Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality

Anush Krishna Moorthy and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—Our approach to blind image quality assessment (IQA) is based on the hypothesis that natural scenes possess certain statistical properties which are altered in the presence of distortion, rendering them *un-natural*; and that by characterizing this un-naturalness using scene statistics, one can identify the distortion afflicting the image and perform no-reference (NR) IQA. Based on this theory, we propose an (NR)/blind algorithm—the Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) index—that assesses the quality of a distorted image without need for a reference image. DIIVINE is based on a 2-stage framework involving distortion identification followed by distortion-specific quality assessment. DIIVINE is capable of assessing the quality of a distorted image across multiple distortion categories, as against most NR IQA algorithms that are distortion-specific in nature. DIIVINE is based on natural scene statistics which govern the behavior of natural images. In this paper, we detail the principles underlying DIIVINE, the statistical features extracted and their relevance to perception and thoroughly evaluate the algorithm on the popular LIVE IQA database. Further, we compare the performance of DIIVINE against leading full-reference (FR) IQA algorithms and demonstrate that DIIVINE is *statistically superior* to the often used measure of peak signal-to-noise ratio (PSNR) and *statistically equivalent* to the popular structural similarity index (SSIM). A software release of DIIVINE has been made available online: http://live.ece.utexas.edu/research/quality/DIIVINE_release.zip for public use and evaluation.**

***Index Terms*—Blind quality assessment, image quality, natural scene statistics, no-reference.**

I. Introduction

WE live in an increasingly visual digital world. Advances in technology have allowed for images and videos to be captured easily and efficiently, transmitted, stored, shared, and viewed over a range of devices, from pocket-size handheld ones to large HD screens. This progress coupled with the fact that humans are highly visual creatures necessitates that the reproduced 2-D world of images and videos be accurately rendered without deviating from what humans perceive as acceptable. The limits on bandwidth and the physical properties of devices used to capture these visual signals imply that loss of information and the introduction of extraneous artifacts are bound to occur. Researchers in visual quality assessment have endeavored to understand how the presence of these distortions affects the viewing experience. The ideal approach to measure the effect of distortions on the quality of viewing experience is to solicit opinions from a sufficiently large sample of the human populace. Averaging across these opinions produces a mean opinion score (MOS) which is considered to be the perceived quality of the stimulus. Such *subjective* assessment of visual quality is the best indicator of how distortions affect perceived quality; however, they are time-consuming, cumbersome, and impractical. Hence, one seeks to develop algorithms that produce quality estimates of these distorted visual stimuli with high correlation with MOS. Such *objective* image quality assessment (IQA) is the focus of this paper.

Objective quality assessment can be divided into three categories depending on the amount of information provided to the algorithm [1]. *Full-reference* (FR) algorithms are provided with the original undistorted visual stimulus along with the distorted stimulus whose quality is to be assessed. *Reduced-reference* (RR) approaches are those in which the algorithm is provided with the distorted stimulus and some additional information about the original stimulus, either by using a auxiliary channel or by incorporating some information in the distorted stimulus (such as a watermark). Finally, *no-reference* (NR)/blind approaches to quality assessment are those in which the algorithm is provided only with the distorted stimulus. In this paper, we propose a method for NR/blind IQA.

Even though NR QA is potentially the most useful goal, the difficulty of creating algorithms that accurately predict visual quality, especially without any information about the original image, has led to greater activity in the FR QA area [2]. These studies have yielded considerable insights into the perception of image distortions and image quality that may prove useful in creating algorithms that assess quality without need for a reference. RR QA remains attractive, not only as a solution to the QA problem on its own, but also as a stepping stone towards solving the NR QA problem [3], [4]. Indeed, our approach to NR QA finds inspiration from previously proposed models for FR QA [5], [6] as well as for RR QA [3], [4].

We have developed a computational theory for NR IQA based on the statistics of natural images[1] [7]–[10]. Natural un-distorted images possess certain statistical properties that hold across different image contents. For example, it is well known that the power spectrum of natural scenes fall-off as (approximately) $1/f^{\gamma}$, where f is frequency [9]. Natural scene statistic (NSS) models seek to capture those statistical

Manuscript received November 29, 2010; revised March 21, 2011; accepted April 08, 2011. Date of publication April 25, 2011; date of current version November 18, 2011. This work was supported in part by the National Science Foundation under grant CCF-0728748 and in part by Intel and Cisco corporation under the VAWN program. The associate editor coordinating the review of this manuscript and approving it for publication was Hsueh-Ming Hang.

The authors are with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: anushmoorthy@mail.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.2011.2147325

[1]By natural, we mean any image that can be obtained from a camera—these include pictures of man-made objects as well as forest/natural environments.

properties of natural scenes that hold across different contents. Our approach to NR IQA is based on the hypothesis that the presence of distortions in natural images alters the natural statistical properties of images, thereby rendering them (and consequently their statistics) *unnatural* [5]. The goal of an NR IQA algorithm based on NSS is to capture this "unnatural-ness" in the distorted image and relate it to perceived quality. In the past, such NSS-based QA algorithms have been successfully deployed for FR IQA [5], [6], for RR IQA [3], [4], and to a small extent, for NR IQA [11]. We explore such an NSS-based approach for NR IQA.

Our NR IQA model utilizes a 2-stage framework for blind IQA that we introduced in [12]. In this framework, scene statistics extracted from a distorted natural image are used to first explicitly classify the distorted image into one of n distortions (distortion identification—stage 1). Then, the same set of statistics are used to evaluate the distortion-specific quality (distortion-specific QA—stage 2) of the image. A combination of the two stages leads to a quality score for the image which, as we shall soon demonstrate, correlates quite well with human perception and is competitive with leading FR IQA algorithms. The proposed approach we call Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE). The name is appropriate as the algorithm resulting from the modeling framework succeeds at "divining" image quality without any reference information or the benefit of distortion models.

The DIIVINE approach to NR IQA is a full-fledged realization of the preliminary framework that we had proposed in [12]. In [12], we had primarily proposed the 2-stage framework and demonstrated simple implementations of the framework as examples. Apart from the fact that the DIIVINE approach performs much better than those realizations, the main contribution of this work is the series of statistical features that we extract, which go beyond the simple marginal descriptions that the previous primary realizations extracted.

Before proceeding, we state some salient aspects of DIIVINE. Present-day NR IQA algorithms are distortion-specific, i.e., the algorithm is capable of assessing the quality of images distorted by a particular distortion type. For example, the algorithm in [13] is for JPEG compressed images, that in [14] is for JPEG2000 compressed images, and that in [15] is for blur. DIIVINE, however, is not bound by the distortion-type affecting the image since we do not seek distortion-specific indicators of quality (such as edge strength at block boundaries) but provide a modular strategy that adapts itself to the distortion in question. Indeed, our framework is ostensibly distortion-agnostic.

Further, since we do not use distortion-specific models, DIIVINE can easily be extended to handle distortions beyond those considered here. Finally, by performing a thorough analysis of our algorithm, we demonstrate that DIIVINE is competitive with present-day NR and FR IQA algorithms *across* commonly encountered distortions. In fact, we shall demonstrate that DIIVINE is not only statistically superior to the full-reference peak signal-to-noise-ratio (PSNR) measure of quality, but is also statistically indistinguishable from a popular full-reference measure—the structural similarity index (SSIM) [16].

The rest of the paper is organized as follows. In Section II, we review previous work in NR IQA. In Section III, we describe the various scene statistics we extract from the image. In Section IV, we describe how these features are used for distortion identification, as well as for distortion-specific quality assessment. We evaluate the performance of the developed approach in Section V and conclude the paper in Section VI.

II. Previous Work

Most present-day NR IQA algorithms assume that the distorting medium is known—for example, compression, loss induced due to noisy channel, etc. Based on this assumption, distortions specific to the medium are modeled and quality is assessed. By far the most popular distorting medium is compression, which implies that blockiness and bluriness should be evaluated. In the following, we study blind QA algorithms that target three common distortion categories: JPEG compression, JPEG2000 compression, and blur. We also survey blind QA algorithms that operate holistically.

A. Distortion-Specific IQA Algorithms

1) JPEG IQA: The general approach to NR JPEG IQA is to measure edge strength at block boundaries and relate this strength and possibly some measure of image activity to perceived quality. JPEG NR IQA algorithms include those that use a hermite transform-based approach to model blurred edges [13], those that estimate first-order differences and activity in an image [17], those that utilize an importance map weighting of spatial blocking scores [18], those that use a threshold-based approach on computed gradients [19], and those that compute block strengths in the Fourier domain [20]. Each of these approaches measures a subset of blocking, blur, and activity and computes perceptual quality, either using a training set, or by combining features in an intelligent fashion.

2) JPEG2000 IQA: For JPEG2000, ringing artifacts in an image are generally modeled by measuring edge-spread using an edge-detection-based approach and this edge spread is related to quality [21]–[23]. Other approaches include those that compute simple features in the spatial domain [14], or those that utilize natural scene statistics [11]. In [11], the authors exploit the dependency between a wavelet coefficient and its neighbors, and the fact that the presence of distortion will alter these dependencies. The dependencies are captured using a threshold + offset approach, where the parameters are estimated using a training set.

3) Sharpness/Blur IQA: Blur IQA algorithms model edge spreads and relate these spreads to perceived quality, similar to the approach followed by NR JPEG2000 IQA algorithms. Edge strengths are quantified using a variety of techniques, including block kurtosis of DCT coefficients [15], iterative thresholding of a gradient image [24], and measuring the probability of blur detection [25] or model the just-noticeable-blur [26] in an image. Researchers have also explored the use of saliency models for NR blur IQA [27]. A noise-robust blur measure was also proposed in [28] that utilizes a gradient-based approach coupled with the singular value decomposition.

It should be clear to the reader that each of these distortion-specific NR IQA algorithms attempt to model indicators of quality for the distortion in question, and hence are unsuitable for use in a general-purpose (distortion-agnostic) scenario.

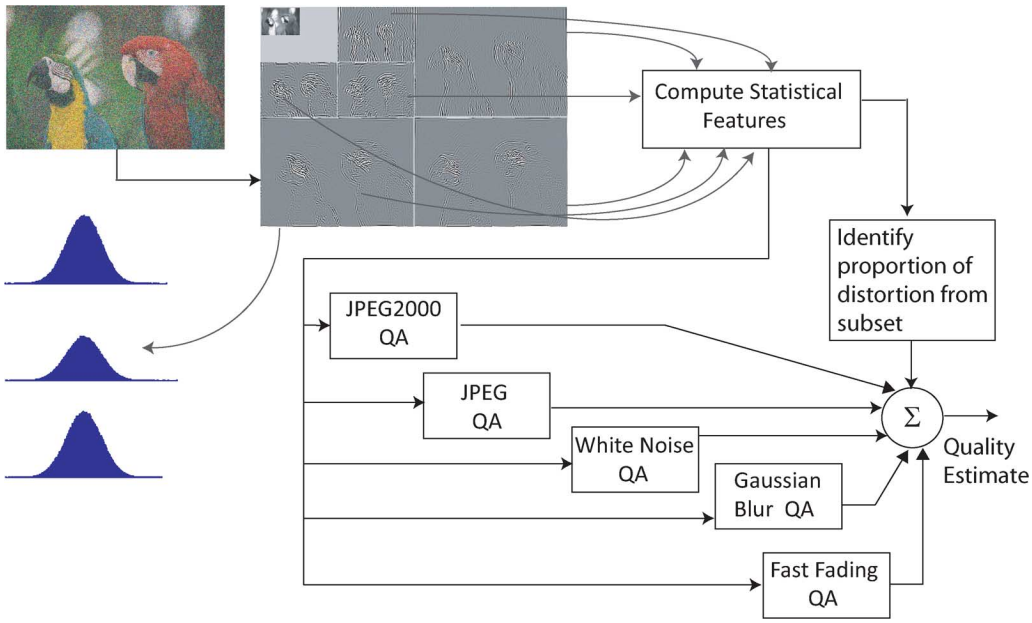

Fig. 1. Proposed Distortion identification-based Image Verity and INtegrity Evaluation (DIIVINE) index consists of two stages: probabilistic distortion identification followed by distortion-specific quality assessment as illustrated here.

B. Holistic IQA Algorithms

Li proposed a series of heuristic measures to characterize image quality based on three quantities—edge sharpness, random noise level (impulse/additive white Gaussian noise), and structural noise [29]. Edge sharpness is measured using an edge-detection approach, while the random noise level is measured using a local smoothness approach (impulse noise) and PDE-based model (Gaussian noise). Structural noise as defined by Li relates to blocking and ringing from compression techniques such as JPEG and JPEG2000. Unfortunately, the author does not analyze the performance of the proposed measures, nor propose a technique to combine the measures to produce a general-purpose quality assessment algorithm.

Gabrada and Cristobal proposed an innovative strategy for blind IQA which utilized the Renyi entropy measure [30] along various orientations to measure anisotropy. The proposed approach is attractive since natural images are anisotropic in nature and possesses statistical structure that distortions destroy. They measure mean, standard deviation, and range of the Renyi entropy along four pre-defined orientations in the spatial domain and demonstrate their correlation with perceived quality. Unfortunately, a thorough evaluation of the proposed measure is again lacking.

Recently, Saad and Bovik proposed a general-purpose blind quality assessment algorithm that computes four features in the DCT domain: DCT kurtosis, DCT contrast, and two anisotropy measures inspired from [30]—maximum and variance of the Renyi entropy along four orientations [31]. Features are extracted over two scales and a Gaussian distribution is used to model the relationship between the DMOS and the extracted features. The measure was shown to perform well in terms of correlation with human perception across distortion categories.

Blind/NR *video* quality assessment (VQA) is an important problem that has followed a similar trajectory. Some authors have proposed techniques which measure blockiness, blur, corner outliers, and noise separately, and use a Minkowski sum to pool the measures of quality together [32], [33]. In both these approaches, distortion-specific indicators of quality are computed and pooled using a variety of pre-fixed thresholds and training, as against our approach that uses concepts from NSS to produce a modular and easily extensible approach that can be modified to include other distortions than those discussed here. We anticipate that the approach taken here could be eventually extended to video to achieve good results.

III. Scene Statistics of Distorted Images

The DIIVINE approach for NR IQA proceeds as follows. The distorted image is first decomposed using a scale-space-orientation decomposition (loosely, a wavelet transform) to form oriented band-pass responses. The obtained subband coefficients are then utilized to extract a series of statistical features. These statistical features are stacked to form a vector which is a statistical description of the distortion in the image. Our goal is to utilize these feature vectors across images to perform two tasks in sequence: 1) identify the probability that the image is afflicted by one of the multiple distortion categories, then 2) map the feature vector onto a quality score for each distortion category, i.e., build a regression model for each distortion category to map the features onto quality, conditioned on the fact that the image is impaired by that particular distortion category (i.e., distortion-specific QA). The probabilistic distortion identification estimate is then combined with the distortion-specific quality score to produce a final quality value for the image. The method described here is illustrated in Fig. 1 and is labeled as the DIIVINE index.

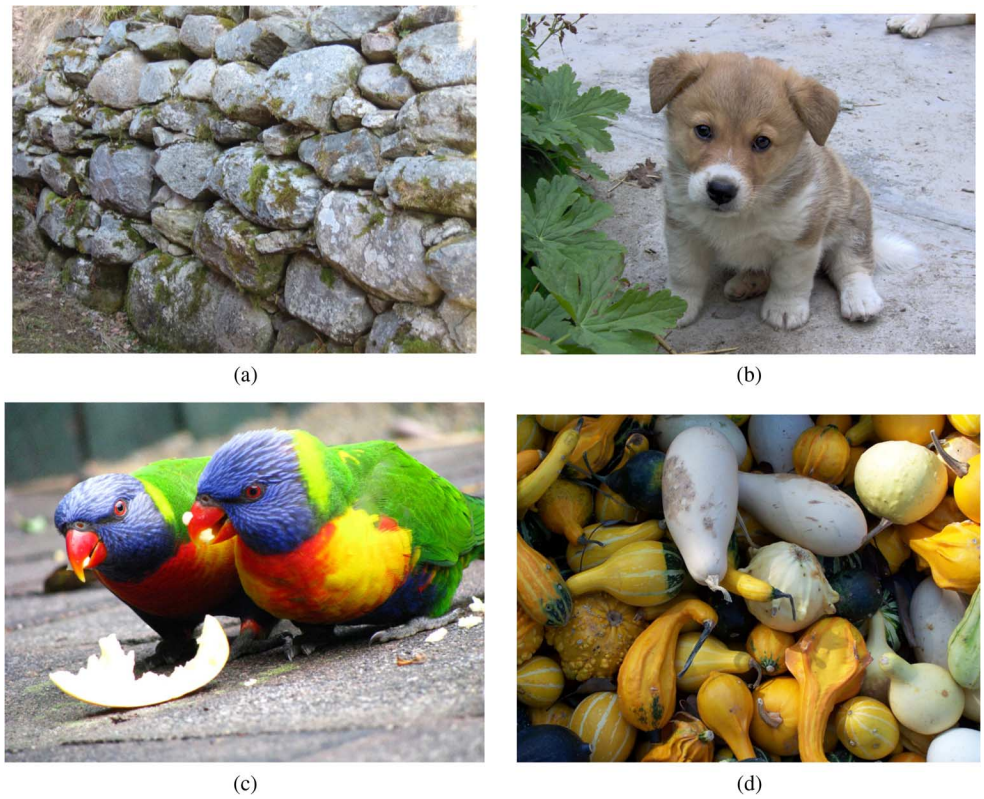

Fig. 2. (a)–(d) Images used to demonstrate features derived under NSS models.

A. Statistical Model for Wavelet Coefficients

In the DIIVINE framework, a set of neighboring wavelet coefficients are modeled using the Gaussian scale mixture (GSM) model [34]. An N-dimensional random vector Y is a GSM if $Y \equiv z \cdot U$ where $\equiv$ denotes equality in probability distribution, U is a zero-mean Gaussian random vector with covariance C_U, and z is a scalar random variable called a mixing multiplier. The density of Y is then given by

$$p_Y(y) = \int \frac{1}{(2\pi)^{N/2}\left|z^2 C_U^{1/2}\right|} exp\left(\frac{-Y^T C_U^{-1} Y}{z^2}\right) p_Z(z)dz.$$

The GSM model has been used to model the marginal and joint statistics of the wavelet coefficients of natural images [3], [34], where the vector Y is formed by clustering a set of neighboring wavelet coefficients within a subband, or across neighboring subbands in scale and orientation.

Next, we shall describe the statistical features that we extract from the distorted image and motivate their choice. In order to illustrate how each of these features behaves in natural and distorted images, we shall use the natural un-distorted reference images shown in Fig. 2 and their distorted counterparts, a subset of which are shown in Fig. 3. The distortions in Fig. 3 are exactly the same as the ones that we consider in this paper and that span the set of distortions in the LIVE Image Quality Assessment Database [35]—JPEG and JPEG2000 (JP2k) compression, additive white noise (WN), Gaussian blur (blur), and a Rayleigh fading channel labeled fast fading (FF).

B. Extracting Scene Statistics

In order to extract statistics from distorted images, we utilize the steerable pyramid decomposition [36]. The steerable pyramid is an overcomplete wavelet transform that allows for increased orientation selectivity. The choice of the wavelet transform was motivated by the fact that the scale-space-orientation decomposition that the wavelet transform performs mirrors models of spatial decomposition that occurs in area V1 of the primary visual cortex [37], [38]. The steerable pyramid has been previously used for FR IQA [5] as well as RR IQA [3] with success. Note that we do not use the complex version of the steerable pyramid as in [39], but that used in [5].

Given an image whose quality is to be assessed, the first step is to perform a wavelet decomposition using a steerable pyramid over two scales and six orientations. We have found that an increased degree of orientation selectivity is beneficial for the purpose of QA—more so than selectivity over more than two scales. The choice of steerable filters was also motivated by its increased orientation selectivity. Our experiments have indicated that increasing the number of scales beyond two does not improve performance. The resulting decomposition results in 12 subbands across orientations and scales labeled s_α^θ, where $\alpha \in \{1, 2\}$ and $\theta \in \{0°, 30°, 60°, 90°, 120°, 150°\}$.

The next step is to perform the perceptually significant process of divisive normalization [40]. Divisive normalization or contrast-gain-control was proposed in the psychovisual literature in order to account for the nonlinear behavior of certain cortical neurons. Such normalization accounts for interactions between neighboring neurons and governs the response of a

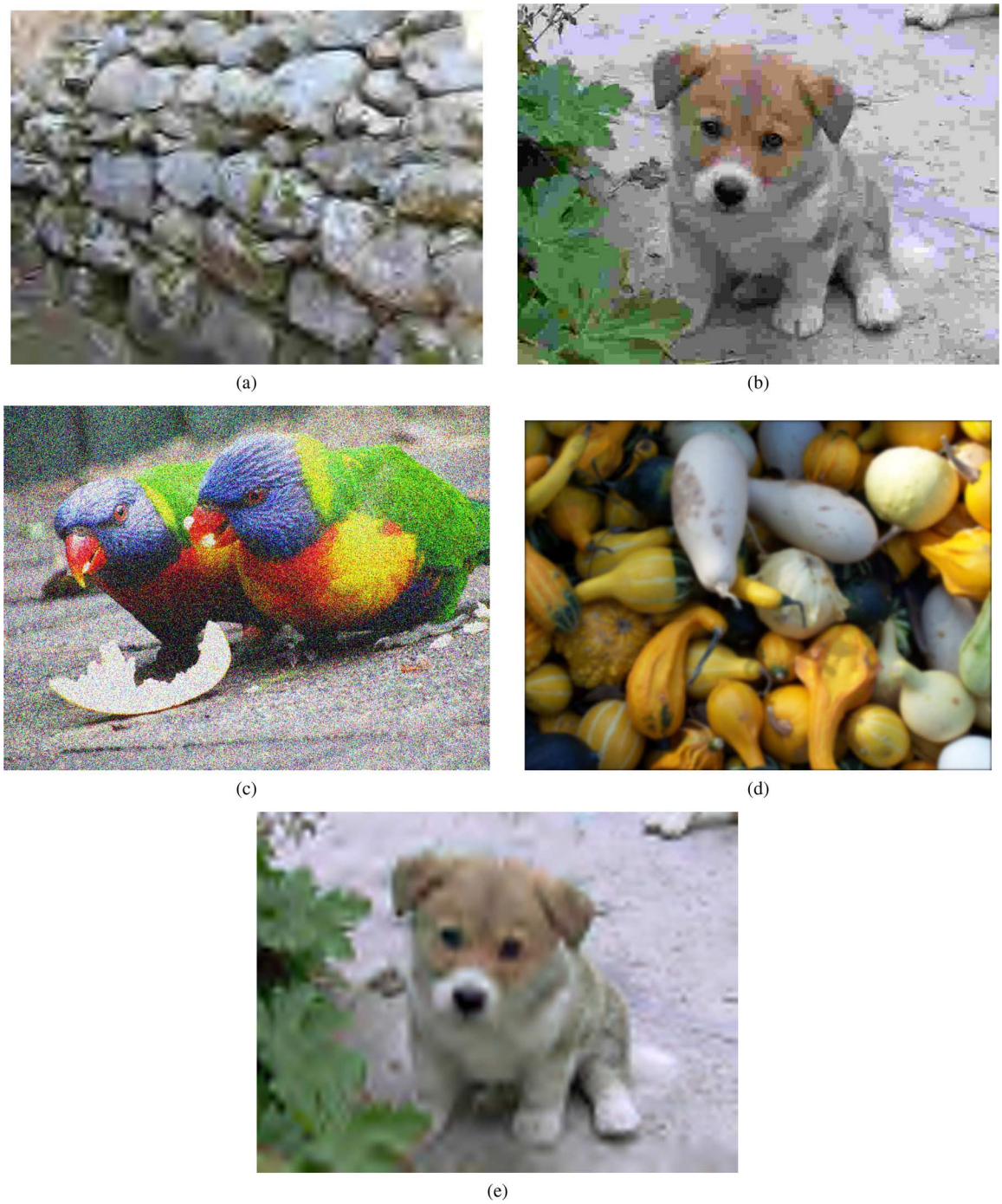

Fig. 3. Subset of the distorted versions of images in Fig. 2. (a)–(e) correspond to the following distortions: (a) JP2k compression, (b) JPEG compression, (c) white noise, (d) Gaussian blur, and (e) fast fading distortion.

neuron based on the responses of a pool of neurons surrounding it [40]. Divisive normalization also reduces the statistical dependencies between subbands thereby de-coupling subband responses to a certain degree [34], [40]. Further, divisive normalization models partially account for contrast masking [38]—an essential ingredient in QA algorithm design. Divisive normalization has been explicitly used for RR IQA in the past [3]. FR IQA techniques such as the visual information fidelity index (VIF) [5] and the structural similarity index (SSIM) [16] also utilize divisive normalization, albeit in an implicit manner [41]. Finally, the successful MOVIE index, a recently proposed FR VQA algorithm [42], also utilizes such a technique (drawing inspiration from the Teo and Heeger model [43]). Here, divisive normalization is implemented as described in [3].

Specifically, given a subband coefficient y, our goal is to compute a normalization parameter p, based on responses from neighboring subbands in order to finally compute $\hat{y} = y/p$. To estimate p, we utilize the previously defined local statistical model for natural images—the GSM model [34]. In our implementation, for a center coefficient y_c at each subband, we define

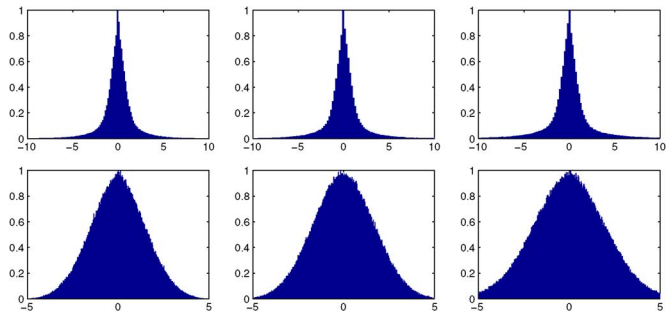

Fig. 4. Figure demonstrating the effect of divisive normalization on the subband statistics of image in Fig. 2(a). The first row shows the histogram of subband coefficient distributions before divisive normalization, while the second row is the distribution after normalization. Divisive normalization makes the subband statistics of *natural images* more Gaussian-like, as compared to the Laplacian nature of the pre-normalized subband coefficients.

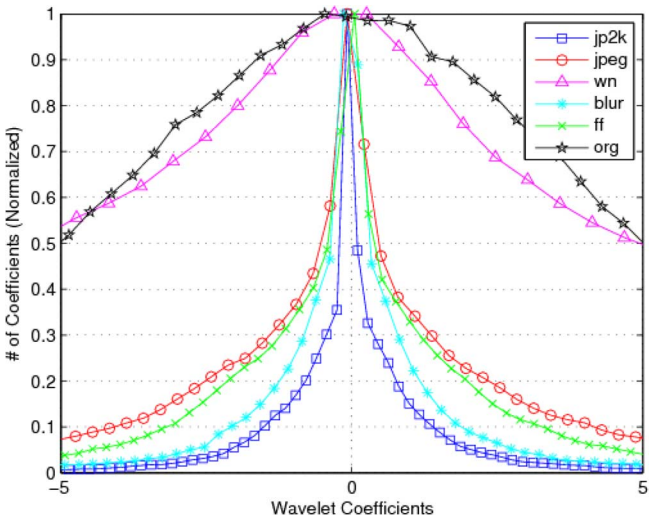

Fig. 5. Subband statistics from $d_1^{0^\circ}$ of the image in Fig. 2(c) for different distortions. Notice how each distortion affects the statistics in a characteristic way.

a divisive normalization transform (DNT) neighborhood vector Y that contains 15 coefficients, including 9 from the same subband (3×3 neighborhood around y_c), 1 from the parent band, and 5 from the same spatial location in the neighboring bands at the same scale. Given this vector Y, the normalization coefficient p is computed as $p = \sqrt{Y^T C_U^{-1} Y / N}$. This computation is undertaken at each coefficient in each subband to produce a divisively-normalized set of subbands—d_α^θ, where $\alpha \in \{1, 2\}$ and $\theta \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$. The interested reader is referred to [3] for details on the divisive normalization procedure.

In order to visualize how divisive normalization affects the statistics of the subband coefficients, Fig. 4 plots a histogram of coefficients from s_1^θ and d_1^θ, where $\theta \in \{0^\circ, 30^\circ, 60^\circ\}$. The normalization makes the subband statistics more Gaussian-like for natural images.

In order to demonstrate that subband statistics are affected by each distortion in a particular fashion, Fig. 5 plots the coefficient distributions from $d_1^{0^\circ}$ of the image in Fig. 2(c) for each distortion considered here. It should be clear that each distortion affects the statistics in a characteristic way which is essentially independent of the content (e.g., WN always increases the variance of subband coefficients).

Given that each distortion affects subband statistics characteristically, the goal is to compute marginal and joint statistics across subbands in order to extract features that are relevant to the perceived quality of the image.

1) Scale and Orientation Selective Statistics ($f_1 - f_{24}$): Subband coefficients from each of the 12 subbands are parametrized using a generalized Gaussian distribution (GGD). The GGD is

$$f_X(x; \mu, \sigma^2, \gamma) = a e^{-[b|x-\mu|]^\gamma} \quad x \in \Re$$

where μ, σ^2, and γ are the mean, variance, and shape-parameter of the distribution and

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)}$$
$$b = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}$$

and $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \quad x > 0.$$

The shape parameter γ controls the "shape" of the distribution. For example, $\gamma = 2$ yields a Gaussian distribution and $\gamma = 1$ yields a Laplacian distribution. The parameters of the distribution (μ, σ^2, and γ) are estimated using the method proposed in [44]. GGD has also been used before to model the subband statistics of natural images in RR IQA [4]. Since wavelet subband responses are zero mean, we have to estimate σ^2 and γ for each subband leading to a total of 24 features. $f_1 - f_{12}$ correspond to σ^2 across subbands and $f_{13} - f_{24}$ correspond to γ across subbands.

At this juncture, it may be prudent to explain the choice of the GGD. The divisive normalization procedure tends to produce coefficients distributed in a Gaussian manner for *natural* images. In the presence of distortion however, this Gaussianity at the output of the normalization procedure is not guaranteed. For example, from Fig. 5, it should be clear that distortions such as JPEG, JP2k, Blur, and FF lead to highly kurtotic (non-Gaussian) distributions even after the divisive normalization procedure. Since the shape parameter of the GGD will capture this non-Gaussian nature, the GGD fit is utilized here as against a simple Gaussian fit. We note that a similar procedure was used for RR IQA in [3].

In order to demonstrate how these subband features affect quality, Fig. 6 shows a plot of $log_e(\sigma^2)$ versus γ for one of the subbands for each of the reference images in Fig. 2 and their associated distorted versions.

We have previously shown that these simple marginal statistics when used in a simple, preliminary blind IQA algorithm—the Blind Image Quality Index (BIQI) [12]—do a good job of identifying the distortion afflicting the image and predicting the perceived quality of an image [12], [45]. Here we full develop the 2-stage NSS-based IQA concept introduced in [12], by deploying a much richer set of NSS-based features that

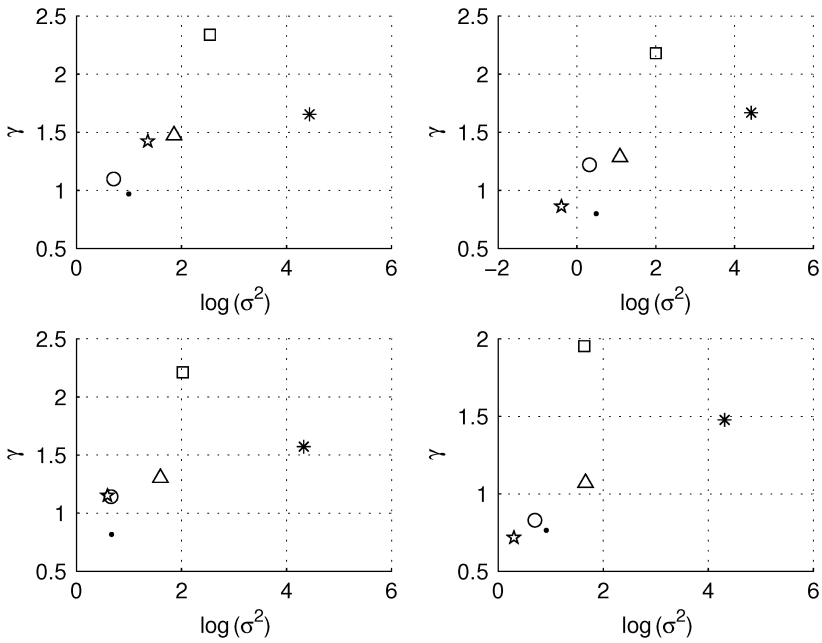

Fig. 6. Plot of $log_e(\sigma^2)$ versus γ for $d_1^{120^\circ}$ for each of the images considered in Fig. 2 and their associated distorted versions. Notice how each distortion seems to cluster in a particular region, immaterial of the image content. Images (a)–(d): left to right, top to bottom. Reference image ($\square$), JPEG ($\circ$), JPEG2K ($\triangle$), WN ($*$), blur ($\cdot$), and ff ($\star$).

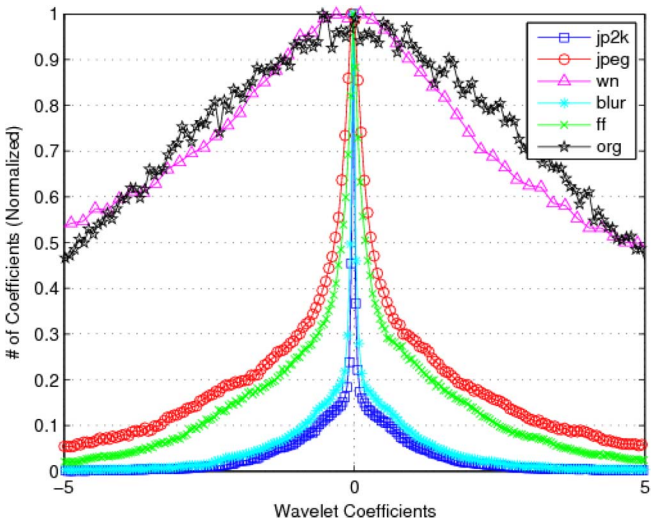

Fig. 7. Histogram (normalized) of coefficients from $d_1^{0^\circ}$ and $d_2^{0^\circ}$ for the image in Fig. 2(c) and its various distorted versions. Notice the difference in distributions of these across-scale coefficients for natural and distorted images.

capture the dependencies between subband coefficients over scales and orientations, as well as utilizing the perceptually relevant divisive normalization procedure.

2) Orientation Selective Statistics ($f_{25} - f_{31}$): Images are naturally multiscale. Further, there exists a relationship between subbands at the same orientation and across different scales. Distortions in an image will affect these across-scale statistics. For example, in Fig. 7, we plot a histogram of coefficients from $d_1^{0^\circ}$ and $d_2^{0^\circ}$ for the image in Fig. 2(c) and its various distorted versions. In order to plot these distributions in 1-D, these subbands were stacked together to form a large vector, whose histogram we plot. Notice the difference in distributions of these across-scale coefficients for natural and distorted images.

In order to capture the variation seen in Fig. 7, we again utilize a GGD fit. The 1-D GGD is now fit to the coefficients obtained by stacking together coefficients from subbands at the same orientation but at different scales. Specifically, 6 GGD fits corresponding to each one of $\{d_1^\theta, d_2^\theta\}$, $\theta \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$ are computed. Again, these fits are zero-mean and we compute two parameters—σ^2 and γ. In our experiments, σ^2 does not add any information about the perceived quality, and hence, we use only the computed γ's as features. Further, we also compute a GGD fit when all of the subbands are stacked together (i.e., $\{d_\alpha^\theta\}, \forall \alpha, \theta$) and use the γ parameter again as our feature. Thus, $f_{25} - f_{30}$ correspond to γ from the statistics across scales over different orientations, while f_{31} corresponds to γ from the statistics across subbands. In Fig. 8, we plot these computed γ values for each of the images in Fig. 2 and their associated distorted versions.

3) Correlations Across Scales ($f_{32} - f_{43}$): One of the primary stages in human visual processing is filtering of the visual stimulus by the retinal ganglion cells [38]. These cells have center-surround-difference properties and have spatial responses that resemble difference of Gaussians (DoG) functions [38], [46]. The responses of these cells serve a variety of likely purposes including dynamic range compression, coding, and enhancement of features such as edges [38], [46]. Image compression algorithms such as EZT and SPIHT [47], [48] offer evidence of correlations across scales as well. Statistics of edges have been used for blur quality assessment [49]. Given that edges are important, it is reasonable to suppose that there exist elegant statistical properties between high-pass (HP) responses of natural images and their band-pass (BP) counterparts. Indeed, in our experiments, we found that such

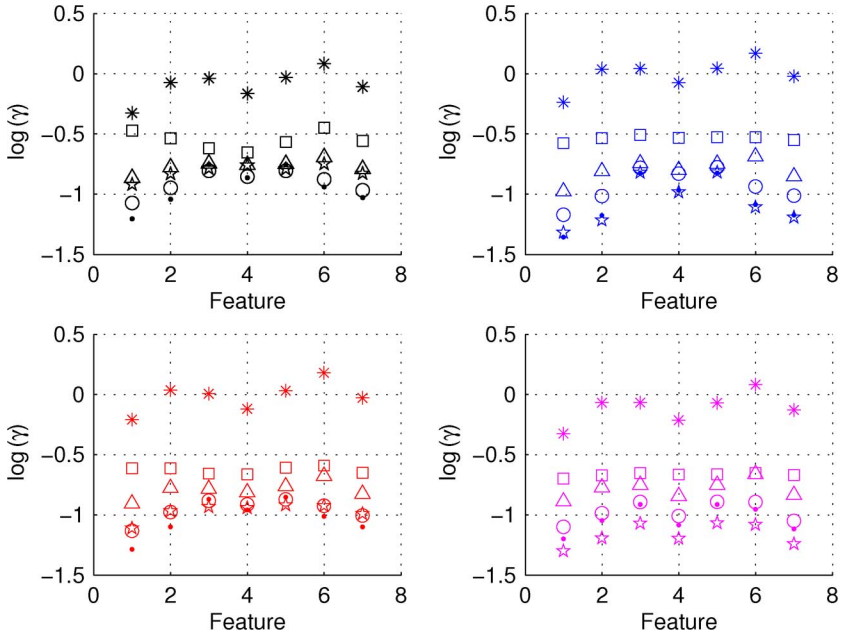

Fig. 8. Orientation selective statistics (γ) for reference and distorted images. Images (a)–(d): left to right, top to bottom. Reference image ($\square$), JPEG ($\circ$), JPEG2K ($\triangle$), WN ($*$), blur ($\cdot$), and ff ($\star$). 1–7 on the x-axis correspond to $f_{25} - f_{31}$.

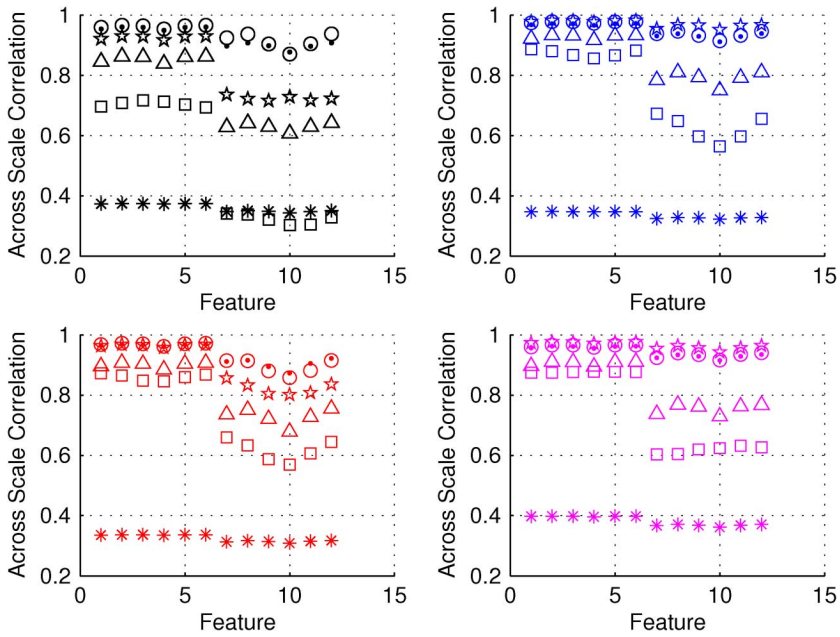

Fig. 9. Across scale correlation statistics for reference and distorted images. Images (a)–(d): left to right, top to bottom. Reference image ($\square$), JPEG ($\circ$), JPEG2K ($\triangle$), WN ($*$), blur ($\cdot$), and ff ($\star$). 1–12 on the x-axis correspond to $f_{32} - f_{43}$.

a relationship exists for natural images and this relationship is affected by the presence of distortion. We model high-pass band-pass correlations in order to capture these dependencies.

Each BP subband is compared with the HP residual band (obtained from the steerable pyramid transform) using a windowed structural correlation [16]. Specifically, the BP and HP bands are filtered using a 15×15 Gaussian window with $\sigma = 1.5$ [16]. The structural correlation is then computed as

$$\rho = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

where σ_{xy} is the cross-covariance between the windowed regions from the BP and HP bands, and σ_x^2, σ_y^2 are their windowed variances, respectively; C_2 is a stabilizing constant that prevents instabilities from arising when the denominator tends to 0, and its value is the same as that used in [16]. The mean of the correlation map so obtained is used as the correlation feature.

Fig. 9 plots the value of the correlation coefficient for each of the 12 subbands and for all images considered in Fig. 2 and their associated distorted versions. Again, distortion-specific clustering immaterial of content is evident.

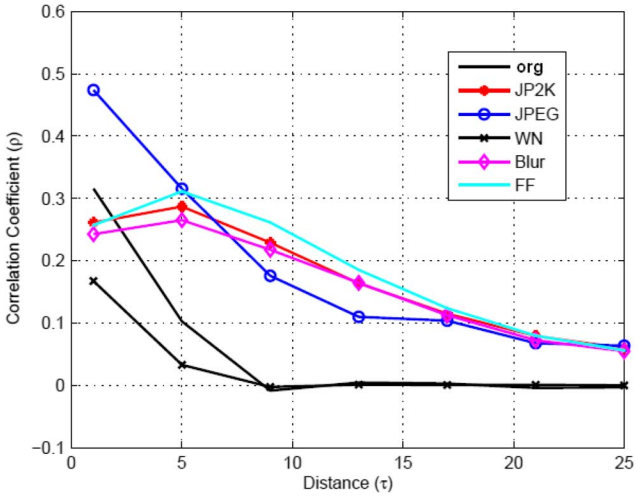

Fig. 10. Plot of spatial correlation coefficient $(\rho(\tau))$ for various distance τ for one subband of an image, across distortions.

Since there are 12 subbands, 12 such correlations are computed, yielding features $f_{32} - f_{43}$.

4) Spatial Correlation $(f_{44} - f_{73})$*:* Throughout this discussion, we have emphasized the observation that natural images are highly structured and that distortions modify this structure. While we have captured many such modifications in the subband domain, one particular form of scene statistics that remains neglected is the spatial structure of the subbands. Natural images have a correlation structure that, in most places, smoothly varies as function of distance.

In order to capture spatial correlation statistics, we proceed as follows. For each τ, $\tau \in \{1, 2, \ldots, 25\}$, and for each d_1^θ, $\theta \in \{0°, 30°, 60°, 90°, 120°, 150°\}$, we compute the joint empirical distribution between coefficients at (i, j) and $\mathcal{N}_8^\tau(i, j)$, where $\mathcal{N}_8^\tau$ denotes the set of spatial locations at a distance of τ (chessboard distance). The joint distribution attained for a value of τ can be thought of as the joint distribution $p_{XY}(x, y)$ between two random variables X and Y. To estimate correlation between these two variables, we compute

$$\rho(\tau) = \frac{E_{p_{XY}(x,y)}\left[\left(X - E_{p_X(x)}[X]\right)^T \left(Y - E_{p_Y(y)}[Y]\right)\right]}{\sigma_X \sigma_Y}$$

where $E_{p_X(x)}[X]$ is the expectation of X with respect to the marginal distribution $p_X(x)$ obtained from the computed joint distribution, and similarly for Y and (X, Y). In order to visualize this $\rho(\tau)$ for different distortions, in Fig. 10, we plot ρ as a function of τ for the image in Fig. 2(b) and its distorted versions in Fig. 3. Notice how the presence of distortion alters the spatial correlations statistics.

Once $\rho(\tau)$ is obtained, we parameterize the obtained curve by fitting it with a 3rd order polynomial, where τ is the distance at which the estimate of ρ is computed. Such a fit is computed for $d_1^\theta, \forall\theta$. The coefficients of the polynomial and the error between the fit and the actual $\rho(\tau)$ form the features—$f_{44} - f_{73}$.

5) Across Orientation Statistics $(f_{74} - f_{88})$*:* One set of statistics that remains unexplored are statistical correlations that natural images exhibit across orientations. In order to capture the distortion-induced modifications to these statistical correlations across orientations, we compute windowed structural correlation (same as the across scale statistics) between all possible pairs of subbands at the coarsest scale. The set of features is the lowest 5% [50], [51] of the structural correlation values so obtained for each pair, leading to a total of ${}^6C_2 = 15$ features—$f_{74} - f_{88}$. In Fig. 11, we plot the value of these across orientation features for each of the images considered in Fig. 2 and their associated distorted versions. Notice clustering of distortions independent of content.

All of the features described here are listed in Table I for reference.

Until now, we defined a series of statistical features that we extracted from subband coefficients, and we described how each of these statistics is affected in the presence of distortion. However, the relationship to quality for each of these features requires clarification. Hence, in Fig. 12, we plot the Spearman's rank ordered correlation coefficient (SROCC) across each of the distorted categories across all distorted images in the LIVE image database. Note that *no training is undertaken here*; the plot is simply to justify the choice of the features as good indicators of quality. As is clear, some features predict perceived quality with greater correlation with human perception than others.

IV. Distortion-Identification-Based Image Verity and Integrity Evaluation

Our 2-stage approach to NR IQA—as realized here in constructing the DIIVINE index—consists of utilizing the features extracted as described above for distortion-identification as well as for distortion-specific quality assessment [12]. Both these stages require a calibration process that relates the computed feature to the distortion-class associated with it and the human opinion score associated with it. This calibration is achieved using training, where a set of images whose ground truth class of distortion as well the associated human opinion score (i.e., perceived quality score) is known. Given this training set, we calibrate the two stages of distortion-identification and distortion-specific quality assessment. Once calibrated, DIIVINE is capable of assessing the quality of any distorted image without the need for the reference. Note that the calibration stage also does not require the reference image.

Given a training set of images with known distortion class, spanning the range of distortions (n) the algorithm is being calibrated for, we train a classifier with the true class and the feature vector as inputs. The classifier "learns" the mapping from feature space to class label, and once calibration is achieved, the trained classifier produces an estimate of the class of distortion given an input image (i.e., the feature vector associated with the input image).

Similarly, given a set of training images with known quality scores for each of the n distortion classes, we train n regression modules that map the feature vector to the associated quality score. Since each module is trained specifically for each distortion, these regression modules, once trained, function as distortion-specific assessors of quality, i.e., each trained module will produce an estimate of quality (when given as input an

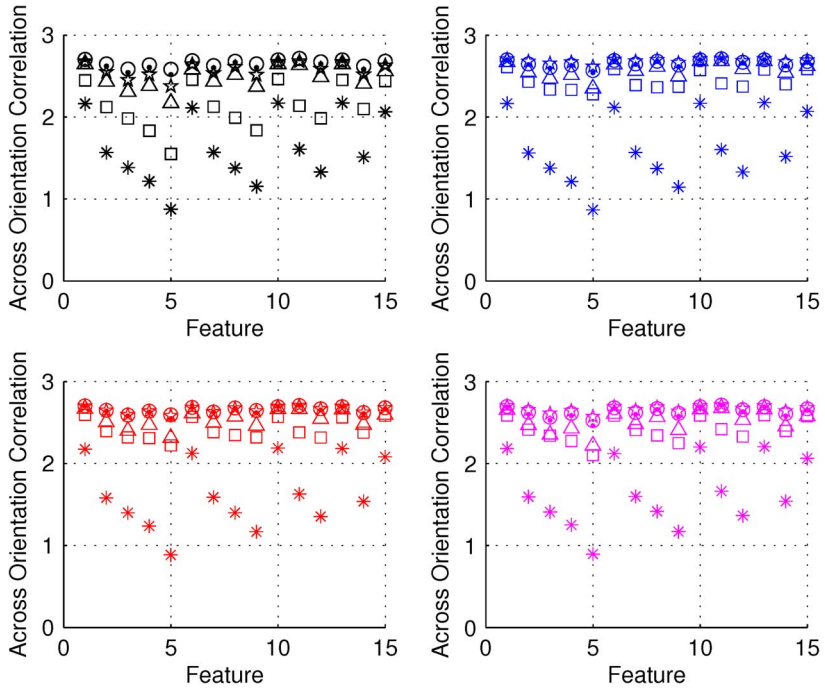

Fig. 11. Across-orientation statistics for reference and distorted images. Images (a)–(d): left to right, top to bottom. Reference image ($\square$), JPEG ($\circ$), JPEG2K (Δ), WN ($*$), blur ($\cdot$), and ff ($\star$). 1–15 on the x-axis correspond to $f_{74} - f_{88}$.

TABLE I
TABLE LISTING EACH OF THE FEATURES CONSIDERED HERE AND THE METHOD IN WHICH THEY WERE COMPUTED

Feature ID	Feature Description	Computation Procedure
$f_1 - f_{12}$	Variance of subband coefficients	Fitting a generalized Gaussian to subband coefficients
$f_{13} - f_{24}$	Shape parameter of subband coefficients	Fitting a generalized Gaussian to subband coefficients
$f_{25} - f_{31}$	Shape parameter across subband coefficients	Fitting a generalized Gaussian to orientation subband coefficients
$f_{32} - f_{43}$	Correlations across scales	Computing windowed structural correlation between filter responses
$f_{44} - f_{73}$	Spatial correlation across subbands	Fitting a polynomial to the correlation function
$f_{74} - f_{88}$	Across orientation statistics	Computing windowed structural correlation between adjacent orientations at same scale

image/feature vector) under the assumption that the image is distorted with that particular distortion. The input image whose quality is to be assessed is passed through each of these trained distortion-specific quality assessment modules, and hence, we receive $\vec{q}$, an n-dimensional vector corresponding to the quality estimates from each of these n regression modules.

In our approach, the classifier does not produce a hard classification. Instead, probability estimates are extracted from the classifier, which indicate the confidence that the trained classifier demonstrates in placing the input in each of the n classes. Thus, given an input image/feature vector, the trained classifier produces an n-dimensional vector $\vec{p}$, which represent probabilities of the input belonging to each of the n classes.

Given the two vectors $\vec{p}$ and $\vec{q}$, $\text{DIIVINE} = \vec{p}^T\vec{q}$—i.e., each distortion-specific quality score is weighted by the probability of that distortion being present in the image.

Obviously, one can choose to utilize any classifier and any regression tool to map the feature vectors onto classes/quality scores. In this implementation, we utilize a support vector machine (SVM) for classification and support vector regression (SVR) for regression [52], [53]. The choice of SVM and SVR were motivated by the fact that these tools have been shown to perform well on high-dimensional hard classification/regression problems [54]. The interested reader is directed to [52]–[54] for detailed explanations of SVMs and SVRs.

We utilize the libSVM package [55] in order to implement the SVM and the SVRs. The kernel used for both classification and regression is the radial basis function (RBF) kernel, whose parameters are estimated using cross-validation on the training set.

V. PERFORMANCE EVALUATION

A. LIVE IQA Database

We tested the DIIVINE index on the popular LIVE IQA database [35], which consists of 29 reference images and 779

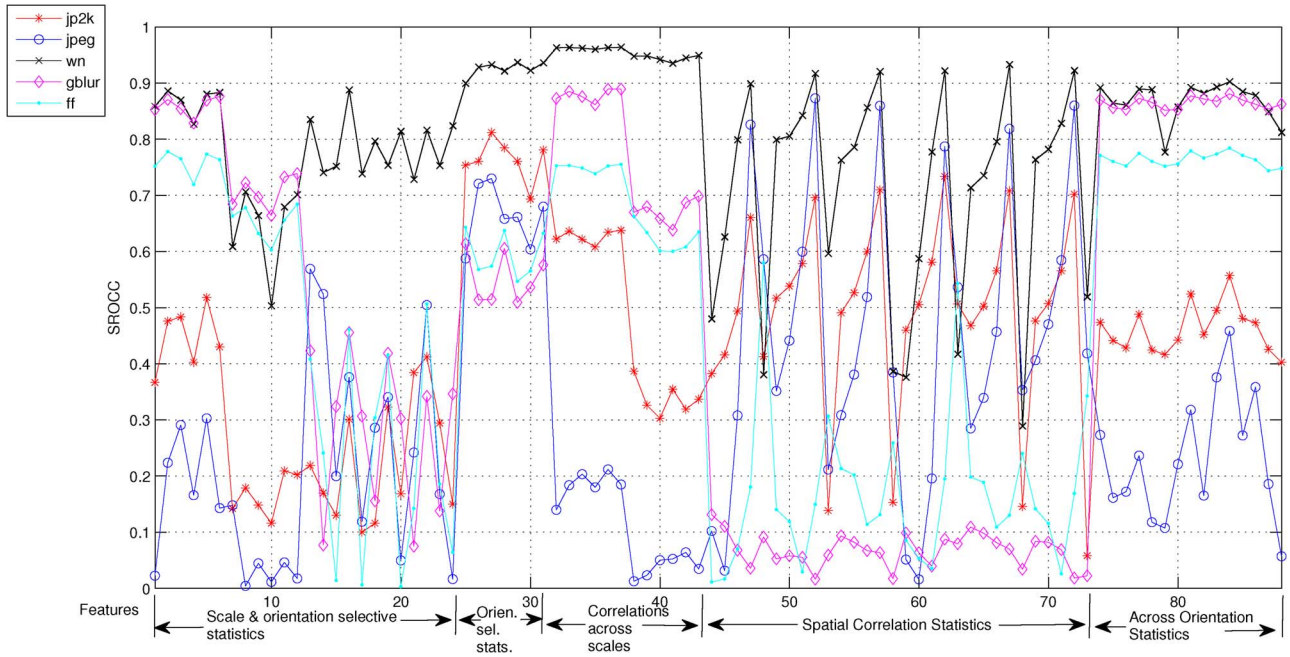

Fig. 12. Spearman's rank ordered correlation coefficient (SROCC) for each of the features from Table I on the LIVE image database.

TABLE II
MEDIAN SPEARMAN'S RANK ORDERED CORRELATION COEFFICIENT (SROCC) ACROSS 1000 TRAIN-TEST TRIALS ON THE LIVE IMAGE QUALITY ASSESSMENT DATABASE. *ITALICIZED* ALGORITHMS ARE NR IQA ALGORITHMS, OTHERS ARE FR IQA ALGORITHMS

	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.868	0.885	0.943	0.761	0.875	0.866
SSIM (SS)	0.938	0.947	0.964	0.907	0.940	0.913
BIQI-PURE	*0.736*	*0.591*	*0.958*	*0.778*	*0.700*	*0.726*
BIQI-4D	*0.802*	*0.874*	*0.958*	*0.821*	*0.730*	*0.824*
Anisotropic IQA	*0.173*	*0.086*	*0.686*	*0.595*	*0.541*	*0.323*
BLIINDS	*0.805*	*0.552*	*0.890*	*0.834*	*0.678*	*0.663*
DIIVINE	*0.913*	*0.910*	*0.984*	*0.921*	*0.863*	*0.916*

TABLE III
MEDIAN LINEAR CORRELATION (LCC) ACROSS 1000 TRAIN-TEST TRIALS ON THE LIVE IMAGE QUALITY ASSESSMENT DATABASE. *ITALICIZED* ALGORITHMS ARE NR IQA ALGORITHMS, OTHERS ARE FR IQA ALGORITHMS

	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.879	0.903	0.917	0.782	0.880	0.862
SSIM (SS)	0.940	0.947	0.983	0.902	0.952	0.906
BIQI-PURE	*0.750*	*0.630*	*0.968*	*0.800*	*0.722*	*0.740*
BIQI-4D	*0.819*	*0.879*	*0.968*	*0.843*	*0.771*	*0.833*
Anisotropic IQA	*0.130*	*0.083*	*0.490*	*0.469*	*0.420*	*0.187*
BLIINDS	*0.807*	*0.597*	*0.914*	*0.870*	*0.743*	*0.680*
DIIVINE	*0.922*	*0.921*	*0.988*	*0.923*	*0.888*	*0.917*

distorted images that span various distortion categories—JPEG and JPEG2000 compression, white noise, Gaussian blur, and a Rayleigh fading channel (fast fading), along with the associated human differential mean opinion scores (DMOS), which are representative of the perceived quality of the image.

Since DIIVINE requires a training stage in order to calibrate the relationship between the extracted statistical features and the distortion category, as well as DMOS, we split the LIVE dataset into two non-overlapping sets—a training set and a testing set. The training set consists of 80% of the reference images and their associated distorted versions while the testing set consists of the remaining 20% of the reference images and their associated distorted versions. The classification and regression modules are trained on the training set and the results are then tested on the testing set. In order to ensure that the proposed approach is robust across content and is not governed by the specific train-test split utilized, we repeat this random 80% train—20% test split 1000 times on the LIVE dataset and evaluate the performance on each of these test sets. The figures reported here are the median of the indices used for performance across these 1000 train-test iterations.[2]

TABLE IV
MEDIAN ROOT-MEAN-SQUARED ERROR (RMSE) ACROSS 1000 TRAIN-TEST TRIALS ON THE LIVE IMAGE QUALITY ASSESSMENT DATABASE. *ITALICIZED* ALGORITHMS ARE NR IQA ALGORITHMS, OTHERS ARE FR IQA ALGORITHMS

	JP2K	JPEG	WN	Gblur	FF	All
PSNR	11.87	13.60	11.14	11.25	13.33	13.89
SSIM (SS)	8.59	10.11	5.17	7.96	8.74	11.56
BIQI-PURE	*16.54*	*24.58*	*6.93*	*11.10*	*19.48*	*18.36*
BIQI-4D	*14.34*	*15.06*	*6.94*	*9.90*	*17.90*	*15.05*
Anisotropic IQA	*24.65*	*31.40*	*24.41*	*16.19*	*25.44*	*26.68*
BLIINDS	*14.78*	*25.32*	*11.27*	*9.08*	*18.62*	*20.01*
DIIVINE	*9.66*	*12.25*	*4.31*	*7.07*	*12.93*	*10.90*

The indices used to measure performance of the algorithm are the Spearman's rank ordered correlation coefficient (SROCC), the linear (Pearson's) correlation coefficient (LCC), and the root mean squared error (RMSE) between the predicted score and the DMOS. LCC and RMSE are computed after passing the algorithmic scores through a logistic nonlinearity as in [35]. A value close to 1 for SROCC and LCC and a value close to 0 for RMSE indicates superior correlation with human perception. The median SROCC, LCC, and RMSE values across these 1000 train-test trials are tabulated in Tables II–IV, for each distortion category, as well as across distortion categories.

[2]We use the realigned DMOS scores as recommended in [35] and report results only on the distorted images, as in [35].

TABLE V
MEDIAN CLASSIFICATION ACCURACY OF CLASSIFIER ACROSS 1000 TRAIN-TEST TRIALS ON THE LIVE IMAGE DATABASE

	JP2K	JPEG	WN	Gblur	FF	All
Class. Acc.(%)	80.00	81.10	100	90.00	73.33	83.75

We also report the performance of two FR IQA algorithms—PSNR and the SSIM. The former has been used (despite much criticism [56], [57]) as a measure of quality for many years, and the latter is now gaining popularity as a good-yet-efficient assessor of perceived image quality. We also tabulate the performances of several NR IQA algorithms, including original algorithms used to demonstrate the concept of the two-stage framework—the Blind Image Quality Index (BIQI)—BIQI-PURE and BIQI-4D,[3] and the two holistic NR IQA algorithms that we have previously discussed—Anisotropy based NR IQA [30] and the BLind Image Integrity Notator using DCT Statistics (BLIINDS) index [31]. The BIQI realizations are available online [58], and the implementation of the anisotropy measure[4] was obtained from [59]. We implemented the BLIINDS index as described in [31].

It should be clear that DIIVINE performs well in terms of correlation with human perception. Further, DIIVINE improves upon the BIQI realizations, and is superior to the two other holistic NR IQA approaches. Remarkably, DIIVINE also trumps the *full-reference* PSNR, for each distortion separately as well as across distortion categories. However, the most salient observation from Tables II–IV is that the proposed *no-reference* approach is competitive with the *full-reference* SSIM index. This is no mean achievement, since the SSIM index is currently one of the most popular FR IQA algorithms.

Although distortion-identification/classification is not explicitly performed in the two-stage framework used here (recall that we use a probabilistic classification in which the probability of an image belonging to a particular distortion category is estimated), in order to demonstrate that the features are capable of identifying the distortion afflicting the image with high accuracy, in Table V, we list the median classification accuracy of the classifier for each distortion category and the overall accuracy as well. The caveat here is that the actual accuracy of the classifier is not of great import for the proposed approach, since hard classification is never performed. The classification accuracies are reported for completeness.

B. Statistical Significance Testing

We tabulated the median correlation values of DIIVINE as well as other NR and FR IQA algorithms in the previous section. Although the presented results show some differences in terms of the median correlation, in this section, we evaluate if this difference in correlation is statistically significant. Our analysis here is based on the SROCC values across all distortions.

Recall that we computed correlations for each of the algorithms over 1000 test sets. Thus, apart from the median score tabulated before, we have at our disposal the mean SROCC value and the standard error associated with the 1000 correlation values. In Fig. 13, we plot this mean correlation value across the dataset along with error bars one standard deviation wide for each of the algorithms evaluated in Table II.

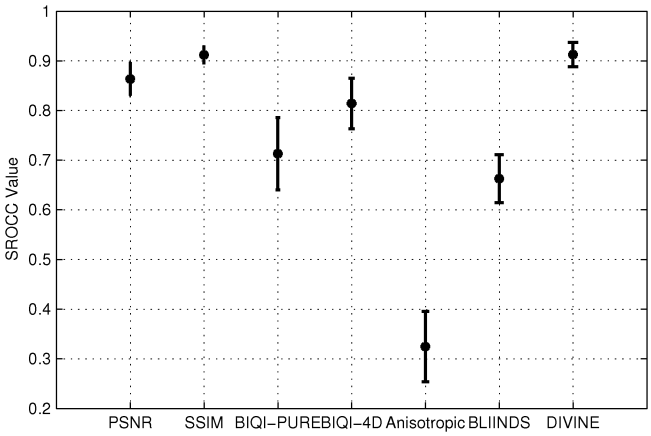

Fig. 13. Mean SROCC and error bars one standard deviation wide for the algorithms evaluated in Table II, across 1000 train-test trials on the LIVE IQA database.

In order to evaluate statistical significance, we utilize the one-sided t-test between the correlation scores generated by the algorithms across the 1000 train-test trials [60]. In Table VI, we tabulate the results of such statistical analysis. The null hypothesis is that the mean correlation of the row is equal to the mean correlation of the column at the 95% confidence level. The alternative hypothesis is that the mean correlation of the row is greater (or lesser) than the mean correlation of the column. Table VI indicates which row is statistically superior ("1"), statistically equivalent ("0"). or statistically inferior ("−1") to which column.

From Table VI, it is obvious that DIIVINE is statistically better than other no-reference approaches to IQA. Further, *DIIVINE is statistically superior to the full-reference PSNR*. This is a significant result indeed, for we are unaware of any NR IQA algorithm that is not only capable of assessing quality across many distortion categories, but also performs statistically better than the full-reference PSNR. Indeed, DIIVINE, which predicts perceived quality given ONLY the distorted image produces correlations with human subjective judgments at a level that is *statistically indistinguishable from the full-reference SSIM* that needs both the reference and distorted image in order to assess quality. This suggests that one can safely replace the FR SSIM with the NR DIIVINE without any loss in performance, provided that the distortions encountered are well represented by the dataset used to train DIIVINE (here—the LIVE IQA database).

C. Database Independence

Since NR IQA algorithms are generally trained and tested on various splits of a single dataset (as described above), it is natural to wonder if the trained set of parameters are database-specific. In order to demonstrate that the training process is simply a calibration, and once such training is performed, DIIVINE is capable of assessing the quality of any distorted image (from the

[3]The reader is referred to [12] for details on these realizations of the BIQI-framework.

[4]We note that in [30], the authors mention a correction for JPEG images, which we do not implement here. The variance parameter as suggested is used for NR IQA.

TABLE VI
RESULTS OF THE ONE-SIDED T-TEST PERFORMED BETWEEN SROCC VALUES. A VALUE OF "1" INDICATES THAT THE ALGORITHM (ROW) IS STATISTICALLY SUPERIOR TO THE ALGORITHM (COLUMN). A VALUE OF "0" INDICATES STATISTICAL EQUIVALENCE BETWEEN THE ROW AND COLUMN, WHILE A VALUE OF "−1" INDICATES THAT THE ALGORITHM (ROW) IS STATISTICALLY INFERIOR TO THE ALGORITHM (COLUMN). *ITALICIZED* ALGORITHMS ARE NR IQA ALGORITHMS, OTHERS ARE FR IQA ALGORITHMS

	PSNR	SSIM	*BIQI-PURE*	*BIQI-4D*	*Anisotropic IQA*	*BLIINDS*	*DIIVINE*
PSNR	0	-1	1	1	1	1	-1
SSIM	1	0	1	1	1	1	0
BIQI-PURE	*-1*	*-1*	*0*	*-1*	*1*	*1*	*-1*
BIQI-4D	*-1*	*-1*	*1*	*0*	*1*	*1*	*-1*
Anisotropic IQA	*-1*	*-1*	*-1*	*-1*	*0*	*-1*	*-1*
BLIINDS	*-1*	*-1*	*-1*	*-1*	*1*	*0*	*-1*
DIIVINE	*1*	*0*	*1*	*1*	*1*	*1*	*0*

TABLE VII
SPEARMAN'S RANK ORDERED CORRELATION COEFFICIENT (SROCC) ON THE TID2008 DATABASE. *ITALICIZED* ALGORITHMS ARE NR IQA ALGORITHMS, OTHERS ARE FR IQA ALGORITHMS

	JP2K	JPEG	WN	Gblur	All
PSNR	0.825	0.876	0.918	0.934	0.870
SSIM (SS)	0.963	0.935	0.817	0.960	0.902
DIIVINE	*0.924*	*0.866*	*0.851*	*0.862*	*0.889*

set of distortions it is trained for) we evaluate the performance of DIIVINE on an alternate database—the TID2008 [61].

The TID database consists of 25 reference images and 1700 distorted images over 17 distortion categories. Of these 25 reference images, only 24 are natural images and we test our algorithm only on these 24 images. Further, of the 17 distortion categories, we test DIIVINE only on those categories it has been trained for—JPEG, JPEG2000 compression (JP2k), additive white noise (WN), and Gaussian blur (blur).[5] In order to evaluate DIIVINE on the TID database, we train the parameters of DIIVINE using the entire LIVE IQA database as described previously. The trained model is then tested for its performance on the TID database. In Table VII, we tabulate the SROCC values obtained for such testing for each distortion as well as across distortion categories. Further, we also list the performance of the FR PSNR and SSIM for comparison purposes. It is clear from Table VII that the performance of DIIVINE is NOT database dependent and that once trained, DIIVINE is capable of assessing the quality of images across the distortions that it is trained for.

D. Computational Analysis

Although DIIVINE was not developed under the constraint of real-time analysis of images, given that the performance of DIIVINE is as good as leading FR QA algorithms, its computational complexity is relevant when one considers applications of DIIVINE. Hence, it is prudent to perform an informal analysis of the computations needed to predict the quality of an image without a reference using DIIVINE.

An unoptimized MATLAB code takes approximately 60 s to produce a quality estimate on a 1.8-GHz processor with 2 GB of RAM running Windows XP and MATLAB R2008a for a 512 $\times$ 768 image. The amount of time taken for training the SVM/SVRs is negligible as is the time taken to predict the quality by the trained classifier/regressors compared to that of feature extraction. In Table VIII, we tabulate the percentage of time devoted to each of the steps in DIIVINE.

[5]Although DIIVINE has been trained for FF, the JP2k transmission loss distortion on the TID database does not correspond to this kind of fading channel model and hence is not considered here.

TABLE VIII
INFORMAL COMPLEXITY ANALYSIS OF DIIVINE. TABULATED VALUES REFLECT THE PERCENTAGE OF TIME DEVOTED TO EACH OF THE STEPS IN DIIVINE

Step	Percentage of Time
Steerable Pyramid Decomposition	2.52
Divisive Normalization	10.83
Or. & Scale selective Statistics	0.10
Orientation selective Statistics	0.48
Across scale Correlations	8.42
Spatial Correlation	69.72
Across Orientation Statistics	7.92

As is clear from Table VIII, spatial correlation statistics occupy a considerable chunk of the processing time. This is primary because constructing the 2-D PDFs needed for various spatial shifts is a computationally intensive process. One would imagine that implementing this efficiently in compile-able code (such as C) would cut down the time needed considerably. Further, the steerable pyramid decomposition in this version of DIIVINE is performed using the MATLAB toolbox from the authors [36], without using MEX code as recommended. Given that there exists C code for the same, it is not wrong to suppose that the time take for this section may also be reduced drastically. Similar arguments hold for the divisive normalization process. The across-orientation statistics and the across scale correlations are based on windowed structural correlation computation, whose current implementation is in MATLAB. Recently, however, faster real-time implementations of such windowed correlations have been made available, which would reduce the computation associated with these steps as well [62].

Thus, it seems that DIIVINE can be re-coded efficiently in order to achieve close-to-real-time (if not real-time) performance. Thus, application of DIIVINE should not suffer owing to its complexity.

VI. CONCLUSION

We proposed a no-reference (NR)/blind IQA framework and integrated algorithm based on natural scene statistics, that assesses the quality of an image without need for a reference across a variety of distortion categories. This algorithm—the DIIVINE index—utilizes the previously proposed two-stage

framework which first identifies the distortion present in the image and then performs distortion-specific quality assessment to provide an ostensibly distortion-independent measure of perceptual quality, using extracted natural scene statistic features. We detailed the statistical features extracted, along with motivations drawn from vision science and image processing, and demonstrated that the DIIVINE index correlates well with human perception of quality. We undertook a thorough analysis of the proposed index on the publicly available LIVE IQA Database, and showed that the proposed measure is statistically superior to other NR IQA algorithms that function across distortion categories. Further, we compared the performance of DIIVINE with two standard full-reference QA algorithms: the PSNR and the single-scale SSIM. We showed that DIIVINE is *statistically superior* to the FR PSNR and *statistically indistinguishable* from the FR SSIM. To the best of our knowledge, DIIVINE is the only IQA algorithm that not only assesses quality across a range of distortions, but also correlates with human perception judgments at a level that is statistically equivalent to good FR measures of quality. Finally, we demonstrated that DIIVINE performance is database-independent and can easily be extended to distortions beyond those considered here, and performed an informal complexity analysis.

The proposed approach is modular, and can easily be extended beyond the set of distortions considered here. Importantly, DIIVINE does not compute specific distortion features (such as blocking), but instead extracts statistical features which lend themselves to a broad range of distortion measurements. Future work will involve increasing the subset of distortions beyond those considered here, in an effort to further relax any distortion dependence. A software release of DIIVINE has been made available online: http://live.ece.utexas.edu/research/quality/DIIVINE_release.zip.

REFERENCES

[1] A. C. Bovik and Z. Wang, *Modern Image Quality Assessment*. New York: Morgan and Claypool, 2006.

[2] K. Seshadrinathan and A. C. Bovik, "Video quality assessment," in *The Essential Guide to Video Processing*. New York: Academic, 2009.

[3] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Select. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[4] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[5] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[6] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

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

[8] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

[9] W. Geisler, "Visual perception and the statistical properties of natural scenes," *Ann. Rev. Neurosci.*, vol. 59, pp. 167–192, 2007.

[10] B. Olshausen and D. Field, "Natural image statistics and efficient coding," *Netw.: Computat. Neural Syst.*, vol. 7, pp. 333–339, 1996.

[11] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG 2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[12] 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.

[13] L. Meesters and J. B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Process.*, vol. 82, no. 3, pp. 369–387, 2002.

[14] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Process.: Image Commun.*, vol. 23, no. 4, pp. 257–268, 2008.

[15] J. Caviedes and S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," in *Proc. IEEE Int. Conf. Image Processing*, 2002, vol. 3, pp. 53–56.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

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

[18] R. Barland and A. Saadane, "A reference free quality metric for compressed images," in *Proc. 2nd Int. Workshop Video Processing and Quality Metrics for Consumer Electronics*, 2006.

[19] J. Chen, Y. Zhang, L. Liang, S. Ma, R. Wang, and W. Gao, "A no-reference blocking artifacts metric using selective gradient and plainness measures," in *Proc. 9th Pacific Rim Conf. Multimedia: Advances in Multimedia Information Processing*, 2008.

[20] S. Suthaharan, "No-reference visually significant blocking artifact metric for natural scene images," *Signal Process.*, vol. 89, no. 8, pp. 1647–1652, 2009.

[21] E. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, and L. Jiang, "No-reference JPEG-2000 image quality metric," in *Proc. Int. Conf. Multimedia and Expo*, 2003, vol. 1, pp. 6–9.

[22] H. Tong, M. Li, H. Zhang, and C. Zhang, "No-reference quality assessment for jpeg2000 compressed images," in *Proc. IEEE Int. Conf. Image Processing,* Citeseer, 2004, pp. 24–27.

[23] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process.: Image Commun.*, vol. 19, no. 2, pp. 163–172, 2004.

[24] S. Varadarajan and L. Karam, "An improved perception-based no-reference objective image sharpness metric using iterative edge refinement," in *Proc. 15th IEEE Int. Conf. Image Processing*, 2008, pp. 401–404.

[25] N. Narvekar and L. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Proc. 1st Int. Workshop Quality of Multimedia Experience (QoMEX)*, 2009.

[26] R. Ferzli and L. 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, p. 717, Apr. 2009.

[27] N. Sadaka, L. Karam, R. Ferzli, and G. Abousleman, "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," in *Proc. 15th IEEE Int. Conf. Image Processing*, 2008, pp. 369–372.

[28] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. 1st Int. Workshop Quality of Multimedia Experience (QoMEX)*, 2009.

[29] X. Li *et al.*, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Processing*, 2002, vol. 1, pp. 449–452.

[30] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Optic. Soc. Amer. A*, vol. 24, no. 12, pp. 42–51, 2007.

[31] M. A. Saad, A. C. Bovik, and C. Charrier, "A perceptual DCT statistics based blind image quality metric," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[32] M. Farias and S. Mitra, "No-reference video quality metric based on artifact measurements," in *Proc. IEEE Int. Conf. Image Processing*, 2005, vol. 3, pp. 141–144.

[33] J. Caviedes and J. Jung, "No-reference metric for a video quality control loop," in *Proc. 5th World Multiconf. Systemics, Cybernetics and Informatics*, 2001.

[34] M. Wainwright and E. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inf. Process. Syst.*, vol. 12, no. 1, pp. 855–861, 2000.

[35] 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.

[36] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[37] B. A. Olshausen and D. J. Field, "How close are we to understanding V1?," *Neural Computat.*, vol. 17, no. 8, pp. 1665–1699, 2005.

[38] R. Sekuler and R. Blake, *Perception*. New York: McGraw-Hill, 2002.

[39] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2385–2401, Oct. 2009.
[40] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization: Modeling nonlinearities and adaptation in cortical neurons," *Statist. Theories of the Brain*, pp. 203–222, 2002.
[41] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," in *Proc. 15th IEEE Int. Conf. Image Processing (ICIP 2008)*, 2008, pp. 1200–1203.
[42] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
[43] P. Teo and D. Heeger, "Perceptual image distortion," in *SID Int. Symp. Digest of Technical Papers*, 1994, vol. 25, pp. 209–209.
[44] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Tran. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
[45] A. K. Moorthy and A. C. Bovik, "Statistics of natural image distortions," in *Proc. 2010 IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 962–965.
[46] S. Palmer, *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press, 1999.
[47] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
[48] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
[49] M. J. Chen and A. C. Bovik, "No-reference image blur assessment using multiscale gradient," in *Proc. 1st Int. Workshop Quality of Multimedia Experience (QoMEX)*, 2009.
[50] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Select. Topics Signal Process., Special Issue on Visual Media Quality Assessment*, vol. 3, no. 2, pp. 193–201, Apr. 2009.
[51] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–313, Sep. 2004.
[52] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 2000.
[53] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural Computat.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[54] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
[55] C. Chang and C. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm/.
[56] B. Girod, "What's wrong with mean-squared error?," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.
[57] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?—A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[58] A. K. Moorthy and A. C. Bovik, "Perceptually significant spatial pooling techniques for image quality assessment," in *Proc. SPIE Human Vision and Electronic Imaging XIV*, Jan. 2009, vol. 7240.
[59] S. Gabarda and G. Cristobal, Nov. 2010. [Online]. Available: http://www.iv.optica.csic.es/page49/page51/page51.html.
[60] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL: CRC, 2004.
[61] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008—A database for evaluation of full reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30–45, 2009.
[62] M. J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," in *Proc. 2010 IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 994–997.

Anush Krishna Moorthy received the B.E. degree in electronics and telecommunication with a Silver Medal from the University of Pune, Pune, India, in June 2007 and the M.S. degree in electrical engineering from the University of Texas at Austin in May 2009, and he is currently pursuing the Ph.D. degree.

He joined the Laboratory for Image and Video Engineering (LIVE), University of Texas at Austin, in January 2008 and is currently the Assistant Director with LIVE, Department of Electrical and Computer Engineering, University of Texas. His research interests include image and video quality assessment, image and video compression, and computational vision.

Mr. Moorthy is the recipient of the Continuing Graduate Fellowship from The University of Texas at Austin for 2010-2011, the Professional Development Award, Fall 2009, Fall 2010, the Center for Perceptual Systems Travel Grant, Spring 2010, and the TATA scholarship for higher education abroad.

Alan Conrad Bovik (F'96) is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published about 600 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (New York: Academic, 2009).

Dr.. Bovik was named the SPIE/IS&T Imaging Scientist of the Year for 2011. He has also received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009); the Education Award (2007); the Technical Achievement Award (2005); and the Meritorious Service Award (1998). He received the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008), the IEEE Third Millennium Medal (2000), and two journal paper awards from the international Pattern Recognition Society (1988 and 1993). He is a Fellow of the Optical Society of America (OSA), a Fellow of the Society of Photo-Optical and Instrumentation Engineers (SPIE), and a Fellow of the American Institute of Medical and Biomedical Engineering (AIMBE). He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996–1998; co-founder and Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; Editorial Board, PROCEEDINGS OF THE IEEE, 1998–2004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003-present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, Texas, in November, 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.