i7 CPU. Fig. 9 shows a scatter plot of SROCC vs. execution time for the considered NR-IQA algorithms. The plot makes quite clear the high efficiency achieved by the HIGRADE models.

Table IV shows the root-mean-squared-errors (RMSE), reduced $\tilde{\chi}^2$ statistic between scores predicted by the algorithms and MOS (after logistic regression) and the outlier ratios (expressed in percentage). The top performing algorithms yielded lower values of RMSE and outlier ratio. HIGRADE-1 (L), HIGRADE-2 (L), and HIGRADE-3 (L) only used features extracted on the L-component after conversion to LAB color space.

We believe that this is because of the fact that adding more number of features may not always result in better performance of the algorithms because the features might be capturing overlapping or redundant information from the images.

Fig. 10 shows box plots of the distributions of the 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 enabled us to study the robustness of the algorithms against variations of the choice of training set. The HIGRADE model predictions had smaller variation of correlation against the human subjective scores.

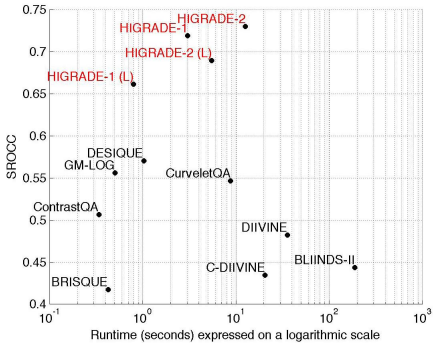


Fig. 9. Scatter plot of SROCC of selected NR-IQA algorithms with images in the ESPL-LIVE HDR Image Quality Database vs. runtime. Red indicates the proposed algorithms.

To analyze the degree of variation of SROCC between the scores predicted by the algorithm and MOS, the percentage of train/test splits was varied from in 10% increments, 90% of the content used for training and the remaining 10% used for testing, to 10% of the content used for training to 90% used for testing. As shown in Fig. 11, the knee of the curve occurs roughly at 60:40 train:test splits for the proposed HIGRADE algorithms. This shows that the results are not affected by overfitting or underfitting to the training data.

D. Experiments on Other Databases

We also tested the performance of the proposed HIGRADE algorithms on the legacy LIVE IQA database [9], TID2013 database [10], LIVE Multiply Distorted Image Database [59] and on the large new crowd-sourced LIVE In the Wild

TABLE VI

SPEARMAN'S RANK ORDERED CORRELATION COEFFICIENT (SROCC) AND PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC) BETWEEN ALGORITHM SCORES AND DMOS FOR VARIOUS NR IQA ALGORITHMS ON A SUBSET OF THE TID 2013 DATABASE AFTER TRAINING THE ALGORITHMS ON THE LIVE IQA DATABASE. '−' INDICATES THAT THE ORIGINAL PAPER DID NOT REPORT THESE VALUES. THE TABLE WAS SORTED IN DESCENDING ORDER OF SROCC OF THE 'OVERALL' CATEGORY. THE BOLD VALUES INDICATE THE BEST PERFORMING ALGORITHM. RED INDICATES THE PROPOSED MODELS. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	IQA	JPEG			JP2K			Gaussian Noise			Blur			Overall		
		SROCC	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
1	BLIINDS-II	0.889	-	-	0.915	-	-	0.696	-	-	0.857	-	-	**0.854**	-	-
2	Tang's IQA	**0.900**	-	-	**0.950**	-	-	0.830	-	-	0.900	-	-	0.83	-	-
3	HIGRADE-1	0.864	**0.913**	**0.176**	0.909	0.926	0.185	**0.874**	**0.884**	**0.095**	**0.927**	**0.922**	**0.139**	0.839	**0.883**	**0.188**
4	HIGRADE-2	0.840	0.886	0.201	0.914	**0.938**	**0.169**	0.851	0.857	0.105	0.889	0.883	0.168	0.823	0.868	0.199
5	HIGRADE-1 (L)	0.777	0.805	0.271	0.871	0.898	0.215	0.743	0.751	0.134	0.817	0.813	0.208	0.789	0.806	0.237
6	BRISQUE	0.835	0.867	0.215	0.870	0.890	0.252	0.696	0.699	0.145	0.816	0.805	0.212	0.779	0.819	0.230
7	HIGRADE-2 (L)	0.787	0.845	0.231	0.772	0.800	0.293	0.701	0.697	0.146	0.875	0.880	0.170	0.767	0.793	0.244
8	DESIQUE	0.762	0.831	0.240	0.812	0.805	0.290	0.752	0.770	0.130	0.714	0.757	0.246	0.727	0.749	0.265
9	DIQI	-	-	-	-	-	-	-	-	-	-	-	-	0.673	-	-

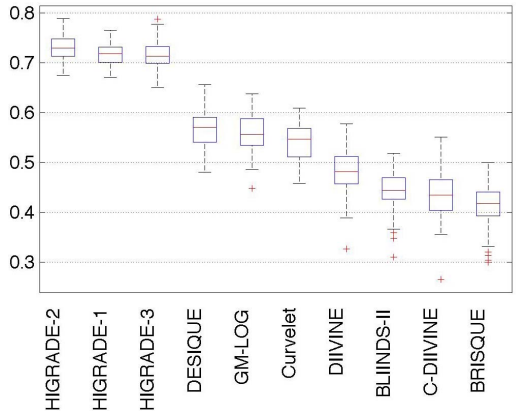


Fig. 10. Box plot of SROCC of selected learning based NR IQA algorithms on ESPL-LIVE HDR Image Database using 4:1 train-test splits over 100 trials. For each box, median is the central box, edges of the box represent 25th and 75th percentiles, whiskers span the most extreme non-outlier data points, and outliers are plotted individually.

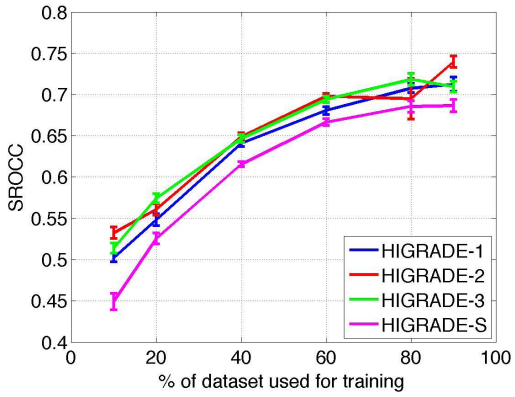


Fig. 11. Median SROCC between predicted scores and subjective DMOS scores for HIGRADE-1, HIGRADE-2, HIGRADE-3, and HIGRADE-S (and the associated 95% confidence intervals) as a function of the percentage of the content used to train on the ESPL-LIVE HDR Image Database over 25 trials. Note that the ordinate is plotted over [0.4,0.8].

Challenge database [60]. None of these databases contain HDR processed images. In each of these experiments, a mapping was obtained from the feature space to MOS/DMOS scores

TABLE VII

SPEARMAN'S RANK-ORDER CORRELATION COEFFICIENT (SROCC), PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC), AND ROOT MEAN SQUARE ERROR (RMSE) BETWEEN VARIOUS IQA ALGORITHM SCORES AND THE MOS ACROSS 100 TRAIN-TEST (4:1) COMBINATIONS ON THE LIVE MULTIPLY DISTORTED IMAGE DATABASE. THE TABLE IS SORTED IN DESCENDING ORDER OF SROCC FOR PART 2 OF THE DATABASE. RED INDICATES THE PROPOSED MODELS. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	IQA	Part 1			Part 2		
		SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
1	BRISQUE	0.873	0.909	8.578	**0.881**	**0.906**	**8.132**
2	HIGRADE-1	0.863	0.895	8.956	0.879	0.900	8.326
3	HIGRADE-1 (L)	0.892	0.918	7.654	0.871	0.899	8.407
4	C-DIIVINE	**0.901**	**0.927**	**7.236**	0.870	0.897	8.686
5	DESIQUE	0.894	0.926	7.670	0.850	0.881	8.935
6	CurveletQA	0.782	0.846	10.936	0.847	0.879	9.325
7	HIGRADE-2	0.754	0.814	11.355	0.834	0.848	9.999
8	GM-LOG	0.813	0.860	10.173	0.831	0.865	9.478
9	HIGRADE-2 (L)	0.812	0.871	9.891	0.805	0.848	10.518
10	DIIVINE	0.795	0.854	10.432	0.790	0.812	11.467
11	ContrastQA	0.248	0.219	18.397	0.200	0.176	18.324

using a support vector regressor (SVR) to implement ϵ-SVR with the radial basis function kernel. The same protocol was followed of splitting the data into disjoint training and testing sets at 4:1 ratio, randomized over a number of trials.

Table V shows the performance of the HIGRADE algorithms on the LIVE database [9]. The high degrees of correlation obtained against the subjective data shows that the HIGRADE methods are also able to capture typical artifacts arising in SDR images. Several blind IQA models that use sophisticated deep learners are included in Table V, [34]–[36]. These models can achieve the outer bounds of performance on this legacy database.

In order to explore the database independence of HIGRADE, it was trained on the legacy LIVE IQA database of natural images, then evaluated on the TID 2013 database [10]. From among the distortions present in the TID2008 database only the ones corresponding to those in the LIVE database were selected: JPEG2000, JPEG, gaussian noise, and blur. Table VI shows the results obtained on those types of artifacts.

TABLE IX

MEDIAN SPEARMAN'S RANK ORDERED CORRELATON COEFFICIENT (SROCC) AND PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC) BETWEEN ALGORITHM SCORES AND MOS ACROSS 50 TRAIN-TEST COMBINATIONS ON THE ESPL-LIVE HDR DATABASE WHEN HIGRADE-1 FEATURES FROM EACH COMPONENT OF EACH OF THREE DIFFERENT COLOR SPACES WERE INDEPENDENTLY USED TO TRAIN AN SVR, WAS WELL AS USING ALL THREE CHANNELS OF EACH COLOR SPACE

Color Space	IQA	Tone Mapping		Multi-Exposure Fusion		Post Processing		Overall	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
LAB space	HIGRADE-1 (L)	0.675	0.712	0.644	0.636	0.554	0.591	0.664	0.660
	HIGRADE-1 (A)	0.389	0.466	0.573	0.575	0.541	0.570	0.509	0.529
	HIGRADE-1 (B)	0.457	0.519	0.589	0.596	0.478	0.497	0.533	0.542
	All	0.730	0.762	0.710	0.698	0.617	0.651	**0.719**	**0.719**
YUV space	HIGRADE-1 (Y)	0.711	0.743	0.609	0.615	0.497	0.545	0.654	0.657
	HIGRADE-1 (U)	0.608	0.638	0.615	0.611	0.524	0.560	0.620	0.614
	HIGRADE-1 (V)	0.689	0.728	0.567	0.582	0.473	0.520	0.624	0.626
	All	0.762	0.788	0.689	0.685	0.541	0.576	**0.728**	**0.730**
LMS space	HIGRADE-1 (L)	0.665	0.696	0.580	0.581	0.479	0.512	0.616	0.615
	HIGRADE-1 (M)	0.681	0.717	0.596	0.596	0.536	0.583	0.644	0.651
	HIGRADE-1 (S)	0.612	0.647	0.622	0.628	0.519	0.548	0.610	0.618
	All	0.718	0.752	0.643	0.655	0.553	0.581	**0.686**	**0.688**

TABLE X

MEDIAN SPEARMAN'S RANK ORDERED CORRELATON COEFFICIENT (SROCC) AND PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC) BETWEEN ALGORITHM SCORES AND MOS ACROSS 50 TRAIN-TEST COMBINATIONS ON THE ESPL-LIVE HDR DATABASE WHEN HIGRADE-2 FEATURES FROM EACH COMPONENT OF EACH COLOR SPACE WERE INDEPENDENTLY USED TO TRAIN AN SVR, AS WELL AS WHEN USING ALL THREE CHANNELS OF EACH COLOR SPACE

Color Space	IQA	Tone Mapping		Multi-Exposure Fusion		Post Processing		Overall	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
LAB space	HIGRADE-2 (L)	0.726	0.752	0.687	0.678	0.517	0.567	0.696	0.694
	HIGRADE-2 (A)	0.387	0.457	0.530	0.550	0.513	0.554	0.481	0.508
	HIGRADE-2 (B)	0.442	0.498	0.576	0.573	0.480	0.492	0.525	0.534
	All	0.764	0.794	0.717	0.702	0.590	0.609	**0.731**	**0.728**
YUV space	HIGRADE-2 (Y)	0.734	0.773	0.619	0.629	0.501	0.545	0.675	0.680
	HIGRADE-2 (U)	0.701	0.733	0.662	0.647	0.506	0.546	0.679	0.669
	HIGRADE-2 (V)	0.690	0.740	0.556	0.585	0.446	0.482	0.618	0.628
	All	0.755	0.792	0.680	0.689	0.533	0.581	**0.726**	**0.728**
LMS space	HIGRADE-2 (L)	0.701	0.727	0.612	0.621	0.498	0.538	0.647	0.649
	HIGRADE-2 (M)	0.694	0.731	0.612	0.626	0.500	0.540	0.648	0.650
	HIGRADE-2 (S)	0.678	0.716	0.622	0.632	0.461	0.496	0.636	0.637
	All	0.739	0.771	0.658	0.666	0.529	0.576	**0.695**	**0.696**

TABLE XI

RESULTS OF THE F-TEST PERFORMED ON THE RESIDUALS BETWEEN MODEL PREDICTIONS AND MOS SCORES ON THE 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" PROCESSING ALGORITHMS. "1"("0") INDICATES THAT THE PERFORMANCE OF THE ROW IQA ALGORITHM IS BETTER(WORSE) TO THAT OF THE COLUMN IQA ALGORITHM. - INDICATES THAT THE STATISTICAL PERFORMANCE OF THE ROW IQA IS EQUIVALENT TO THAT OF THE COLUMN IQA. THE MATRIX IS SYMMETRIC. THE PROPOSED ALGORITHMS HAVE BEEN SHOWN IN RED

	HIGRADE-2	HIGRADE-1	DESIQUE	GM-LOG	Curvelet	ContrastQA	DIIVINE	BLIINDS-II	C-DIIVINE	BRISQUE
HIGRADE-2	- - - -	- - - -	1 - - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1
HIGRADE-1	- - - -	- - - -	1 - - 1	1 1 - 1	1 1 - 1	- 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1	1 1 - 1
DESIQUE	0 - - 0	0 - - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	1 - - 1
GM-LOG	0 0 - 0	0 0 - 0	- - - -	- - - -	- - - -	0 - - -	- - - -	- - - -	- - - -	- - - -
Curvelet	0 0 - 0	0 0 - 0	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	1 - - -
ContrastQA	0 0 - 0	- 0 - 0	- - - -	1 - - -	- - - -	- - - -	1 - - -	1 - - -	1 - - -	1 - - -
DIIVINE	0 0 - 0	0 0 - 0	- - - -	- - - -	- - - -	0 - - -	- - - -	- - - -	- - - -	- - - -
BLIINDS-II	0 0 - 0	0 0 - 0	- - - -	- - - -	- - - -	0 - - -	- - - -	- - - -	- - - -	- - - -
C-DIIVINE	0 0 - 0	0 0 - 0	- - - -	- - - -	- - - -	0 - - -	- - - -	- - - -	- - - -	- - - -
BRISQUE	0 0 - 0	0 0 - 0	0 - - 0	- - - -	0 - - -	0 - - -	- - - -	- - - -	- - - -	- - - -

For the sake of comparison, results obtained under the same train-test setup from some other well-known NR IQA models were also included, including the Deep learning based Image Quality Index (DIQI) in [33]. The correlation figures did drop somewhat relative to the results obtained when training and testing on disjoint sets of the same database, which is to be expected given the diversity of distortion ranges across databases, but the agreement with human quality judgments remained very strong.

The LIVE Multiply Distorted Image database contains images distorted by mixtures of blur, noise, and JPEG compression artifacts and offers a more realistic simulation of the images impaired by distortions arising from multiple sources. 'Part 1' of the database contains images distorted

TABLE VIII

MEDIAN SPEARMAN'S RANK-ORDER CORRELATION COEFFICIENT (SROCC), PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC) AND OUTLIER RATIO (OR) BETWEEN VARIOUS IQA ALGORITHM SCORES AND MOS ACROSS 50 TRAIN-TEST (4:1) COMBINATIONS ON THE LIVE CHALLENGE DATABASE. BOLD VALUES INDICATE THE BEST PERFORMING ALGORITHM. THE TABLE IS SORTED IN DESCENDING ORDER OF SROCC. RED INDICATES THE PROPOSED MODELS. (L) DENOTES ALGORITHM IS APPLIED ON LUMINANCE ONLY

	NR IQA algorithms	SROCC	PLCC	OR
1	FRIQUEE-ALL	**0.66 ± 0.04**	**0.70 ± 0.04**	**0.04**
2	HIGRADE-1	0.64 ± 0.01	0.65 ± 0.01	0.05
3	C-DIIVINE	0.63 ± 0.04	0.66 ± 0.04	0.05
4	HIGRADE-1 (L)	0.62 ± 0.01	0.64 ± 0.01	0.05
5	HIGRADE-2	0.61 ± 0.01	0.62 ± 0.01	0.06
6	HIGRADE-2 (L)	0.59 ± 0.01	0.60 ± 0.01	0.06
7	BRISQUE	0.58 ± 0.05	0.61 ± 0.06	0.07
8	DIIVINE	0.56 ± 0.05	0.59 ± 0.05	0.06
9	NIQE	0.42 ± 0.05	0.48 ± 0.05	-
10	BLIINDS-II	0.40 ± 0.05	0.45 ± 0.05	0.09
11	ContrastQA	0.27 ± 0.01	0.27 ± 0.01	0.08

by blur, followed by JPEG compression, whereas 'Part 2' contains images that are distorted by blur, followed by additive Gaussian white noise. The HIGRADE models again achieved good correlation against MOS on these type of artifacts. These results are summarized in Table VII.

We tested the performance of the HIGRADE algorithms on the challenging LIVE In the Wild Challenge Database [60] containing 1,162 images obtained under a wide variety of photographic and lighting conditions, captured using a variety of imaging devices, and afflicted by highly diverse and complex authentic distortions. The results are shown in Table VIII.[2] Again, the performance of the HIGRADE models was highly competitive with the current state-of-the-art.

In addition to natural images, the proposed HIGRADE models have also been found to show high degrees of correlation with human subjective scores on the ESPL Synthetic Image Database [61] composed to pristine and distorted computer graphics images obtained from video games and animation movies. The details may be found at [62].

E. Experiments on Different Color Spaces

In order to see the impact of choosing different color spaces on the performance of the proposed algorithms, HIGRADE-1 and HIGRADE-2 features were also extracted from the same images represented in the YUV and LMS color spaces. For the images in the ESPL-LIVE HDR Image Quality database, the correlations between the HIGRADE model predictions against MOS are summarized in Tables IX and X. In the LAB color space, using only the L-component resulted in a higher degree of correlation with human subjective scores as compared to only extracting features from the A or B chromatic channels. Interestingly, when testing on images represented in the YUV and LMS color spaces, extracting all the features from any of the three channels in isolation yielded almost equal degrees of correlation between the predicted scores and MOS. Unsurprisingly, the best performance on each color space was achieved when all three channels were used. Overall, using the LAB color space produced the best prediction results for HIGRADE-2 and YUV color space for HIGRADE-1.

[2]As stated in [43], outlier ratio on the predicted scores is not computed for NIQE since it is not trained on MOS values.

F. Determination of Statistical Significance

In order to study the statistical significance of the correlation results on the ESPL-LIVE HDR Image Quality database, ten representative NR IQA algorithms were selected. The statistical significance tests were carried out over multiple training-test splits, using random 4:1 train-test splits of the database each time, and similar results were obtained over all trials. Table XI outlines the results obtained for one such representative trial. To determine whether results delivered by the IQA algorithms were significantly different from each other, the F-statistic, as in [9], was used to determine the degree of statistical significance between the variances of the residuals after a non-linear logistic mapping between the two IQA algorithms, at the 95% confidence interval. Overall, the proposed HIGRADE algorithms were found to be statistically better to the other NR IQA algorithms.

V. CONCLUSION

We devised a new gradient-based NSS framework and proposed two algorithms that predict the perceptual quality of tonemapped/fused/post-processed HDR images. We also studied the relevance of various color spaces of the input HDR maps when perceiving their quality. We validated the new model(s) on a new, very large crowdsourced study of HDR images that deployed 5,462 unique participants who reported 327,720 image evaluations. The proposed HIGRADE algorithms were found to be the top performing predictors of human perceptual judgments of visual HDR artifacts, among the 12 NR-IQA models that were tested. The HIGRADE features were also shown to be effective at assessing the artifacts arising in SDR images, as corroborated by the experimental results on the legacy image quality databases [9], [10], [59], [60]. In the future, we believe that the HIGRADE models or features may prove to be quite useful for the perceptual optimization [63] of HDR processed pictures.

REFERENCES

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

[2] D. Kundu, D. Ghadiyaram, B. L. Evans, and A. C. Bovik, "Large-scale crowdsourced study for tone mapped HDR pictures," *IEEE Trans. Image Process.*, to be published. [Online]. Available: http://users.ece.utexas.edu/~bevans/papers/2017/crowdsourced/index.html

[3] 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.

[4] H. Z. Nafchi, A. Shahkolaei, R. F. 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.

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

[6] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[7] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 167–192, Jan. 2008.

[8] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[9] 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.

[10] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.

[11] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[12] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Proc. Int. Workshop Quality Multimedia Exper. (QoMEX)*, Jul. 2009, pp. 87–91.

[13] R. Hassen, Z. Wang, and M. M. A. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.

[14] S. A. Golestaneh and D. M. Chandler, "No-reference quality assessment of JPEG images via a quality relevance map," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 155–158, Feb. 2014.

[15] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Sep. 2002, pp. I-477–I-480.

[16] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[17] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," in *Proc. NIPS*, 1993, pp. 551–558.

[18] 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.

[19] 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, Mar. 2013.

[20] Y. Zhang and D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," *J. Electron. Imag.*, vol. 22, no. 4, p. 043025, 2013.

[21] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[22] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.

[23] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[24] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[25] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Process., Image Commun.*, vol. 29, no. 7, pp. 725–747, Aug. 2014.

[26] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[27] L. Liu, H. Dong, H. Huang, and A. C. Bovik, "No-reference image quality assessment in curvelet domain," *Signal Process., Image Commun.*, vol. 29, no. 4, pp. 494–505, Apr. 2014.

[28] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793–799, May 2011.

[29] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Amer.*, vol. 24, no. 12, pp. B42–B51, Dec. 2007.

[30] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. CVPR*, Jun. 2012, pp. 1098–1105.

[31] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 75–78, Feb. 2012.

[32] Y. Lu, F. Xie, T. Liu, Z. Jiang, and D. Tao, "No reference quality assessment for multiply-distorted images based on an improved bag-of-words model," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1811–1815, Oct. 2015.

[33] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Deep learning network for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Mar. 2014, pp. 511–515.

[34] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.

[35] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Mar. 2014, pp. 1733–1740.

[36] H. Tang, N. Joshi, and A. Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2877–2884.

[37] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurogr. Conf.*, Munich, Germany. 2009.

[38] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graphics*, vol. 3, no. 4, pp. 291–306, Oct. 1997.

[39] *Photomatix*, accessed on May 2016. [Online]. Available: http://www.hdrsoft.com/

[40] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *Proc. ACM SIGGRAPH Papers*, 2008, pp. 67:1–67:10. [Online]. Available: http://doi.acm.org/10.1145/1399504.1360666

[41] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[42] 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.

[43] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, p. 32, 2016.

[44] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[45] 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.

[46] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Image Commun.*, vol. 40, pp. 1–15, Jan. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.image.2015.10.005

[47] J. Bigun, "Optimal orientation detection of linear symmetry," in *Proc. IEEE Int. Conf. Comp. Vis.*, Great Britain, U.K., Jun. 1987, pp. 433–438.

[48] 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.

[49] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[50] A. Baghaie and Z. Yu, "Structure tensor based image interpolation method," *AEU Int. J. Electron. Commun.*, vol. 69, no. 2, pp. 515–522, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1434841114003033

[51] J. Weickert, *Anisotropic Diffusion in Image Processing* (European Centre for Minority Issues). Stuttgart, Germany: Teubner Verlag, 1998. [Online]. Available: http://opac.inria.fr/record=b1105544

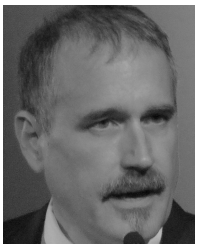
[52] G. Kuhne, J. Weickert, O. Schuster, and S. Richter, "A tensor-driven active contour model for moving object segmentation," in *Proc. Int. Conf. Image Process.*, vol. 2. Oct. 2001, pp. 73–76.

[53] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. New York, NY, USA: Springer-Verlag, 2010.

[54] 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
[55] F. Banterle. *HDR Toolbox for MATLAB*, accessed on May 2016. [Online]. Available: https://github.com/banterle/HDR_Toolbox
[56] S. Paul, I. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain," *J. Circuits, Syst., Comput.*, vol. 25, no. 10, pp. 1650123-1–1650123-18, 2016. [Online]. Available: http://ee.ucr.edu/~supaul/Webpage_files/JCSC2016.pdf, doi: 10.1142/S0218126616501231.
[57] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," *J. Virtual Reality Broadcast.*, vol. 10, no. 2, Mar. 2013. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:0009-6-36506
[58] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm
[59] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1693–1697.
[60] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jun. 2016. [Online]. Available: http://dx.doi.org/10.1109/TIP.2015.2500021
[61] D. Kundu and B. L. Evans, "Full-reference visual quality assessment for synthetic images: A subjective study," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2374–2378.
[62] D. Kundu, "Subjective and objective quality evaluation of synthetic and high dynamic range images," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Texas Austin, Austin, TX, USA, May 2016. [Online]. Available: http://users.ece.utexas.edu/ bevans/students/phd/debarati_kundu/
[63] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 857–872, Jun. 2008.

Debarati Kundu received the B.E. degree in electronics and telecommunications engineering from Jadavpur University, Kolkata, India, in 2010, and the M.Sc. and Ph.D. degrees from the Department of Electrical and Computer Engineering, The University of Texas at Austin, in 2012 and 2016, respectively. She is currently a Senior Engineer with Qualcomm Research Bangalore, India. Her research interests include image and video quality assessment, computer graphics, computer vision, machine learning, and prototyping of real-time systems. She was a recipient of the RGM Advisors Research Award for best poster at Graduate and Industry Networking, UT Austin 2016, the top 10% paper award at the IEEE International Conference on Image Processing 2015, and the Qualcomm Roberto Padovani fellowship 2014 awarded to the top 1% of the interns.

Deepti Ghadiyaram received the B.Tech. degree in computer science from the International Institute of Information Technology, Hyderabad, in 2009, and the M.S. degree from The University of Texas at Austin in 2013, where she is currently pursuing the Ph.D. degree with the Laboratory for Image and Video Engineering. Her research interests broadly include image and video processing, particularly perceptual image and video quality assessment, computer vision, and machine learning, and their applications to the aspects of information retrieval and enhancing end users quality of experience. She was a recipient of the Microelectronics and Computer Development Fellowship from 2013 to 2014 and a recipient of Graduate Student Fellowship offered to the top 1% of the students by the Department of Computer Science for the academic years from 2013 to 2016.



Alan C. Bovik (F'96) holds the Ernest J. Cockrell Endowed Chair in engineering with The University of Texas at Austin, where he is currently a Professor with the Department of Electrical and Computer Engineering and The Institute for Neurosciences and the Director of the Laboratory for Image and Video Engineering. His interests include image and video processing, digital television, computational vision, and modeling of biological visual perception. He has also authored *The Handbook of Image and Video Processing*, (Elsevier Academic Press, 2005), *Modern Image Quality Assessment* (Morgan & Claypool, 2006), *The Essential Guide to Image Processing* (Elsevier Academic Press, 2009), and *The Essential Guide to Video Processing* (Elsevier Academic Press, 2009). He has authored over 800 technical articles in these areas and holds several U.S. patents.

Dr. Bovik is a fellow of the Optical Society of America and SPIE, and a member of both the Television Academy (ATAS) and the National Academy of Television Arts and Sciences (NATAS). He is also a member of the Society of Motion Picture and Television Engineers and the Royal Society of Photography. He received a Televisions highest honor, a Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Academy of Television Arts and Sciences (The Television Academy) in 2015, for his work on the development of video quality prediction models, which are standard tools throughout the global cable, satellite, broadcast, and Internet television industries. He has also received a number of major awards from the IEEE Signal Processing Society, including the Society Award; the Technical Achievement Award; the Best Paper Award; the *Signal Processing Magazine* Best Paper Award; the Education Award; the Distinguished Lecturer Award; the Meritorious Service Award; the Sustained Impact Paper Award; and (Co-Author) the Young Author Best Paper Award. He also received the IEEE Circuits and Systems for Video Technology Best Paper Award. He received the Honorary Member Award of the Society for Imaging Science and Technology, the Society of Photo-Optical and Instrumentation Engineers (SPIE) Technology Achievement Award, and the IS&T/SPIE Imaging Scientist of the Year. He was also a recipient of the Joe J. King Professional Engineering Achievement Award and the Hocott Award for Distinguished Engineering Research, both from the Cockrell School of Engineering, The University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign–Urbana (2008), and the IEEE Third Millennium Medal.

He founded and served as the first General Chair of the IEEE International Conference on Image Processing, held in Austin, TX, in 1994. He was also the General Chair of the 2014 Texas Wireless Symposium, held in Austin in 2014. Among many professional activities, he has served on the IEEE Signal Processing Society Board of Governors; the Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002; an Overview Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, served on the Editorial Board of The Proceedings of the IEEE, and the Senior Editorial Board of the IEEE JOURNAL ON SPECIAL TOPICS IN SIGNAL PROCESSING.

He is a registered Professional Engineer in the State of Texas (License # 114706) and a frequent Consultant to legal, industrial, and academic institutions.

Brian L. Evans (F'09) received the B.S. degree in EECS from the Rose-Hulman Institute of Technology in 1987 and the M.S. and Ph.D. degrees in EE from the Georgia Institute of Technology in 1988 and 1993, respectively. From 1993 to 1996, he was a Post-Doctoral Researcher with the University of California at Berkeley, Berkeley. In 1996, he joined the Faculty at UT Austin. He holds the Engineering Foundation Professorship with UT Austin. Embedded Signal Processing Laboratory is hosted under University of Texas at Austin.

His research bridges digital signal processing theory and embedded real-time implementation. Applications include smart grid communications, cellular communications, and image/video acquisition/display.

He has authored over 240 refereed conference and journal papers, and graduated 27 Ph.D. and ten M.S. students. He received the Best Paper Award at the 2013 IEEE International Symposium on Power Line Communications and Its Applications, and Top 10% Paper Awards at the 2015 EEE International Conference on Image Processing and the 2012 IEEE Multimedia Signal Processing Workshop. He received three teaching awards at UT Austin, and a 1997 U.S. National Science Foundation CAREER Award.