No-reference image quality assessment based on log-derivative statistics of natural scenes

Yi Zhang
Damon M. Chandler

No-reference image quality assessment based on log-derivative statistics of natural scenes

Yi Zhang
Damon M. Chandler
Oklahoma State University
School of Electrical and Computer Engineering
Laboratory of Computational Perception and Image Quality
Stillwater, Oklahoma, 74078
E-mail: yi.zhang@okstate.edu

Abstract. *We propose an efficient blind/no-reference image quality assessment algorithm using a log-derivative statistical model of natural scenes. Our method, called DErivative Statistics-based QUality Evaluator (DESIQUE), extracts image quality-related statistical features at two image scales in both the spatial and frequency domains. In the spatial domain, normalized pixel values of an image are modeled in two ways: pointwise-based statistics for single pixel values and pairwise-based log-derivative statistics for the relationship of pixel pairs. In the frequency domain, log-Gabor filters are used to extract the fine scales of the image, which are also modeled by the log-derivative statistics. All of these statistics can be fitted by a generalized Gaussian distribution model, and the estimated parameters are fed into combined frameworks to estimate image quality. We train our models on the LIVE database by using optimized support vector machine learning. Experiment results tested on other databases show that the proposed algorithm not only yields a substantial improvement in predictive performance as compared to other state-of-the-art no-reference image quality assessment methods, but also maintains a high computational efficiency.* © 2013 SPIE and IS&T [DOI: 10.1117/1.JEI.22.4.043025]

1 Introduction

Rapid advances in digital image processing technology have revolutionized the way in which images are captured, stored, transmitted, and accessed. A crucial task for any image processing system is the ability to assess image quality in a manner that is consistent with human subjective judgment. To address this need, numerous image quality assessment (IQA) algorithms have been proposed over the past several decades, using a wide variety of image analysis techniques. Based on the availability of a reference image, these algorithms can be divided into three main categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) IQA.

In this paper, we address the task of NR IQA, in which no reference image is available to the IQA algorithm. Although humans can easily assess the quality of a distorted image in an NR setting, this task remains extremely challenging for an algorithm. The vast majority of NR IQA algorithms have been designed for specific distortion types. These distortion-specific methods often assume that a particular distortion

type is known, such as additive white noise (WN) or Gaussian blur (Gblur) (e.g., Refs. 1 to 4), or JPEG/ JPEG2000 compression (e.g., Refs. 5 to 8). Then, based on the assumed distortion type, specific distortion-related features are extracted to assess image quality.

Another class of NR IQA methods has been recently developed which does not require knowledge of the distortion type, but instead evaluates image quality assuming that the distorted image shares some properties with those in the training database. Machine-learning techniques are used to develop a mapping from the input features to the target quality scores in the training database, and then the quality of the distorted image is estimated based on the mapping. Such non-distortion-specific approaches usually follow (1) a learning-based approach; (2) a natural scene statistics-based (NSS-based) approach (e.g., Refs. 9 to 11) usually rely on a large number of quality-related features that are designed based on the physical properties of image distortions. Assuming that image quality can be measured by various features (e.g., wavelet coefficient magnitude and phase, texture statistics), and assuming that certain features that are effective for one particular distortion type may not be applicable to others, various regression models are learned to combine these features such that high performance is always achieved for a variety of distortions. In comparison, the NSS-based NR IQA approaches (e.g., Refs. 12 to 15) also extract quality-related features, but these features are based on a statistical representation of the image's coefficients in the original/transformed domain under certain distribution models. Quality is estimated based on the extent to which these statistics match those of natural images. In Sec. 2, we provide a more detailed review of existing NR IQA algorithms.

Despite the difference of the statistical models and image transforms employed, these NSS-based NR IQA algorithms share a common thread: they estimate image quality based on single-domain properties. However, we argue that information from both the spatial domain and the frequency domain can play important roles in representing distortions. For instance, it is well known that a blurred image will lose some sharp edges in the spatial domain, which means a reduction in the variance of edge pixel values throughout the image, while in the frequency domain, blurring results in a reduction of the high-frequency components. As we have previously

demonstrated in Ref. 16, the quality of a blurred image can be estimated by utilizing both the spatial and spectral properties. We argue that this dual-domain approach can be useful for general NR IQA.

In this paper, we present an NSS-based non-distortion-specific approach called derivative statistics-based quality evaluator (DESIQUE), which builds upon BRISQUE[15] by utilizing statistical features measured in both the spatial and spectral domains. Our work draws on the concept of log-derivative statistics proposed by Huang and Mumford,[17] and on the concepts proposed by Ruderman.[18] We employ combined frameworks consisting of both one- and two-stage frameworks that are commonly used in NSS-based NR IQA algorithms to predict image quality.

For the spatial domain analysis, images are first decorrelated through local mean subtraction and divisive normalization[18] to obtain normalized pixel values [also called mean-subtracted contrast-normalized (MSCN) coefficients]. Then, these normalized pixel values are statistically modeled in two ways: (1) pointwise-based statistics for single pixel values (following Ref. 15) and (2) pairwise-based log-derivative statistics for the relationship of pixel pairs. Figure 1 gives an example of three blurred images [Fig. 1(a)] and their computed histogram profiles [Fig. 1(b)] based on the two above-mentioned statistics. As shown in the figure, different distortion levels give rise to different histogram shapes in the spatial domain. *Image3* is the most blurred, and thus we see its spatial-domain histograms are most narrow at the base and have sharp peaks. In comparison, the histograms for *image1* appear to be less peaked, a result of the much sharper content in edge areas. *Image2* is somewhat in between: it is clearly sharper than *image3*, but more blurred than *image1*, and thus its histogram's shape is somewhere in between the other two. This figure and other corresponding histograms for another distortion type shown later (see Sec. 3) demonstrate that statistics in the spatial domain can effectively indicate different amounts of distortions.

However, although the spatial-domain-based statistics can be a potential measure of image quality, they do not take into account spectral properties, which can also be a good indicator of perceived distortion. For example, two images can have very similar statistics in the spatial domain but appear to be of different visual quality. We demonstrate this assertion in Fig. 2, which shows two images corrupted by different amounts of high-frequency noise. As shown in the figure, the computed histograms of the spatial-domain-based log-derivative statistics appear to be quite similar. However, when applying such statistics to the high-frequency components of the image in the frequency domain, the resulting histograms can be distinct.

To obtain the spectral properties contained in an image, we utilize log-Gabor filters. Numerous studies have shown that the response properties of simple cells in primary visual cortex can be well-modeled by Gabor filters,[21,22] which offer simultaneous localization of spatial and frequency information. Log-Gabor filters overcome the frequency distribution problem encountered by the classical Gabor filters by maintaining Gaussian passbands on a logarithmic frequency scale.[23] In our implementation, we decompose an image into several spatial frequency bands at two orientations (horizontal and vertical) and compute the log-derivative statistics of the high-spatial frequency log-Gabor

coefficients. As shown in Figs. 1 and 2, the histograms computed in both domains can be fitted by a two-parameter generalized Gaussian distribution (GGD), which has been demonstrated in Ref. 24 to be an effective model of the subband statistics of natural images for RR IQA. The parameters of these GGD fits are utilized as the DESIQUE features, which are then fed into combined frameworks to estimate the image's quality.

Although DESIQUE is built upon BRISQUE, there are key differences between the two algorithms. First, the two algorithms use different types of analyses. As opposed to the paired-product-based analysis used in BRISQUE, the log-derivative-based analysis used in DESIQUE provides the latter with (1) greater sensitivity to local contrast changes and (2) distributions that are symmetric and more easily modeled. To demonstrate these facts, Fig. 3 shows three images [Fig. 3(a)] with different contrast, as well as their corresponding paired product and log-derivative coefficient histograms in four orientations [Fig. 3(b)]. As shown in this figure, the computed histograms of log-derivative values in the spatial domain yield profiles that can be used to distinguish between the three images, whereas the histograms of the paired-product values do not provide such discriminability. Also, as shown in Fig. 3, the MSCN coefficients have a symmetric distribution when modeled by log-derivative statistics, whereas the paired product statistics give rise to asymmetric distributions. Thus, when using log-derivatives, fewer parameters are required for the histogram fitting. This reduction in parameters (and thus in the number of features) allows for the computational resources to be devoted to the frequency-domain analysis to improve the predictive performance without penalizing the computational (runtime) performance.

Second, DESIQUE uses a two-domain approach, whereas BRISQUE uses a single-domain approach. As mentioned previously, BRISQUE operates in the spatial domain only, while we have shown in Figs. 1 and 2 that both the spatial and spectral properties can be important indicators of perceived distortion. It is well known that some distortions are more easily detected in the frequency domain rather than in the spatial domain and vice versa. For example, the fast-fading distortion in the LIVE database is more detectable in the spatial domain, whereas different types of noise are more detectable in the frequency domain. As we will demonstrate, by adopting a dual-domain approach, the proposed DESIQUE algorithm can achieve not only better performance on various distortions across different databases, but also relatively higher robustness to images distorted by different types of noise (see Sec. 4.4).

The main contributions of this work are as follows: (1) DESIQUE analyzes distorted images in both the spatial and frequency domains by extracting NSS-based quality-related features, and the ultimate quality score is predicted based on combined frameworks. We show in this work that distortions affect image statistics in both domains and both can play important complementary roles in assessing image quality. (2) We propose to use the seven types of log-derivative statistics to model the relationship between neighboring pixel/coefficient values, which are sensitive to distortion. All of these statistics have symmetrical distributions and can be characterized by using a two-parameter GGD model. (3) We also show that DESIQUE can be
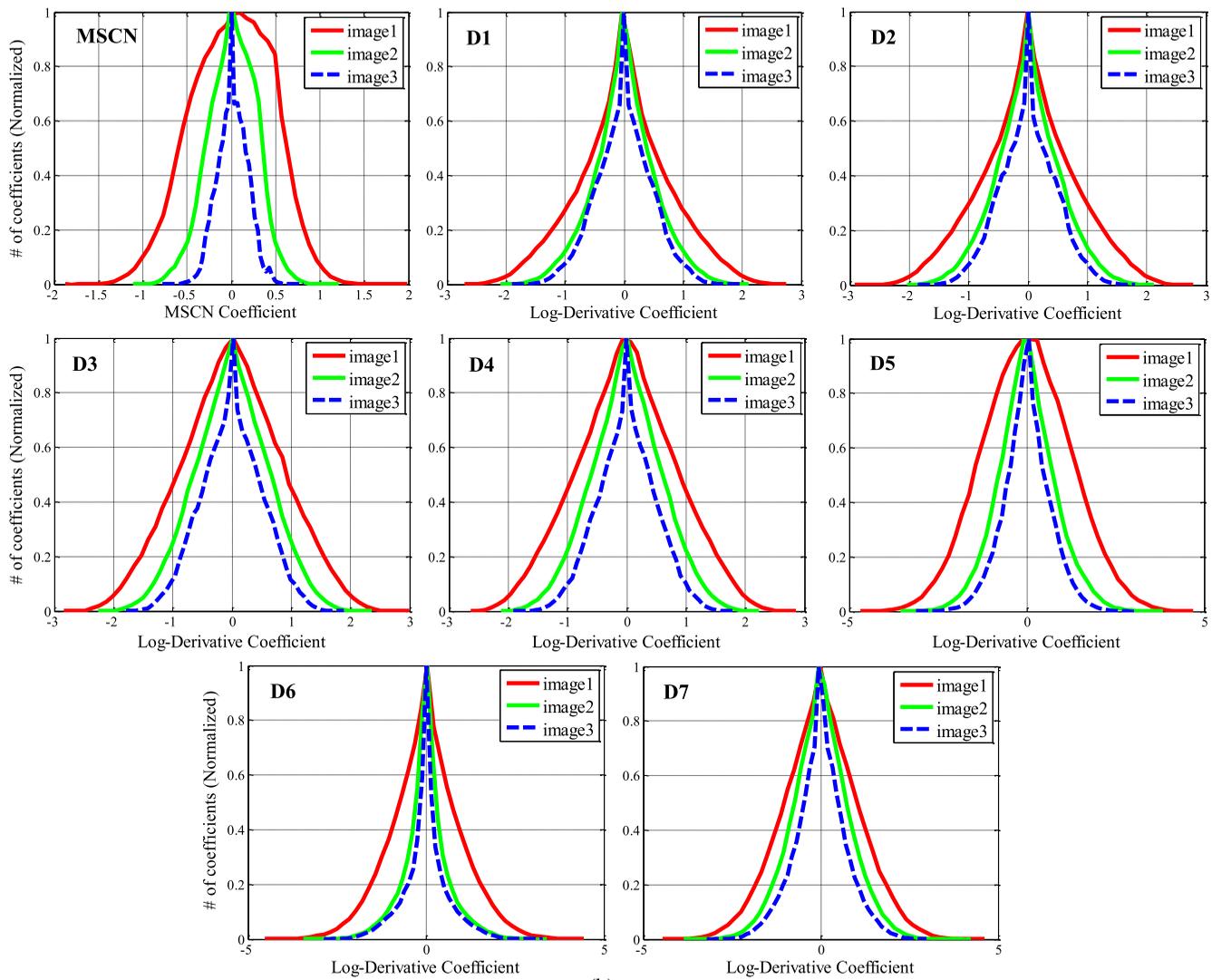
Fig. 1 Three blurred images from the LIVE database[19] with different histogram profiles in the spatial domain. D1 through D7 denote the log-derivatives computed at six orientations with various spatial contexts (see Sec. 3.2.1 for details about computing the spatial-domain-based statistics).
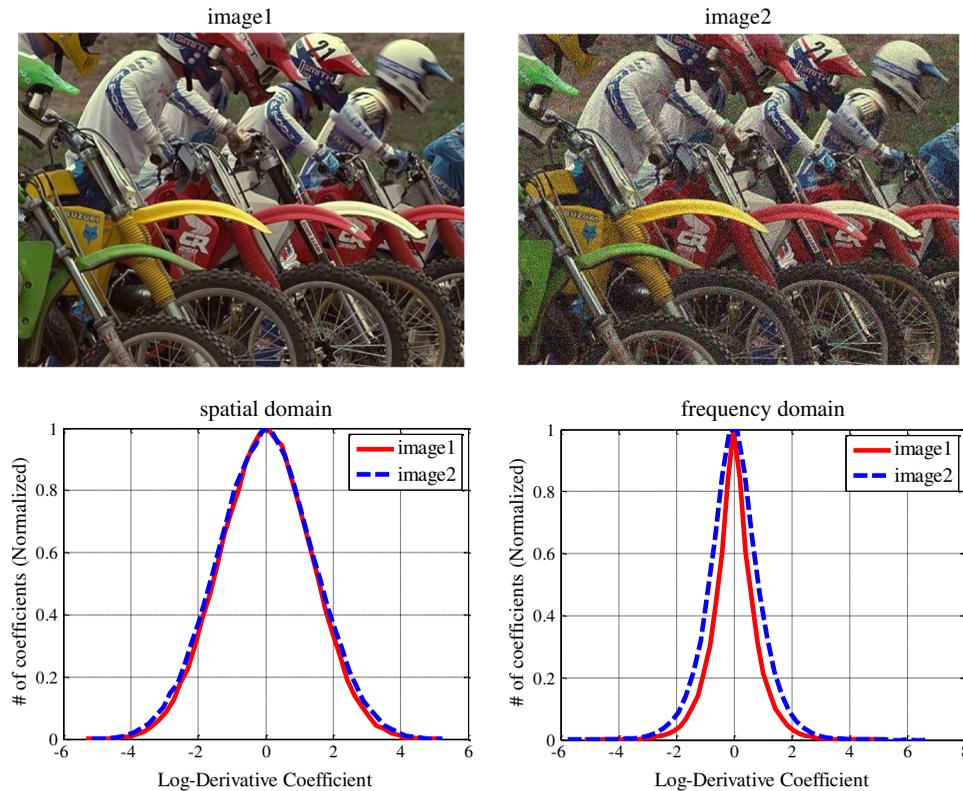
Fig. 2 Two high-frequency noise-corrupted images from the TID database[20] with different visual qualities. Notice that the spatial-domain-based log-derivative statistics have similar histogram profiles, but the histogram shape varies in the frequency domain. (These histograms were generated from measure D6; see Sec. 3.2.2 for details about computing the frequency-domain-based statistics.)

adaptive to images distorted by other types of noise on which it was not trained. Although DESIQUE was trained on additive white Gaussian noise, it still performs well on other types of noise such as additive color noise, spatially correlated noise, and high-frequency noise. (4) DESIQUE not only demonstrates good predictive performance, but also maintains significant computational efficiency; the algorithm can achieve near-real-time performance.

This paper is organized as follows. In Sec. 2, we give a brief review of the previous work on NR IQA. In Sec. 3, we provide details of the DESIQUE algorithm. In Sec. 4, we present results of the proposed algorithm on different image databases and evaluate the performance of different algorithms. General conclusions are provided in Sec. 5.

2 Previous Work

Current NR IQA algorithms can generally be divided into two categories: (1) those which assume that the distortion type is known, only determining the severity and (2) those which do not consider the prior knowledge of distortion type, but assume that the distortion information of the tested image can be similar to the training database. In this section, we provide a brief review of these NR IQA methods.

2.1 Distortion-Specific NR IQA Algorithms
2.1.1 JPEG IQA

JPEG artifacts mainly contain blocking, blurring, and aliasing due to the quantization of discrete cosine transform (DCT) coefficients. Therefore, the general approach to JPEG

IQA is to measure the blocking artifacts or edge strength at block boundaries to estimate image quality.

In Ref. 5, the blocky image is modeled as a non-blocky image distorted by a pure blocky signal, and then the the blocking artifact is measured by detecting and estimating the power of the blocky signal. In Ref. 6, blocking artifacts are measured in DCT domain by modeling the blocking artifacts as two-dimensional (2-D) step functions. This information is then used with a human-visual-system (HVS)-based measurement of blocking artifact severity to esitimate quality. In Ref. 26, image blockiness is measured by detecting the low-amplitude step edges and estimating the edge parameters using a Hermite transform. In Ref. 7, image blockiness is measured based on the average differences across block boundaries and the activity of the image signal. The activity is measured using two factors: (1) the average absolute difference between in-block image samples and (2) the zero-crossing rate. Other JPEG-based NR IQA algorithms have used the DCT (e.g., Refs. 27 and 28), discrete Fourier transform (DFT) (e.g., Ref. 29), Sobel operator (e.g., Refs. 30 and 31), selected gradient and uniformity measures (e.g., Ref. 32), HVS models (e.g., Ref. 33), and machine learning (e.g., Ref. 34). In summary, each algorithm measures blocking artifacts in JPEG compressed images by using either the transformed domain properties or the fundamental characteristics introduced by quantization to the signal.

2.1.2 JPEG2000 IQA

For JPEG2000-compressed images, the predominant artifacts introduced are blurring and ringing. Algorithms for
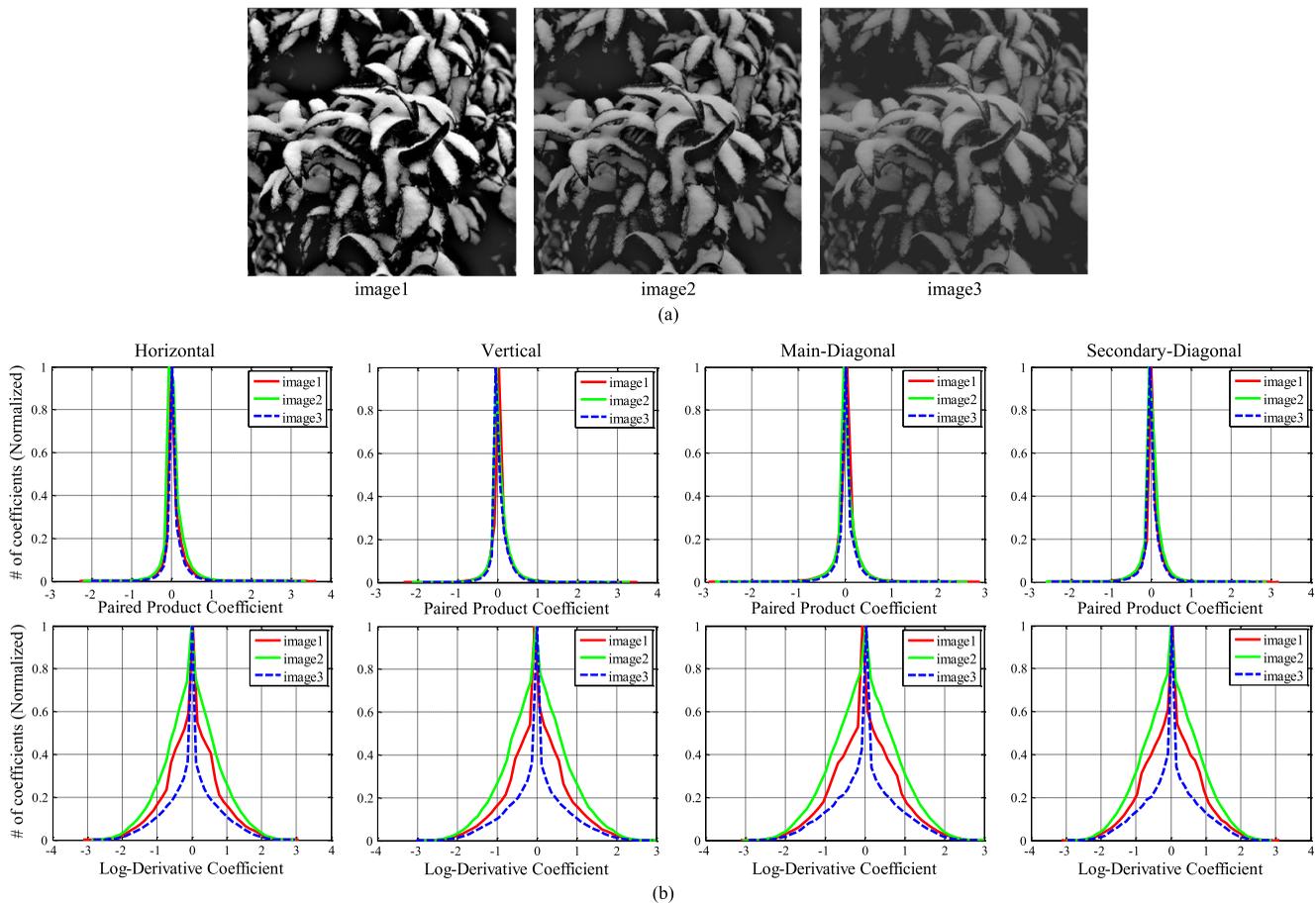
Fig. 3 Three images with different contrast and their corresponding paired product/log-derivative coefficient histograms in four orientations. Note that image3 is from the CSIQ database;[25] image1 and image2 are generated by adjusting the local contrast in image3 using Adobe Photoshop CS3 software.

assessing quality of this type of image are mainly focusing on measuring the amount of blurring or edge spreading by using edge-detection techniques. For example, Li et al.[35] proposed a principal component analysis–based method to assess the quality of JPEG2000 image. First, by viewing all edge points in an image as distorted or undistorted, local features are extracted at each of the edge points, which indicate both blurring and ringing. Then, a model is built to describe the relationship between these local features and a local distortion metric through the probability of the edge point being distorted versus undistorted. The image's quality is estimated based on these edge points (a similar method can be found in Ref. 36). Sazzad et al.[37,38] proposed to use pixel distortion and edge information for the quality evaluation of JPEG2000 images. Their algorithms assume that the HVS is very sensitive to edge information and any kinds of artifacts that create pixel distortion from neighborhood pixels.

Other methods for NR JPEG2000 IQA may utilize certain kinds of transform such as the wavelet transform or DCT. Statistical regularizations of the coefficients are then used to assess image quality. For instance, Sheikh et al.[8] found that wavelet coefficient distributions of JPEG2000 images provide a good indication of the loss of visual quality. Thus, by computing features from all wavelet subbands and applying a nonlinear combination, a weighted average of the transformed features is used for quality prediction.

Zhou et al.[39] proposed to evaluate JPEG2000 images based on three steps: (1) image is divided into blocks, among which textured blocks are employed for quality prediction; (2) projections of wavelet coefficients between adjacent scales with the same orientation are utilized to measure the positional similarity; and (3) a general regression neural network is adopted to conduct quality prediction according to features from (1) and (2). Zhang et al.[40] utilized kurtosis in the DCT domain to blindly appraise quality of JPEG2000 images based on three measurements: frequency band-based one-dimensional (1-D) kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis. The method is simple (without edges/features extraction), parameter free, and correlates well with subjective quality scores.

2.1.3 Blur/noise IQA

Researchers have also developed various methods to measure image quality when other distortion types and combinations are present. Since the most common distortion types in images are noise and blur (due, e.g., to the transmission, storage, and reconstruction), numerous algorithms have been proposed to estimate noise degradation and blurring. For instance, Corner et al.[41] developed a noise estimation technique using data masking. Li et al.[1] proposed an NR IQA technique by joining three different NR measures for three different image distortion types (blur, noise, and

block/ringing artifacts). Brandao and Queluz[3] proposed an NR IQA algorithm for estimating quantization noise introduced by lossy encoding such as JPEG or MPEG based on the statistics of DCT coefficients. Cohen and Yitzhaky[4] presented an NR method to identify noise and blur by assuming that common statistics of natural images can be obtained from their power spectra. Vu and Chandler[16] proposed a spectral and spatial measure of local perceived sharpness in natural images by using the slope of the magnitude spectrum and the total spatial variation.

2.2 Non-Distortion-Specific NR IQA Algorithms

2.2.1 Learning-based IQA

Learning-based NR IQA methods estimate quality by learning a mapping from a high-dimensional feature space to a scalar quality value. The feature space contains a number of quality-related features that are designed to capture relevant factors that affect image quality. Different techniques on feature extraction and regression lead to various algorithms.

Tong et al.[42] proposed a learning-based method that estimates quality by learning from training examples that contain both the high- and low-quality classes. A binary classifier is built and the quality metric of the tested image is denoted by the extent to which it belongs to these two classes. Tang et al.[9] proposed a learning-based blind image quality measurement that first extracts low-level image features from natural image measure and texture statistics. Then, via a machine-learning framework, the algorithm learns a mapping from these features to the subjective quality score. Li et al.[43] developed an NR IQA algorithm based on a general regression neural network (GRNN). The features extracted include the mean value of a phase congruency image, the entropy of the phase congruency image, the entropy of the distorted image, and the gradient of the distorted image. Quality is estimated by learning the functional relationship between these features and subjective mean opinion scores using the GRNN. In Refs. 10 and 11, Ye et al. used the visual codebook to quantize Gabor-filter-based features and then learned a mapping from the quantized feature space to image quality scores using either an example-based method or a support vector machine (SVM). This method extracts patch-level features and has the potential to be used in real-time applications with parallel implementation.

2.2.2 NSS-based IQA

NSS-based NR IQA methods assume that some statistical properties of natural scenes will remain across different reference images, but will change significantly in the presence of distortions. Thus, by measuring how the statistical properties of a distorted image deviate from those of a typical natural image, the distortion types as well as the image quality can be estimated. Typical NSS-based NR algorithms include BIQI,[12] BLIINDS-II,[14] DIIVINE,[13] BRISQUE,[15] and NIQE.[44]

BIQI[12] is a wavelet-based NR IQA method that extracts image features by modeling the wavelet subband coefficients based on a two-parameter GGD model. It utilizes a discrete wavelet transform with three scales and three orientations using Daubechies 9/7 wavelet basis. A total of 18 wavelet-based statistical features are extracted for each image.

BLIINDS-II[14] measures image quality based on the statistics of DCT coefficients of local image patches via two stages. In the first stage, non-overlapping small image blocks are prepared and statistical features are extracted. In the second stage, a multivariate GGD model is used to perform quality prediction. In order to perform prediction in the second stage, BLIINDS-II requires training, during which a Bayesian probabilistic model is obtained containing all information in the training database.

DIIVINE[13] operates based on a steerable pyramid transform, from which the coefficient statistics of the distorted image are used as the feature values for IQA. The algorithm employs a two-stage framework: distortion identification followed by distortion-specific quality assessment, and 88 features are extracted to perform the IQA task through SVM learning.

BRISQUE[15] is another NSS-based NR IQA method that operates in the spatial domain. The underlying features derive from the empirical distribution of the local normalized luminance values, as well as their neighboring products under a spatial NSS model. No required transformation and a small feature number allow the algorithm to run almost in real-time. Despite its simplicity, the algorithm represents the currently best-performing NSS-based NR IQA method.

The recently developed NIQE[44] algorithm is the first NSS-based approach that does not require training on the distorted images. The algorithm extracts local image features based on the work in Ref. 15, and the quality of a given test image is expressed as the distance between the multivariate Gaussian fit of the NSS features extracted from the test image and from the corpus of natural images. Although NIQE does not perform as well as DIIVINE or BRISQUE, it demonstrates the possibility to extend current NSS-based NR IQA to a no-distortion-trained approach.

3 Algorithm

The proposed DESIQUE algorithm is based on the assumption that log-derivative-based statistical properties of natural images in both the spatial and frequency domains are generally consistent across reference images, but vary significantly in the presence of distortions. In this way, it is possible to measure deviations in these expected (natural) statistical features as proxy measures of quality degradation. DESIQUE extracts log-derivative-based statistical features in two domains using seven derivative types and employs combined frameworks consisting of one used in BRISQUE (one-stage framework) and the other one in DIIVINE (two-stage framework).

3.1 Log-Derivative Statistics

The derivative statistics of natural images were first studied in Ref. 17, in which the differences of gray-level values between one pixel and its neighboring pixels were considered as the derivatives. Motivated by the work in Refs. 45 and 15, derivatives between pairs of pixels can have six orientations: horizontal (H), vertical (V), main-diagonal (MD), secondary-diagonal (SD), horizontal-vertical (HV), and combined-diagonal (CD). Based on the six orientations and the spatial location of pixel values, we propose seven types of derivatives defined as

$$\text{H}: \nabla_x I(i,j) = I(i,j+1) - I(i,j), \qquad (1)$$

$$V : \nabla_y I(i,j) = I(i+1,j) - I(i,j), \quad (2)$$

$$MD : \nabla_{xy} I(i,j) = I(i+1,j+1) - I(i,j), \quad (3)$$

$$SD : \nabla_{yx} I(i,j) = I(i+1,j-1) - I(i,j), \quad (4)$$

$$HV : \nabla_x \nabla_y I(i,j) = I(i-1,j) + I(i+1,j) - I(i,j-1)$$
$$- I(i,j+1), \quad (5)$$

$$CD_1 : \nabla_{cx} \nabla_{cy} I(i,j)_1 = I(i,j) + I(i+1,j+1) - I(i,j+1)$$
$$- I(i+1,j), \quad (6)$$

$$CD_2 : \nabla_{cx} \nabla_{cy} I(i,j)_2 = I(i-1,j-1) + I(i+1,j+1)$$
$$- I(i-1,j+1) - I(i+1,j-1), \quad (7)$$

where $I(i,j)$ represents the pixel value (or coefficient) corresponding to spatial location i, j. Note that Eqs. (1) to (4) and Eq. (6) are defined on neighboring pixels/pixel pairs and Eqs. (5) and (7) are defined on a patch size of 3×3.

In order to efficiently model natural images/image subbands by using the distribution of derivative statistics, we compute the logarithm of each pixel's value to create new images/image subbands J by

$$J(i,j) = \log[I(i,j) + K], \quad (8)$$

where K is a small constant that prevents $I(i,j)$ from being zero. Thus, based on Eqs. (1) to (8), we define seven types of log-derivatives as

$$D1 : \nabla_x J(i,j) = J(i,j+1) - J(i,j), \quad (9)$$

$$D2 : \nabla_y J(i,j) = J(i+1,j) - J(i,j), \quad (10)$$

$$D3 : \nabla_{xy} J(i,j) = J(i+1,j+1) - J(i,j), \quad (11)$$

$$D4 : \nabla_{yx} J(i,j) = J(i+1,j-1) - J(i,j), \quad (12)$$

$$D5 : \nabla_x \nabla_y = J(i-1,j) + J(i+1,j) - J(i,j-1)$$
$$- J(i,j+1), \quad (13)$$

$$D6 : \nabla_{cx} \nabla_{cy} J(i,j)_1 = J(i,j) + J(i+1,j+1)$$
$$- J(i,j+1) - J(i+1,j), \quad (14)$$

$$D7 : \nabla_{cx} \nabla_{cy} J(i,j)_2 = J(i-1,j-1) + J(i+1,j+1)$$
$$- J(i-1,j+1) - J(i+1,j-1). \quad (15)$$

In the following sections, we will show that histograms computed from these seven types of log-derivative statistics are effective in modeling natural images; they exhibit consistent profiles across different scenes but vary significantly in the presence of distortions. We, thus, extract statistical features that capture changes in these profiles.

3.2 Log-Derivative Statistics–Based Features

As mentioned in Sec. 2, recent work has focused on modeling natural-scene statistics either in the spatial domain[15] or in a transform domain (see, e.g., Refs. 12 to 14). However, we believe that the visual quality of an image can be influenced by the statistical information contained in both the spatial and frequency domains. Thus, the proposed DESIQUE features consist of two parts: the spatial domain features and the frequency domain features, and we will demonstrate that these statistical features are generally consistent across different images but change in the presence of distortion. Figure 4 shows a block diagram illustrating how DESIQUE features are extracted in these two domains.

3.2.1 Modeling log-derivative statistics in the spatial domain

The features extracted in the spatial domain consist of two types: (1) pointwise-based statistics for single pixel values (following Ref. 15) and (2) pairwise-based log-derivative statistics for the relationship of pixel pairs. Specifically, given an image $I(i,j)$, we first compute locally normalized pixel values via local mean subtraction and divisive normalization[18] defined as

$$\hat{I} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \quad (16)$$

where $i \in 1, 2, \cdots, M$, $j \in 1, 2, \cdots, N$ are spatial indices; M, N are the image height and width, respectively; and $C = 1$ is a constant that prevents division by zero. The quantities $\mu(i,j)$ and $\sigma(i,j)$ are defined as

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j), \quad (17)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} [I_{k,l}(i,j) - \mu(i,j)]^2}, \quad (18)$$
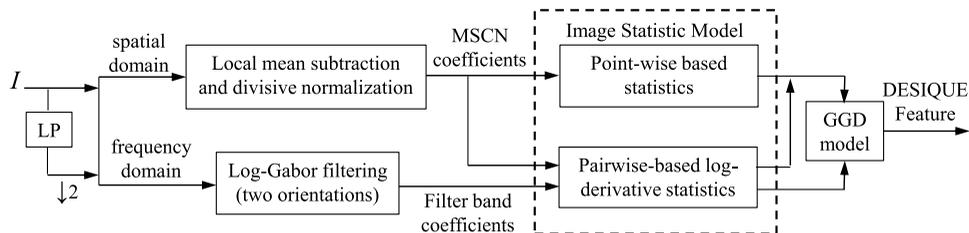
Fig. 4 A block diagram of DESIQUE feature extraction. Note that the filter band coefficients will only contain the first scale of the log-Gabor filter subbands corresponding to the high-frequency components of the image. LP denotes low-pass filter.

where $\omega = \{\omega_{k,l} | k = -K, \cdots, K, l = -L, \cdots, L\}$ is a 2-D circularly symmetric Gaussian weighting function sampled out to three standard deviations and rescaled to unit volume. As in Ref. 15, we also define $K = L = 3$.

According to Ref. 15, the MSCN coefficients $\hat{I}(i,j)$ can be modeled by a zero-mean GGD given by

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp[-(|x|/\beta)^\alpha], \qquad (19)$$

where $\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$ and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt, x > 0$ is the gamma function. The parameter α controls the general shape of the distribution and σ controls the variance. We estimate the two-parameter GGD model by using the moment-matching-based approach proposed in Ref. 46, and these two values form the first set of features in the spatial domain that will be used to capture image distortion, denoted by α_0 and σ_0.

The other set of features we extract in the spatial domain are formed by modeling the relationship of neighboring MSCN coefficient pairs based on the log-derivative statistics. In Ref. 15, statistical relationships between neighboring pixels are modeled by using the empirical distributions of pairwise products of neighboring MSCN coefficients along four orientations. These paired products can be parameterized by an asymmetric GGD.

In our implementation, we model the relationship between two adjacent MSCN coefficients by using the seven types of log-derivatives previously defined by Eqs. (9) to (15). Here $J(i,j) = \ln(|\hat{I}(i,j)| + K)$ and $K = 0.1$ is a constant that prevents $\hat{I}(i,j)$ from being zero. Under the Gaussian coefficient model, and assuming that the MSCN coefficients are zero mean and unit variance, these log-derivative values also follow a generalized Gaussian distribution and thus their parameters (α, σ) can be estimated by using the method proposed in Ref. 46. These 14 parameters form the second part of DESIQUE features in the spatial
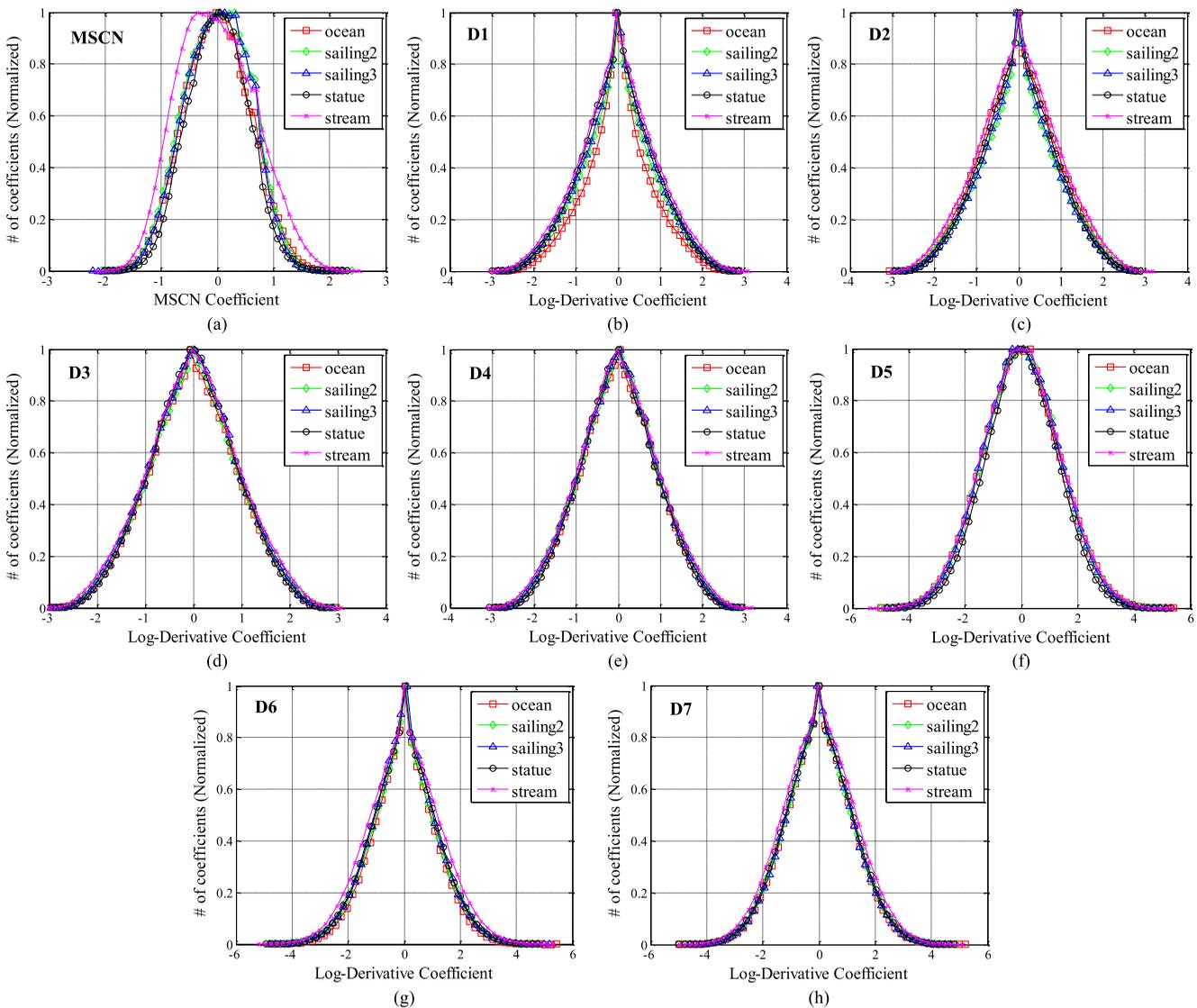
Fig. 5 Histogram of mean-subtracted contrast-normalized (MSCN) coefficients (a) and the seven types of log-derivative statistics [(b) to (h)] for each of the five reference images (shown in Fig. 6). Notice that the histograms exhibit consistent profiles that are largely independent of the particular reference image from which log-derivative statistics are computed.

domain. In all, we extract 16 spatial domain features at one image scale denoted by

$$\mathbf{f}_{\text{spatial}} = [\alpha_0, \alpha_{D_1}, \alpha_{D_2}, \alpha_{D_3}, \alpha_{D_4}, \alpha_{D_5}, \alpha_{D_6}, \alpha_{D_7},$$
$$\sigma_0, \sigma_{D_1}, \sigma_{D_2}, \sigma_{D_3}, \sigma_{D_4}, \sigma_{D_5}, \sigma_{D_6}, \sigma_{D_7}].$$

Here, D_1 through D_7 denote seven types of log-derivatives. In order to visualize how the aforementioned two statistics of MSCN coefficients in the spatial domain vary as a function of distortions, Fig. 5 plots their corresponding histograms for five reference images (shown in Fig. 6), and Fig. 7 plots the histograms for five distorted versions of one reference image (image *sailing2*, shown in Fig. 8). Notice that both the pointwise-based statistics and the pairwise-based log-derivative statistics of MSCN coefficients exhibit almost consistent profiles; they are largely independent of the particular reference image but change significantly in the presence of distortions.

3.2.2 Modeling log-derivative statistics in the frequency domain

Most distortion types have a pronounced effect on the higher-frequency components, resulting in considerable magnitude differences in the high-frequency band. To capture this change, we first decompose an image by using a log-Gabor filter defined by

$$G_{s,o}(\omega, \theta) = \exp\left\{ -\frac{[\log(\omega/\omega_s)]^2}{2[\log(\sigma_s/\omega_s)]^2} \right\} \times \exp\left[-\frac{(\theta - \mu_o)^2}{2\sigma_o^2} \right],$$
(20)

where $G_{s,o}$ is the log-Gabor filter denoted by spatial scale index s and orientation index o. The parameter ω is the normalized radial frequency and θ is the orientation. The decomposition is computed at the highest frequency scale ($s = 1$) and two orientations ($o = 1, 2$). We set the center frequency of that scale to be 1/3, corresponding to $\omega_1 = 2/3$. The value of σ_1/ω_1, which determines the bandwidth, is set to be 0.975. These values result in an approximately 1.5 octave bandwidth. Two orientations are computed using $\mu_o = (o - 1)\pi/2$, resulting in orientations of 0 and 90 deg.

Next, we apply the log-derivative statistics to model the log-Gabor filter subband coefficients. Note that we only apply six types of them corresponding to neighboring pixels/pixel pairs previously defined in Eqs. (9) to (12) and (14) to (15). Here, $J(i, j) = \ln[|g(i, j)| + K]$, where $|g(i, j)|$ represents the coefficient magnitude and $K = 0.1$ is a constant that prevents $g(i, j)$ from being zero. Again, we fit the histogram of log-derivative statistics for each of these subbands with a generalized Gaussian distribution model. The parameters (α, σ) are estimated and yield 24 DESIQUE features (2 subbands × 6 derivative types × 2 parameters/derivative type) extracted in the frequency domain at the original scale.

$$\mathbf{f}_{\text{frequency}} = [\alpha_{D_1,0^\circ}, \alpha_{D_2,0^\circ}, \alpha_{D_3,0^\circ}, \alpha_{D_4,0^\circ}, \alpha_{D_6,0^\circ}, \alpha_{D_7,0^\circ},$$
$$\sigma_{D_1,0^\circ}, \sigma_{D_2,0^\circ}, \sigma_{D_3,0^\circ}, \sigma_{D_4,0^\circ}, \sigma_{D_6,0^\circ}, \sigma_{D_7,0^\circ},$$
$$\alpha_{D_1,90^\circ}, \alpha_{D_2,90^\circ}, \alpha_{D_3,90^\circ}, \alpha_{D_4,90^\circ}, \alpha_{D_6,90^\circ}, \alpha_{D_7,90^\circ},$$
$$\sigma_{D_1,90^\circ}, \sigma_{D_2,90^\circ}, \sigma_{D_3,90^\circ}, \sigma_{D_4,90^\circ}, \sigma_{D_6,90^\circ}, \sigma_{D_7,90^\circ}].$$

Here, D_1 through D_4, D_6, and D_7 denote six types of log-derivatives; 0 deg and 90 deg denote the horizontal and vertical log-Gabor filter subbands, respectively.

To illustrate how the log-derivative statistics of the filter responses behave as a function of distortions, Fig. 9 plots the corresponding histograms for five reference images (shown in Fig. 6) as well as the histograms for five distorted versions

Fig. 6 Five reference images from the LIVE database[19] used to demonstrate the consistency of natural scene statistics in both the spatial and frequency domains. Top row: Images *ocean*, *stream*. Bottom row: Images *sailing2*, *sailing3*, and *statue*.
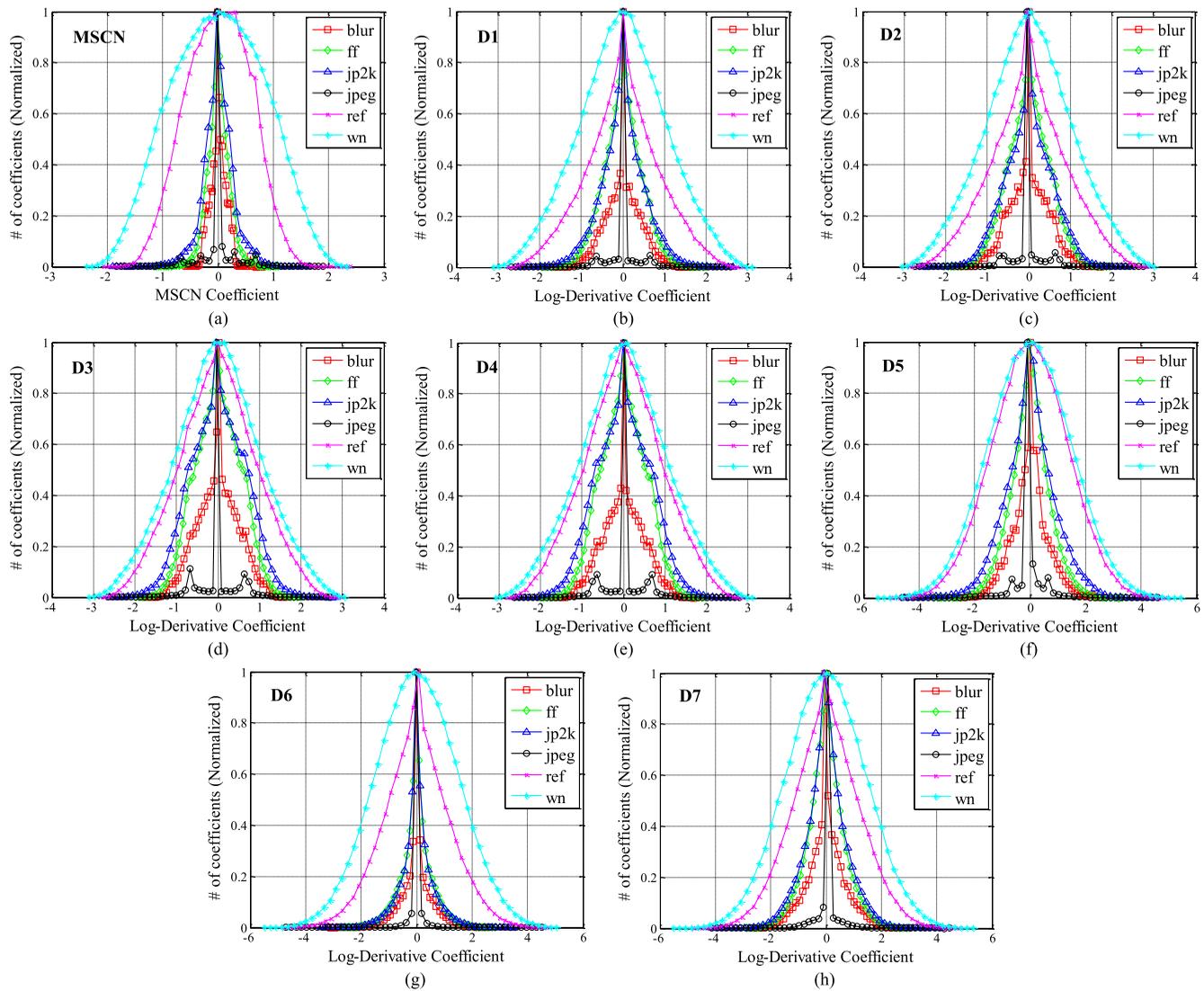
Fig. 7 Histogram of MSCN coefficients (a) and their seven types of log-derivative statistics [(b) to (h)] for each of the five distorted versions of image *sailing2* (shown in Fig. 8). Distortions from the LIVE database—JPEG2000 (JP2K) and JPEG compression (JPEG), additive white noise (WN), Gaussian blur (Gblur), and a Rayleigh fast-fading channel simulation (FF). Notice that the distortions tend to affect the peakiness of the characteristic profile observed for the reference images.

of one reference image (image *sailing2*, shown in Fig. 8) as a comparison. Notice that for both orientations of the log-Gabor filters, the log-derivative statistics of the subband coefficients exhibit consistent profiles that are largely independent of the particular reference image, but change significantly in the presence of distortions. As demonstrated

in Fig. 9, these features extracted in the frequency domain are effective at identifying distortion and thus can be used for measuring image quality.

Since images are naturally multiscale and distortions affect image structure across scales, following from Ref. 15, we extract all spatial features at two scales: the original

Fig. 8 Distorted versions of image *sailing2*. From left to right: Fast-fading, Gaussian blur, JPEG2000 compression, JPEG compression, and Gaussian white noise.
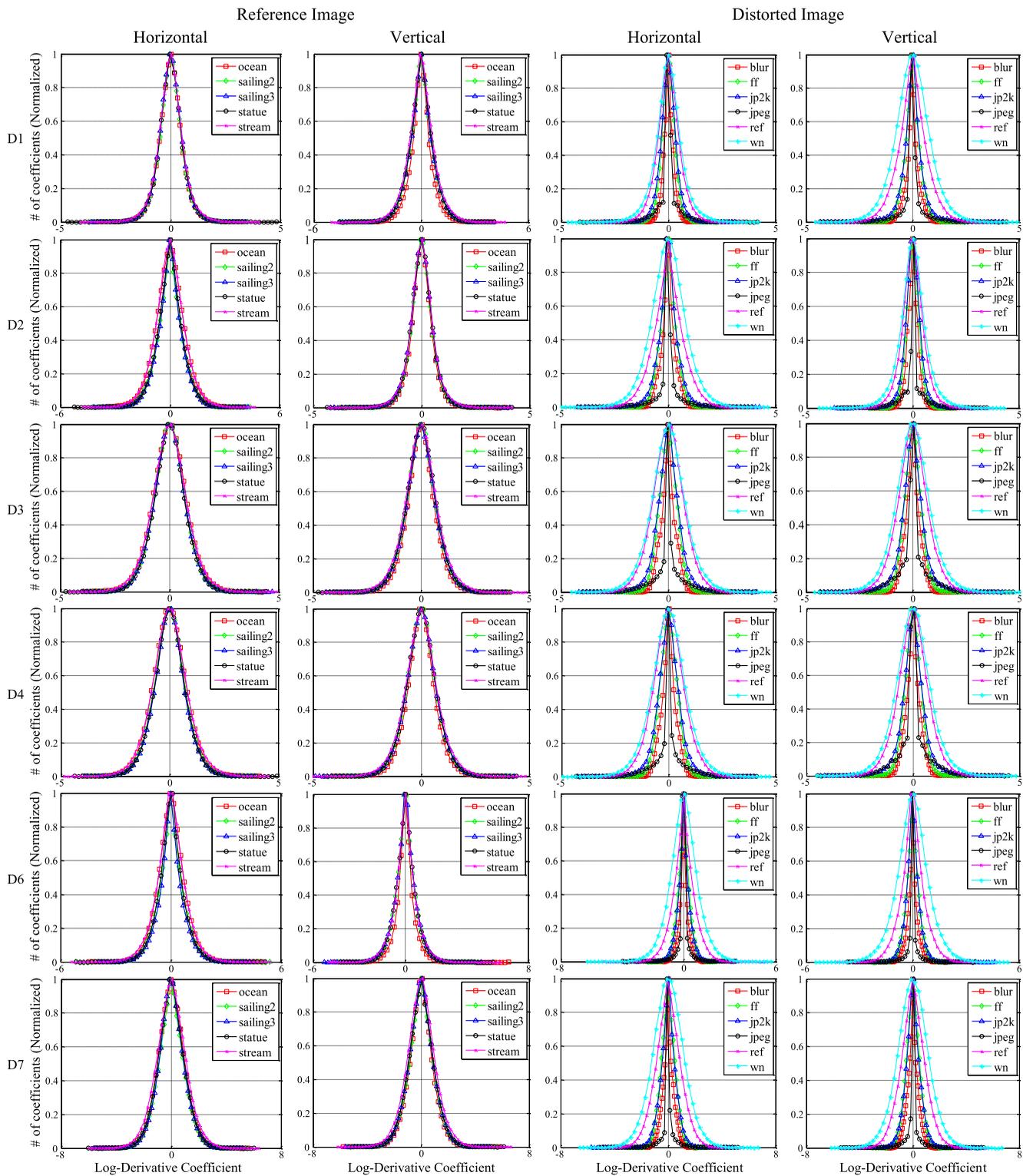
Fig. 9 Histograms of the log-derivative statistics of the log-Gabor filter subband coefficients for five reference images in the LIVE database,[19] and the five distorted versions of one of the reference images (image *sailing2* shown in Fig. 8). D1 to D4 and D6 to D7 correspond to the six log-derivative types defined by Eqs. (9) to (12) and (14) to (15). Distortions from the LIVE database: JPEG2000 (JP2K) and JPEG compression (JPEG), additive white noise (WN), Gaussian blur (Gblur), and a Rayleigh fast-fading channel simulation (FF). Notice that for the five reference images, the histograms exhibit consistent shapes, but they vary significantly when distortions are present.

image scale and a low-pass downsampled (by a factor of 2) scale. The frequency features are extracted based on the six types of log-derivatives at the original scale, and one type (D7) at the low-pass downsampled scale. Thus, a total of 60 features (32 spatial features, 24 frequency features at the original scale, and 4 at the downsampled scale) are used to identify distortions and to perform distortion-specific quality assessment.

3.3 Quality Evaluation

The final step is to map the feature vectors to estimates of image quality. For this task, we employ the existing learning-based approaches commonly used in other NSS-based NR IQA algorithms. We have specifically tested and reported the results of the following approaches: (1) the one-stage framework used in BRISQUE; (2) the two-stage framework used in DIIVINE; (3) a winner-take-all-based variation of (2); and (4) a combination of (1) and (2).

The one-stage framework employs a support vector regression (SVR) to train a regression model, which directly maps the feature vector to an associated quality score, denoted as DESIQUE-I. For the two-stage framework, DESIQUE features are used to perform (1) distortion identification and (2) distortion-specific quality assessment. As in Ref. 13, the distortion identification stage employs support vector classification (SVC) to measure the probability that the distortion in the distorted image falls into one of n distortion classes, and the distortion-specific quality assessment stage employs an SVR to obtain n regression modules, each of which maps the feature vectors to an associated quality score. Note that because each module is trained specifically for each distortion, these regression modules function as distortion-specific quality estimators. Let **p** denote this n-dimensional vector of probabilities and **q** denote the n-dimensional vector of estimated qualities obtained from these n regression modules; then the overall estimated quality, denoted by DESIQUE-II, is computed as follows:

$$\text{DESIQUE-II} = \sum_{i=1}^{n} p(i)q(i), \qquad (21)$$

where $p(i)$ and $q(i)$ denote elements of **p** and **q**, respectively.

We have also investigated a variation of the two-stage framework in which the overall quality estimate is based only on the distortion-specific quality estimate corresponding to the greatest probability. This approach, denoted by DESIQUE-hp, is computed as follows:

$$\text{DESIQUE-hp} = q(i_m), i_m = \arg \max\{p(i)\}. \qquad (22)$$

Another variation that we have investigated, and which is the approach that generally yields the best results (as we will demonstrate shortly) and is thus the approach we recommend, is to use the minimum of the one-stage and two-stage outputs. This approach, denoted by DESIQUE, is computed as follows:

$$\text{DESIQUE} = \min\{\text{DESIQUE-I}, \text{DESIQUE-II}\}. \qquad (23)$$

Smaller values of DESIQUE denote predictions of greater image quality. Note that both the SVC and SVR require training. As we demonstrate next, training these stages on images from one database can yield excellent predictive performance on similarly distorted images from other databases.

4 Results

In this section, the performance of DESIQUE is analyzed in terms of its ability to predict subject ratings of image quality.

4.1 Training

We trained our models on the LIVE database,[19] which contains 29 reference images and 779 distorted images that span various distortion categories: JPEG compression, JPEG2000 compression, white noise, Gblur, and a Rayleigh fading channel (fast fading). Each distorted image has an associated score that indicates a representative of the perceived quality of the image. We use the LIBSVM package[47] to implement the training. To improve predictive performance, optimal radial basis function kernel parameters were used for both classification and regression.

4.2 Testing

To assess its predictive performance, four databases of subjective image quality were used: (1) the LIVE database[19] (note that the LIVE database was used for both training and cross-validation test), (2) the CSIQ database,[25] (3) the TID database,[20] and (4) the Toyama database. The CSIQ database consists of 30 original images distorted using six different types of distortions at four to five different levels of distortion, and the overall ratings are given by 25 different observers in the form of differential mean opinion scores (DMOS). The TID database consists of 25 reference images and 1700 distorted images over 17 distortion types, ratings of which are presented as mean opinion scores (MOS). Among these 25 reference images, only 24 are natural images, and we test our algorithm only on these 24 images. The Toyama database contains 14 original images and 168 distorted versions that cover only two distortion types: JPEG and JPEG2000 compression. The subjective ratings in the Toyama database are provided in the term of MOS values. We compared DESIQUE with various FR and NR quality assessment methods for which code is publicly available. The five FR methods were peak signal-to-noise ratio (PSNR),[48] SSIM,[49] multiscale structure similarity (MS-SSIM),[50] visual information fidelity (VIF),[51] and most apparent distortion (MAD).[52] The three NR methods were DIIVINE,[13] BLIINDS-II,[14] and BRISQUE,[15] all of which are NSS-based and trained on LIVE. Since DESIQUE is trained on the LIVE database containing five distortion types, we tested and compared these algorithms only on those distortion types on which the algorithms have been trained: JPEG compression, JPEG2000 compression, WN, and Gblur.

Before evaluating the performance of a particular quality assessment method on a particular database, we applied a logistic transform to bring the predicted values on the same scales as the DMOS/MOS values. The logistic transform recommended by Video Quality Experts Group[53] is a four-parameter sigmoid given by

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp\left(-\frac{x - \tau_3}{|\tau_4|}\right)} + \tau_2, \qquad (24)$$

where x denotes the raw predicted score and τ_1, τ_2, τ_3, and τ_4 are free parameters selected to provide the best fit of the predicted scores to the MOS/DMOS values. Note that the Spearman rank-order correlation coefficient (SROCC) relies only on the rank-ordering and is thus unaffected by the logistic transform due to the fact that $f(x)$ is a monotonic function of x that does not change the rank-order.

Three criteria were used to measure the prediction monotonicity and prediction accuracy of each algorithm:

(1) SROCC, (2) the Pearson linear correlation coefficient (CC), and (3) the root mean square error (RMSE) after nonlinear regression. As recommended in Ref. 53, the SROCC serves as a measure of prediction monotonicity, while the CC and RMSE serve as measures of prediction accuracy. A value close to 1 for SROCC and CC, and 0 for RMSE, indicates good predictive performance.

Two additional criteria were used to measure the prediction consistency of each algorithm: (1) outlier ratio and (2) outlier distance (OD).[52]

4.3 Cross-Validation Test on LIVE

In this subsection, we performed a cross-validation test on LIVE by splitting the database into two non-overlapping sets—an 80% subset of the database for training and an extra 20% subset for testing. We compared with three NR and three FR IQA methods in terms of median SROCC

and CC values computed over 1000 trials. For comparison, we also include the DESIQUE result of using only one-stage framework, two-stage framework, and two-stage framework with regression module corresponding to the higher probability, denoted by *DESIQUE-I/II/hp*, respectively. The results are shown in Table 1.

In order to evaluate statistical significance, we performed a one-sided *t*-test[54] with 95% confidence level between SROCC values generated by these algorithms across 1000 train-test trials. The results are shown in Table 2, in which "1," "0," "−1" indicate that the mean correlation of the algorithm in row is statistically superior, equivalent, or inferior to the mean correlation of the algorithm in colunm. Also included are DESIQUE-I, DESIQUE-II, and DESIQUE-hp for comparison.

To demonstrate that DESIQUE features can be used for different distortion identification, Table 3 shows the median

Table 1 Median Spearman rank-order correlation coefficient (SROCC) and correlation coefficient (CC) values across 1000 train-test combinations on the LIVE database. Italicized entries denote no-reference image quality assessment (NR IQA) algorithms; others are full-reference (FR) IQA algorithms.

		JPEG2000	JPEG	Additive white noise (WN)	Blur	Rayleigh fast-fading channel	ALL
SROCC	PSNR	0.8646	0.8831	0.9410	0.7515	0.8736	0.8636
	SSIM	0.9389	0.9466	0.9635	0.9046	0.9393	0.9129
	MS-SSIM	0.9627	0.9785	0.9773	0.9542	0.9386	0.9535
	BLIINDS-II	*0.9323*	*0.9331*	*0.9463*	*0.8912*	*0.8519*	*0.9124*
	DIIVINE	*0.9123*	*0.9208*	*0.9818*	*0.9373*	*0.8694*	*0.9250*
	BRISQUE	*0.9139*	*0.9647*	*0.9786*	*0.9511*	*0.8768*	*0.9395*
	DESIQUE-I	*0.9354*	*0.9667*	*0.9849*	*0.9448*	*0.8612*	*0.9406*
	DESIQUE-II	*0.9275*	*0.9683*	*0.9835*	*0.9382*	*0.8661*	*0.9406*
	DESIQUE-hp	*0.9236*	*0.9690*	*0.9822*	*0.9328*	*0.8587*	*0.9374*
	DESIQUE	*0.9359*	*0.9693*	*0.9858*	*0.9482*	*0.8672*	*0.9437*
CC	PSNR	0.8762	0.9029	0.9173	0.7801	0.8795	0.8592
	SSIM	0.9405	0.9462	0.9824	0.9004	0.9514	0.9065
	MS-SSIM	0.9746	0.9793	0.9883	0.9645	0.9488	0.9511
	BLIINDS-II	*0.9386*	*0.9426*	*0.9635*	*0.8994*	*0.8789*	*0.9164*
	DIIVINE	*0.9233*	*0.9348*	*0.9866*	*0.9370*	*0.8916*	*0.9270*
	BRISQUE	*0.9229*	*0.9735*	*0.9851*	*0.9506*	*0.9030*	*0.9424*
	DESIQUE-I	*0.9485*	*0.9782*	*0.9910*	*0.9573*	*0.8729*	*0.9429*
	DESIQUE-II	*0.9416*	*0.9801*	*0.9896*	*0.9496*	*0.8887*	*0.9430*
	DESIQUE-hp	*0.9374*	*0.9803*	*0.9886*	*0.9483*	*0.8835*	*0.9407*
	DESIQUE	*0.9480*	*0.9803*	*0.9915*	*0.9585*	*0.8819*	*0.9465*

Table 2 Results of the one-sided *t*-test performed between SROCC values generated by different measures. "1," "0," "−1" indicate that the algorithm in the row is statistically superior, equivalent, or inferior to the algorithm in the column.

	PSNR	SSIM	MS-SSIM	BLIINDS-II	DIIVINE	BRISQUE	DESIQUE-I	DESIQUE-II	DESIQUE-hp	DESIQUE
PSNR	0	−1	−1	−1	−1	−1	−1	−1	−1	−1
SSIM	1	0	−1	1	−1	−1	−1	−1	−1	−1
MS-SSIM	1	1	0	1	1	1	1	1	1	1
BLIINDS-II	1	−1	−1	0	−1	−1	−1	−1	−1	−1
DIIVINE	1	1	−1	1	0	−1	−1	−1	−1	−1
BRISQUE	1	1	−1	1	1	0	0	0	1	−1
DESIQUE-I	1	1	−1	1	1	0	0	0	1	−1
DESIQUE-II	1	1	−1	1	1	0	0	0	1	−1
DESIQUE-hp	1	1	−1	1	1	−1	−1	−1	0	−1
DESIQUE	1	1	−1	1	1	1	1	1	1	0

Table 3 Mean and median classification accuracy across 1000 train-test trials.

Classification accuracy (%)	JP2K	JPEG	WN	BLUR	FF	ALL
Mean	86.1	97.6	99.9	95.6	82.2	92.3
Median	86.1	100.0	100.0	96.7	83.3	92.6

and mean classification accuracy of the classifier in the two-stage framework for each of the distortions in the LIVE database, as well as across all distortions. To show the performance consistency of each of the algorithms considered here, Fig. 10 plots the mean and standard deviation of SROCC values across these 1000 trials for each of them.
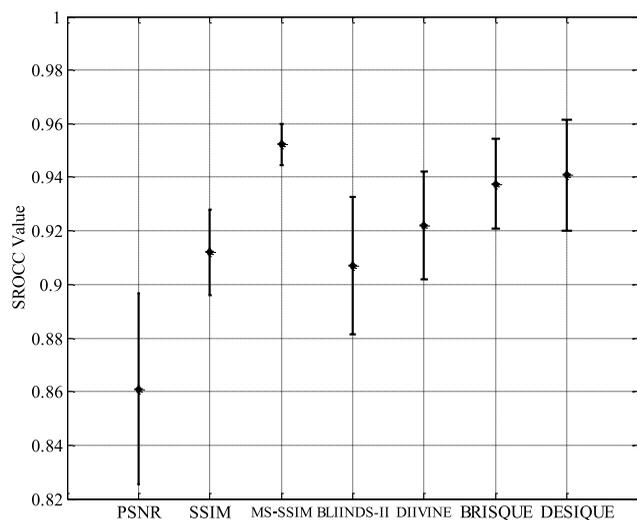
Fig. 10 Mean SROCC and standard error bars for various algorithms across the 1000 train-test trials on the LIVE database.

According to the cross-validation results, DESIQUE performs better than DESIQUE-I and II, which demonstrates that combined frameworks can achieve better performance than using either of them alone. Also, notice that the performance of DESIQUE-I and II are equivalent and are better than DESIQUE-hp. Such a slight dip in the performance of DESIQUE-hp might be attributable to imperfect distortion identification in the first stage of the two-stage framework. Compared with FR IQA methods, DESIQUE is more consistent than PSNR and SSIM, but still remains slightly inferior to MS-SSIM, indicating that there is more space for further improvement. Compared with other three NR IQA algorithms, DESIQUE is statistically the best. However, it is important to note that this cross-validation is not necessarily the best way to gauge the performance of any given algorithm because the amounts of distortions (i.e., the severity levels) are similar in both training and testing sets. Thus, in the following sections, we will show more results of these NR IQA algorithms in assessing image quality on other databases.

4.4 Performances on Other Databases

In this section, we evaluate the performance of DESIQUE on subsets of the CSIQ, TID, and Toyama databases corresponding to the individual distortion types of JPEG, JPEG2000, Gaussian blurring, and additive Gaussian white noise to demonstrate the capability of the proposed

algorithm on assessing quality of images with any distortion types on which it was trained. We also tested DESIQUE on different noise-corrupted images in the TID database to show a degree of adaptiveness to other noise distortion types.

4.4.1 *Overall performance*

The overall testing results on CSIQ, TID, and Toyama are shown in Table 4. Italicized entries denote FR algorithms. Notice that all these FR IQA algorithms compared are only applied on those four distortion types that have been trained. The results of the best-performing FR algorithm in each case are italicized and bolded, and the results of the best-performing NR algorithm are bolded.

From Table 4 it is clear that compared with other NR IQA methods, DESIQUE performs quite well in predicting quality. It improves upon BRISQUE and is superior to DIIVINE and BLIINDS-II. Further, it even challenges some of the FR IQA methods such as PSNR and SSIM. The last rows of the SROCC, CC, and RMSE results in Table 4 show the average values, where the averages were weighted by the number of

distorted images tested in each database. On average, DESIQUE demonstrates the best NR IQA performance. Note that OD is dependent on the dynamic range of the database and, therefore, cannot be used to compare across databases.

Figure 11 shows scatter-plots of logistic-transformed DESIQUE quality predictions versus subjective ratings (MOS or DMOS) on different databases. Although for each database there are some images whose quality scores are predicted far from their true MOS/DMOS values, the proposed DESIQUE algorithm can predict quality for most of them.

In summary, when looking at the overall performance across databases, DESIQUE has a better average performance than other NR IQA methods.

4.4.2 *Statistical significance*

A statistical significance test allows one to quantify whether the numerical difference between IQA performance is statistically significant. To assess the statistical significance, an F-test was employed to compare the variances of two

Table 4 Overall performances of DESIQUE and other algorithms on subsets of the CSIQ, TID, and Toyama databases. Italicized entries denote FR algorithms. Results of the best-performing FR algorithm are italicized and bolded, and results of the best-performing NR algorithm are bolded.

		PSNR	*SSIM*	*MS-SSIM*	*VIF*	*MAD*	DIIVINE	BLIINDS-II	BRISQUE	DESIQUE
CC	CSIQ	*0.908*	*0.851*	*0.950*	*0.967*	***0.974***	0.854	0.901	0.924	**0.942**
	TID	*0.848*	*0.736*	*0.912*	***0.950***	*0.947*	0.877	0.864	0.907	**0.925**
	Toyama	*0.635*	*0.796*	*0.893*	*0.914*	***0.941***	0.634	0.754	0.850	**0.871**
	Average	*0.848*	*0.805*	*0.929*	*0.954*	0.960	0.830	0.867	0.907	**0.926**
SROCC	CSIQ	*0.922*	*0.876*	*0.953*	*0.959*	***0.967***	0.828	0.873	0.900	**0.928**
	TID	*0.870*	*0.767*	*0.897*	***0.940***	*0.935*	0.891	0.840	0.898	**0.919**
	Toyama	*0.613*	*0.786*	*0.886*	*0.909*	***0.936***	0.642	0.724	0.848	**0.872**
	Average	*0.860*	*0.827*	*0.924*	*0.945*	***0.952***	0.822	0.840	0.892	**0.917**
RMSE	CSIQ	*0.119*	*0.148*	*0.088*	*0.072*	***0.064***	0.147	0.123	0.108	**0.095**
	TID	*0.840*	*1.072*	*0.650*	***0.495***	*0.509*	0.760	0.798	0.668	**0.600**
	Toyama	*0.966*	*0.757*	*0.564*	*0.507*	***0.425***	0.968	0.864	0.660	**0.615**
	Average	*0.483*	*0.545*	*0.345*	*0.276*	***0.265***	0.468	0.456	0.375	**0.339**
Outlier ratio	CSIQ	*0.272*	*0.332*	*0.220*	*0.187*	***0.157***	0.365	0.320	0.262	**0.243**
	TID	*0.727*	*0.784*	*0.734*	*0.630*	***0.622***	0.706	0.766	0.693	**0.659**
	Toyama	*0.190*	*0.095*	*0.066*	*0.030*	***0.024***	0.214	0.155	0.066	**0.060**
	Average	*0.412*	*0.448*	*0.369*	*0.312*	***0.293***	0.457	0.444	0.377	**0.355**
Outlier distance	CSIQ	*11.331*	*19.309*	*6.473*	*5.013*	***3.243***	24.561	16.318	12.697	**9.812**
	TID	*176.350*	*233.982*	*120.957*	***78.159***	*79.194*	137.472	162.692	121.259	**101.156**
	Toyama	*20.904*	*9.396*	*3.172*	***1.397***	*1.673*	16.517	12.685	**4.156**	5.671
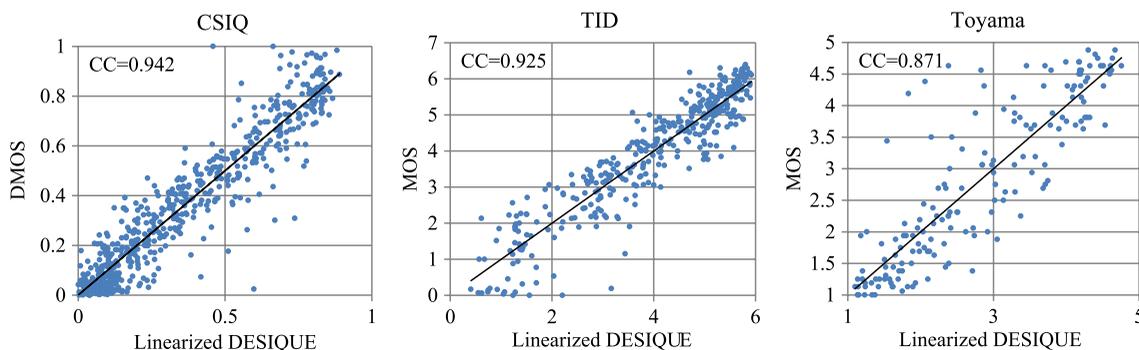
Fig. 11 Scatter plots of objective scores predicted by DESIQUE algorithm after logistic transform versus subjective scores on different image databases. Note that the *x* axis across all three figures represents the predicted value transformed via Eq. (24); the *y* axis represents the true differential mean opinion score (DMOS) value for the CSIQ database, true MOS value for the TID and Toyama databases.

algorithms' residuals (errors in predictions) if the residuals are Gaussian-distributed. The test statistic is the ratio of variance, denoted by $F = \sigma_A^2 / \sigma_B^2$, where σ_A^2 and σ_B^2 denote the variance of the residuals from each IQA algorithm. A smaller residual variance indicates a better prediction. Values of $F > F_{\text{critical}}$ (or $F < 1/F_{\text{critical}}$) indicate that at a given confidence level, algorithm A has significantly larger (or smaller) residuals than algorithm B, where F_{critical} is computed based on the number of residuals and the confidence level. Also, note that if residuals are not Gaussian, then the significance test is often inconclusive. In this paper, a formal test using Jarque-Bera (JB) statistic[55] is used to measure the Gaussianity of the residuals. A smaller value of the JB statistic denotes less deviation from Gaussianity, and vice versa.

Table 5 shows the summary for the overall statistical performance of each NR IQA algorithm on subsets of the CSIQ, TID, and Toyama databases. Each entry is the ratio of the residual variance of the algorithm in the row to the algorithm in the column. Bold entries denote that the algorithm in the row has a statistically smaller residual variance than the algorithm in the column with confidence greater than 95%. Italicized entries denote that the algorithm in the row has statistically greater residual variance with the same confidence. Also contained in Table 5 are the JB statistic measures of Gaussianity. Larger values of the JB statistic denote larger deviations from Gaussianity.

Although DESIQUE has statistically the smallest residual variance on the subsets of the CSIQ and TID databases, note that none of the DESIQUE residuals on these databases can be deemed as Gaussian, which is attributable to several outliers (see Fig. 11). However, the residual variance of DESIQUE is inflated by the existence of these outliers, and thus the fact that it can achieve significantly lower residual variance with these outliers is noteworthy. For the subsets of the Toyama database, the statistical significance is less selective due to the small number of images tested. Nonetheless, DESIQUE still performs competitively well.

4.4.3 Performance on individual distortion types

We also tested the performance of DESIQUE on subsets of the three testing databases corresponding to the individual distortion types of JPEG, JPEG2000, Gaussian blurring, and additive Gaussian white noise. For this evaluation, we used the same logistic transform computed for each full database and then extracted the transformed scores

Table 5 Statistical significance relationships (ratio of residual variance) between NR IQA algorithms on subsets of the CSIQ, TID, and Toyama databases. A value <1 denotes that the algorithm in the row has smaller residuals than the algorithm in the column; a value > 1 denotes larger residuals. See text for details.

		DIIVINE	BLIINDS-II	BRISQUE	DESIQUE
CSIQ	DIIVINE	—	*1.429*	*1.836*	*2.382*
	BLIINDS-II	**0.700**	—	*1.285*	*1.667*
	BRISQUE	**0.545**	**0.778**	—	*1.297*
	DESIQUE	**0.420**	**0.600**	**0.771**	—
	JBSTAT	143.7	78.5	958.6	661.0
TID	DIIVINE	—	*1.649*	*2.351*	*2.912*
	BLIINDS-II	**0.607**	—	*1.426*	*1.766*
	BRISQUE	**0.425**	**0.701**	—	*1.239*
	DESIQUE	**0.343**	**0.566**	**0.807**	—
	JBSTAT	120.7	2.3	5.7	104.7
Toyama	DIIVINE	—	1.255	*2.153*	*2.478*
	BLIINDS-II	0.797	—	*1.716*	*1.975*
	BRISQUE	**0.465**	**0.583**	—	1.151
	DESIQUE	**0.404**	**0.506**	0.869	—
	JBSTAT	2.6	15.0	4.0	96.2

corresponding to each distortion type. Three other NR IQA methods and five FR IQA algorithms were included to make the comparison. The results are shown in Table 6 in terms of SROCC and CC values; highlighted entries indicate the best performance. All of these NR IQA methods were trained on the 779 distorted images in the LIVE database with the same distortion types.

Table 6 SROCC, CC, and RMSE of DESIQUE and other quality assessment algorithms on different types of distortion on the CSIQ, TID, and Toyama databases. Italicized entries denote NR algorithms. Results of the best-performing FR algorithm are italicized and bolded, and results of the best-performing NR algorithm are bolded.

		PSNR	*SSIM*	*MS-SSIM*	*VIF*	*MAD*	DIIVINE	BLIINDS-II	BRISQUE	DESIQUE
					CC					
CSIQ	JPEG2000	*0.947*	*0.923*	*0.977*	*0.978*	***0.983***	0.893	0.912	0.896	**0.925**
	JPEG	*0.891*	*0.940*	*0.981*	***0.988***	*0.983*	0.697	0.912	0.946	**0.973**
	Gblur	*0.925*	*0.900*	*0.959*	*0.974*	***0.976***	0.898	0.897	**0.928**	0.922
	WN	*0.953*	*0.926*	*0.947*	***0.961***	*0.956*	0.786	0.897	0.938	**0.942**
TID	JPEG2000	*0.885*	*0.875*	*0.974*	*0.971*	***0.982***	0.879	0.919	0.906	**0.928**
	JPEG	*0.878*	*0.937*	*0.966*	***0.973***	*0.961*	0.899	0.889	0.950	**0.971**
	Gblur	*0.930*	*0.938*	***0.951***	*0.942*	*0.801*	0.840	0.825	0.873	**0.885**
	WN	***0.942***	*0.807*	*0.810*	*0.907*	*0.819*	0.810	0.714	0.810	**0.873**
Toyama	JPEG	*0.377*	*0.652*	*0.786*	*0.900*	***0.919***	0.709	0.826	0.864	**0.877**
	JPEG2000	*0.858*	*0.918*	*0.949*	***0.962***	*0.961*	0.603	0.686	0.869	**0.880**
					SROCC					
CSIQ	JPEG2000	*0.936*	*0.921*	*0.969*	*0.967*	***0.975***	0.830	0.884	0.867	**0.913**
	JPEG	*0.888*	*0.922*	*0.962*	***0.970***	*0.962*	0.704	0.881	0.909	**0.944**
	Gblur	*0.929*	*0.924*	*0.972*	***0.975***	*0.968*	0.871	0.870	**0.903**	0.901
	WN	*0.936*	*0.925*	*0.947*	***0.957***	*0.954*	0.797	0.886	0.925	**0.930**
TID	JPEG2000	*0.825*	*0.878*	*0.973*	*0.970*	***0.974***	0.907	0.911	0.904	**0.928**
	JPEG	*0.875*	*0.925*	***0.940***	*0.931*	*0.925*	0.871	0.838	0.910	**0.932**
	Gblur	*0.934*	*0.945*	***0.963***	*0.958*	*0.847*	0.859	0.826	0.874	**0.894**
	WN	***0.918***	*0.812*	*0.818*	*0.913*	*0.833*	0.834	0.715	0.823	**0.882**
Toyama	JPEG	*0.285*	*0.626*	*0.835*	*0.907*	***0.916***	0.702	0.820	0.857	**0.880**
	JPEG2000	*0.861*	*0.914*	*0.945*	***0.956***	*0.955*	0.612	0.627	0.867	**0.882**
					RMSE					
CSIQ	JPEG2000	*0.102*	*0.122*	*0.067*	*0.066*	***0.058***	0.130	0.142	0.140	**0.120**
	JPEG	*0.139*	*0.104*	*0.059*	***0.047***	*0.057*	0.125	0.219	0.099	**0.070**
	Gblur	*0.109*	*0.125*	*0.081*	*0.065*	***0.062***	0.127	0.126	**0.107**	0.111
	WN	*0.051*	*0.064*	*0.054*	***0.047***	*0.050*	0.074	0.104	0.058	**0.056**
TID	JPEG2000	*0.891*	*0.928*	*0.430*	*0.459*	***0.366***	1.178	0.756	0.810	**0.712**
	JPEG	*0.809*	*0.592*	*0.435*	***0.388***	*0.468*	0.741	0.774	0.528	**0.403**
	Gblur	*0.428*	*0.402*	***0.359***	*0.390*	*0.694*	0.621	0.655	0.566	**0.540**
	WN	***0.206***	*0.362*	*0.359*	*0.258*	*0.351*	0.343	0.428	0.359	**0.298**
Toyama	JPEG	*1.145*	*0.937*	*1.236*	*0.540*	***0.488***	0.872	0.708	0.622	**0.595**
	JPEG2000	*0.650*	*0.502*	*0.398*	***0.344***	*0.348*	1.007	0.983	0.625	**0.601**

Table 7 Performances on three other noise distortion types in the TID database.

		DIIVINE	BLIINDS-II	BRISQUE	DESIQUE
SROCC	Additive noise in color components (ACN)	0.776	0.737	0.747	0.852
	Spatially correlated noise (SCN)	0.364	0.462	0.569	0.825
	High-frequency noise (HFN)	0.894	0.871	0.755	0.925
CC	ACN	0.790	0.754	0.768	0.866
	SCN	0.403	0.533	0.570	0.811
	HFN	0.908	0.895	0.705	0.946
RMSE	ACN	0.313	0.336	0.328	0.256
	SCN	0.569	0.526	0.511	0.363
	HFN	0.400	0.426	0.679	0.310

As shown in Table 6, DESIQUE provides better predictions in comparison to BLIINDS-II, DIIVINE, and BRISQUE on most distortion types. Notice that for the white noise images in the TID database, DESIQUE demonstrates better predictive performance compared with the other three NSS-based NR IQA algorithms. For some distortion types, such as the JPEG compression, in both the CSIQ and TID databases, DESIQUE performs competitively with many FR IQA methods. In summary, when looking at the performance on individual distortion types, DESIQUE also demonstrates the best NR IQA performance.

4.4.4 *Performance on noise-corrupted images*

Although DESIQUE was trained on WN images in the LIVE database, it can also yield valid predictions of quality when other types of noise are present. To demonstrate the adaptiveness of DESIQUE to different noise types, we tested the algorithm on three other noise distortion types in the TID database: additive noise in color components (ACN), spatially correlated noise (SCN), and high-frequency noise (HFN) (notice that in this test all NR algorithms were still trained on LIVE with its five standard distortion types). We compared with three other NR algorithms in terms of CC, SROCC, and RMSE. The results are listed in Table 7.

As shown in Table 7, DESIQUE demonstrates better performance than DIIVINE, BLIINDS-II, and BRISQUE on these three noise distortion types, despite the fact that none of them were trained on these distortions. These results demonstrate that different noise categories can be detected and measured by using both spatial and spectral feature types. Figure 12 shows the scatter plots of BRISQUE and DESIQUE on these different noise-corrupted images before logistic transform. Notice that BRISQUE even fails on predicting qualities of the spatially correlated noise images, as it produces an increasing relationship with DMOS values in the view of all data points, even though the corresponding SROCC and CC values are computed as 0.569 and 0.570, respectively. Also notice that for the HFN images, DESIQUE provides even more competitive results than

BRISQUE, again owing to DESIQUE's dual-domain analysis. In summary, our proposed DESIQUE algorithm demonstrates not only a competitive performance for the distortion types on which it was trained, but also a degree of adaptiveness to images distorted by different types of noise.

4.4.5 *Performance on reference images*

In this section, we show the performance of DESIQUE on reference images from the three testing databases. Notice that there are no valid human subjective quality scores for the reference images in the CSIQ and TID databases. Although in the Toyama database, human subjective quality scores are provided in terms of MOS values, it still seems to be inappropriate to measure an algorithm's performance using SROCC/CC because the MOS values of reference images are too close to provide any meaningful linear/nonlinear relationships. Therefore, to show the performance of our proposed algorithm, we first predicted the original DMOS value of each reference image using our algorithm and then applied the logistic transform defined in Eq. (24) to bring the predicted values on the same scales as the true DMOS/MOS values for the distorted images. Here, the four parameters of the sigmoid function were estimated by using the distorted images in the same database. Finally, we compared these linearized predicted DMOS values to the corresponding scatter plots in Fig. 11 for each database to see if they are all located in a region representing the highest quality.

Figure 13 shows results of DESIQUE tested on all reference images in CSIQ, TID, and Toyama. Note that we tested only 24 undistorted versions of the natural images in the TID database. By referring to Fig. 11, we conclude that DESIQUE works well for most reference images. Specifically, for the CSIQ database, 28 out of 30 reference images are given quality scores below 0.15 (note in Fig. 11 that the true DMOS values of the reference images in the CSIQ database should be close to zero). Thus, quality scores of only two reference images were predicted slightly higher than their true DMOS values. For the TID databases, all
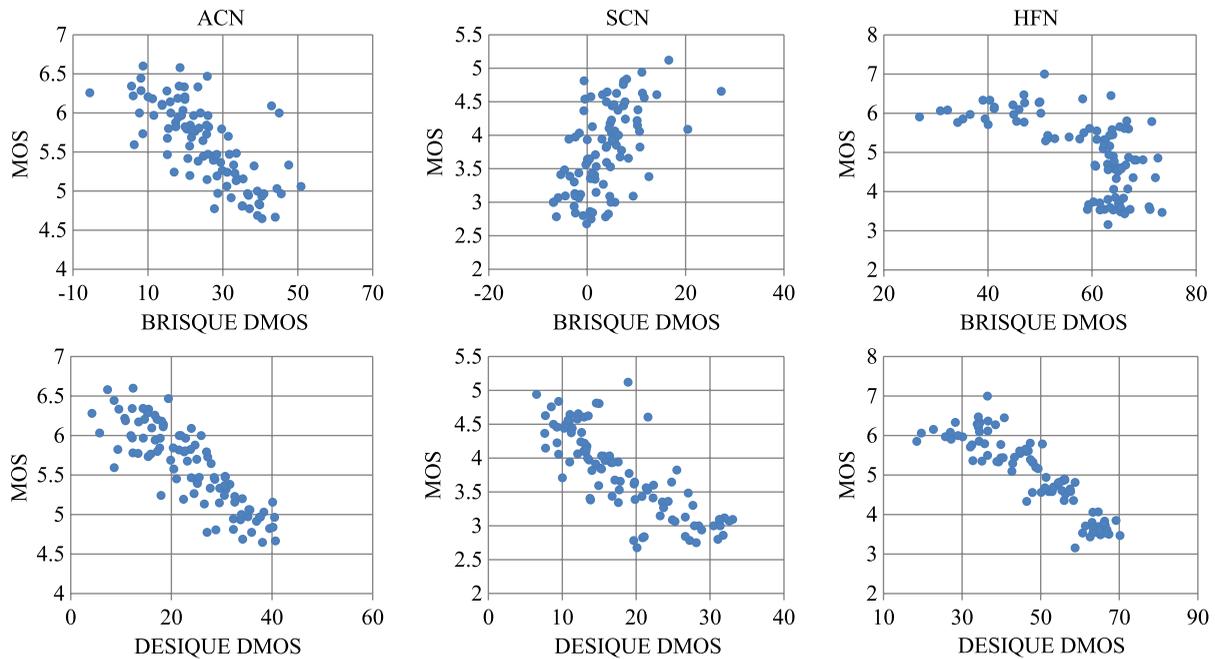
Fig. 12 Scatter plots of objective quality scores (before logistic transform) predicted by BRISQUE and DESIQUE on different noise-corrupted images in the TID database. Note that the *x* axis across all three figures represents the algorithm predicted DMOS value; the *y* axis represents the true MOS value.
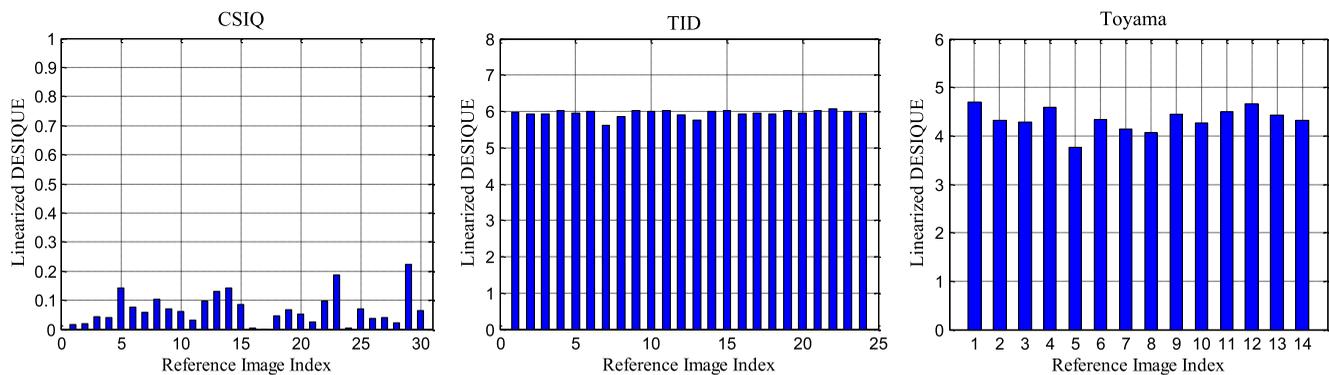
Fig. 13 Objective quality scores predicted by DESIQUE algorithm after logistic transform. Note that the *x* axis across all three figures represents the index of reference images in each database; the *y* axis represents the predicted value transformed via Eq. (24).

reference images seem to be evaluated quite well (note in Fig. 11 that the undistorted/reference images in the TID database should have true MOS values around/above six). For the Toyama database, only one reference image was predicted below four, a bit far away from its true MOS value (note again in Fig. 11 that the undistorted/reference images in the TID database should have true MOS values around/above 4.5). All three figures demonstrate that the proposed DESIQUE algorithm performs quite well on these reference images.

4.5 Contribution Analysis

To analyze the contributions of the log-derivative statistics and each of the two domains (spatial domain and frequency domain) toward the final performance, we performed the prediction test on the aforementioned four databases by using only the spatial-domain-based features (denoted by DESIQUE-S) and only the frequency-domain-based features

(denoted by DESIQUE-F), respectively. Three types of testing were performed: (1) the 1000 cross-validation test of DESIQUE-S and F were performed on the LIVE database; (2) the overall performance of DESIQUE-S and F were tested on subsets of the CSIQ, TID, and Toyama databases; and (3) the performance of DESIQUE-S and F were tested on noise-corrupted images from both the CSIQ and TID databases. Note that in (3) we tested on ACN, SCN, and HFN images from TID, and on WN images from both CSIQ and TID (denoted by WN-CSIQ/TID). Also note that models of DESIQUE-S and F for (2) and (3) were obtained via training on the whole LIVE database, and quality was predicted in all three tests using the combined frameworks as recommended in this paper. For comparison, we also listed the corresponding results of the BRISQUE algorithm and of DESIQUE with combined domains employed. The experiment results are shown in Tables 8 and 9.

Table 8 Median SROCC, CC, and RMSE values of BRISQUE, DESIQUE-S, DESIQUE-F, and DESIQUE algorithms across 1000 train-test combinations on the LIVE database.

		JP2K	JPEG	WN	BLUR	FF	ALL
SROCC	BRISQUE	0.9139	0.9647	0.9786	0.9511	0.8768	0.9395
	DESIQUE-S	0.9361	0.9636	0.9862	0.9519	0.8423	0.9305
	DESIQUE-F	0.9082	0.9623	0.9840	0.9261	0.8000	0.9188
	DESIQUE	0.9359	0.9693	0.9858	0.9482	0.8672	0.9437
CC	BRISQUE	0.9229	0.9735	0.9851	0.9506	0.9030	0.9424
	DESIQUE-S	0.9459	0.9725	0.9914	0.9631	0.8576	0.9354
	DESIQUE-F	0.9181	0.9732	0.9900	0.9389	0.8503	0.9238
	DESIQUE	0.9480	0.9803	0.9915	0.9585	0.8819	0.9465
RMSE	BRISQUE	9.6934	7.1641	4.7856	5.6450	12.1024	9.1315
	DESIQUE-S	8.0470	7.3600	3.6319	4.9680	14.1329	9.6542
	DESIQUE-F	9.9836	7.2440	3.9111	6.3000	14.3949	10.4679
	DESIQUE	7.9692	6.2123	3.6172	5.2285	13.3172	8.7985

According to Table 8, the log-derivative-based analysis in either of the two domains (spatial and frequency) seems to be inferior to the paired-product-based analysis in the spatial domain when performing the cross-validation test on the LIVE database. However, combining the two domain analyses together can improve the performance significantly.

According to Table 9, the log-derivative statistics improve upon paired product statistics in predicting image quality when training on LIVE and testing on other databases (note that although DESIQUE-S only has SROCC and CC values of 0.449 and 0.469 on SCN, BRISQUE actually fails on SCN as illustrated in Sec. 4.4.4), and this fact also

Table 9 SROCC, CC, and RMSE values of BRISQUE, DESIQUE-S, DESIQUE-F, and DESIQUE algorithms on subsets of the CSIQ, TID, and Toyama databases, as well as on different noise-corrupted images in the TID and CSIQ databases.

		CSIQ	TID	Toyama	ACN	SCN	HFN	WN-TID	WN-CSIQ
SROCC	BRISQUE	0.900	0.898	0.848	0.747	0.569	0.755	0.823	0.925
	DESIQUE-S	0.921	0.918	0.894	0.846	0.449	0.905	0.849	0.928
	DESIQUE-F	0.919	0.885	0.816	0.805	0.644	0.903	0.880	0.935
	DESIQUE	0.928	0.919	0.872	0.852	0.825	0.925	0.882	0.930
CC	BRISQUE	0.924	0.907	0.850	0.768	0.570	0.705	0.810	0.938
	DESIQUE-S	0.933	0.925	0.893	0.856	0.469	0.930	0.841	0.939
	DESIQUE-F	0.932	0.877	0.819	0.817	0.664	0.935	0.873	0.945
	DESIQUE	0.942	0.925	0.871	0.866	0.811	0.946	0.873	0.942
RMSE	BRISQUE	0.108	0.668	0.660	0.328	0.511	0.679	0.359	0.058
	DESIQUE-S	0.102	0.600	0.563	0.264	0.549	0.352	0.331	0.058
	DESIQUE-F	0.103	0.760	0.718	0.295	0.465	0.340	0.298	0.055
	DESIQUE	0.095	0.600	0.615	0.256	0.363	0.310	0.298	0.056

Table 10 Runtime requirements (seconds/image) for four NR IQA methods on different image sizes.

	256×256	512×512	1024×1024	1600×1600
BRISQUE	0.160	0.233	0.544	1.158
DIIVINE	8.661	22.428	84.436	225.328
BLIINDS-II	9.843	38.904	156.365	$\approx$380
DESIQUE	0.128	0.367	1.373	3.258

demonstrates that the two statistics are fundamentally different. Although the log-derivative statistics in the frequency domain analysis contributes less than that in the spatial domain when consulting with the overall performance on subsets of the three testing databases, it does help improve or maintain the predicting performance on CSIQ and TID, and especially on those different noise-corrupted images. On the Toyama database, however, the frequency-domain analysis seems to be ineffective. In summary, in regards to the performance, when all databases are evaluated together, the log-derivative statistics proposed in this paper are superior to the paired product statistics in predicting image quality, and both domains (spatial and frequency) are required to achieve better performance across different databases and distortions.

4.6 *Computational Analysis*

DESIQUE also exhibits relatively low computational complexity. Although DESIQUE extracts statistical features in both the spatial and frequency domains to evaluate image quality, and its number of features (60) is larger than that of BRISQUE (which uses only 36 features), it is still quite efficient in computation. To demonstrate this fact, we compared the overall computational complexity of DESIQUE with three other NR IQA methods: BLIINDS-II, DIIVINE, and BRISQUE on different image sizes (256×256, 512×512, 1024×1024, and 1600×1600 pixels). The test was performed on a modern desktop computer (AMD Phenom II X4 965 Processor at 3.39 GHz, 4.00 GB RAM, Windows 7 Pro 64-bit, Matlab 7.8.0). Table 10 shows the average runtime of each algorithm in seconds, where the average was taken over 100 trials.

As shown in Table 10, DESIQUE is only a little slower than BRISQUE when images have large sizes. This is due to the fact that DESIQUE has to compute both the spatial- and frequency-domain features and that the parameter estimation needs to be performed 30 times for an entire image (BRISQUE only has 10 estimations) in total. However, considering its predictive performance improvement, we believe that the time cost is justified, and it is still much faster than DIIVINE and BLIINDS-II.

5 Conclusion

This paper presented an algorithm for NR image quality assessment (DESIQUE), which operates by using log-derivative statistics of natural scenes. DESIQUE extracts log-derivative-based statistical features at two image scales in both the spatial and frequency domains, upon which combined frameworks perform NR IQA. We demonstrated that DESIQUE can achieve better performance in predicting image quality than many other well-known NR IQA methods across various databases. We also showed that DESIQUE demonstrates a degree of adaptiveness to different types of noise and is well performed on reference images. We also showed that DESIQUE is computationally efficient: it is among the fastest NSS-based NR IQA methods currently available. Future work involves building a more comprehensive training database and testing DESIQUE on a wider range of distortion types. Future work will also include exploring other features/feature combinations, as well as other frameworks (e.g., Dempster-Shafer framework[56,57]) to further improve the algorithm performance.

References

1. X. Li, "Blind image quality assessment," in *Proc. of Int. Conf. on Image Processing*, Rochester, New York, Vol. l, pp. I-449–I-452 (2002).
2. S. Gabarda and G. Cristobal, "No-reference image quality assessment through von mises distribution," *J. Opt. Soc. Am. A* **29**(10), 2058–2066 (2012).
3. T. Brandao and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Process.* **88**(4), 822–833 (2008).
4. E. Cohen and Y. Yitzhaky, "No-reference assessment of blur and noise impacts on image quality," *Signal, Image Video Process.* **4**(3), 289–302 (2010).
5. Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. of Int. Conf. on Image Processing*, Vancouver, BC, Vol. 3, pp. 981–984 (2000).
6. A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proc. of IEEE Int. Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, Vol. 3, pp. 1725–1728 (2001).
7. Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. of IEEE Int. Conf. on Image Processing*, Rochester, New York, Vol. 1, pp. 477–480 (2002).
8. H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.* **14**(11), 1918–1927 (2005).
9. H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Int. Conf. on Computer Vision and Pattern Recognition*, Providence, Rhode Island, pp. 305–312 (2011).
10. P. Ye and D. Doermann, "No-reference image quality assessment based on visual codebook," in *18th IEEE Int. Conf. on Image Processing*, Brussels, pp. 3089–3092 (2011).
11. P. Ye and D. Doermann, "No-reference image quality assessment using visual codebook," *IEEE Trans. Image Process.* **21**(7), 3129–3138 (2012).
12. A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.* **17**(5), 513–516 (2010).
13. A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Trans. Image Process.* **20**(12), 3350–3364 (2011).
14. M. A. Saad and A. C. Bovik, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.* **21**(8), 3339–3352 (2012).
15. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012).
16. C. Vu, T. D. Phan, and D. M. Chandler, "S3: a spectral and spatial sharpness measure," *IEEE Trans. Image Process.* **21**(3), 934–945 (2012).
17. J. Huang and D. Mumford, "Statistics of natural images and models," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, Vol. 1, pp. 541–547 (1999).
18. H. R. Sheikh, "The statistics of natural images," *Netw. Comput. Neural Syst.* **5**(4), 517–548 (1994).

19. D. L. RudermanH. R. Sheikh et al., "Image and video quality assessment research at LIVE," http://live.ece.utexas.edu/research/quality/.
20. N. Ponomarenko et al., "TID2008—a database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.* **10**, 30–45 (2009).
21. J. Daugman, "Uncertainty relation for resolution in space, sptial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am.* **2**(7), 1160–1169 (1985).
22. S. Marcelja, "Mathematical description of the response of simple cortical cells," *J. Opt. Soc. Am.* **70**(11), 1297–1300 (1980).
23. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**(12), 2379–2394 (1987).
24. Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.* **3**(2), 202–211 (2009).
25. http://vision.okstate.edu/csiq/.
26. L. Meesters and J. B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Process.* **82**(3), 369–387 (2002).
27. F. Pan et al., "Measuring blocking artifacts using edge direction information," in *2004 IEEE Int. Conf. on Multimedia and Expo*, Taipei, Vol. 2, pp. 1491–1494 (2004).
28. C.-S. Park, J.-H. Kim, and S.-J. Ko, "Fast blind measurement of blocking artifacts in both pixel and DCT domains," *J. Math. Imaging Vis.* **28**(3), 279–284 (2007).
29. C. Chen and J. A. Bloom, "A blind reference-free blockiness measure," in *Proc. of the 11th Pacific Rim Conf. on Advances in Multimedia Information Processing: Part I*, Shanghai, China, pp. 112–123 (2010).
30. C. Perra, F. Massidda, and D. D. Giusto, "Image blockiness evaluation based on sobel operator," in *Int. Conf. on Image Processing*, Genoa, Italy, Vol. 1, pp. 389–392 (2005).
31. H. Zhang, Y. Zhou, and X. Tian, "A weighted sobel operator-based no-reference blockiness metric," in *Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, Wuhan, China, Vol. 1, pp. 1002–1006 (2008).
32. J. Chen et al., "A no-reference blocking artifacts metric using selective gradient and plainness measures," in *Proc. of the 9th Pacific Rim Conf. on Multimedia: Advances in Multimedia Information Processing*, Tainan, Taiwan, Vol. 5353, pp. 894–897 (2008).
33. S. Suthaharan, "No-reference visually significant blocking artifact metric for natural scene images," *Signal Process.* **89**(8), 1647–1652 (2009).
34. S. Suresh, R. V. Babu, and H. J. Kim, "No-reference image quality assessment using modified extreme learning machine classifier," *Appl. Soft Comput.* **9**(2), 541–552 (2009).
35. M. Li, H. Zhang, and C. Zhang, "No-reference quality assessment for JPEG2000 compressed images," in *Int. Conf. on Image Processing*, Singapore, Vol. 5, pp. 3539–3542 (2004).
36. P. M. Frederic et al., "Perceptual blur and ringing metric: application to JPEG2000," *Signal Process.: Image Commun.* **19**(2), 163–172 (2004).
37. Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "Spatial features based no reference quality assessment for JPEG2000," in *IEEE Int. Conf. on Image Processing*, Vol. 3, pp. III–517–III–520 (2007).
38. Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Process.: Image Commun.* **23**(4), 257–268 (2008).
39. J. Zhou, B. Xiao, and Q. Li, "A no reference image quality assessment method for JPEG2000," in *IEEE Int. Conf. on Neural Networks*, Hong Kong, pp. 863–868 (2008).
40. J. Zhang, S. H. Ong, and T. M. Le, "Kurtosis-based no-reference quality assessment for JPEG2000 images," *Signal Process.: Image Commun.* **26**(1), 13–23 (2011).
41. B. R. Corner, R. M. Narayanan, and S. E. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *Int. J. Remote Sens.* **24**(4), 689–702 (2003).
42. H. Tong et al., "Learning no-reference quality metric by examples," in *Proc. 11th Int. Multimedia Modelling Conf.*, Melbourne, Australia, pp. 247–254 (2005).
43. C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.* **22**(5), 793–799 (2011).
44. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a complete blind image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).
45. T.-L. Chen, "On the statistics of natural images," PhD Thesis, Division of Applied Mathematics at Brown University (2005).
46. K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distribution in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.* **5**(1), 52–56 (1995).
47. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Syst. Technol.* **2**(3), 27:1–27:27 (2001).
48. ANSI T1.TR.74-2001, "Objective video quality measurement using a peak-signal-to-noise-ratio (PSNR) full reference technique," T1.TR.74–2001 (2001).
49. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
50. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Conference Record of the Thirty-Seventh Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, California, Vol. 2, pp. 1398–1402 (2003).
51. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
52. E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging* **19**(1), 011006 (2010).
53. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," 2003, http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx.
54. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida (2007).
55. A. K. Bera and C. M. Jarque, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Econ. Lett.* **6**(3), 255–259 (1980).
56. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton (1976).
57. R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Inf. Sci.* **41**(2), 93–137 (1987).

Yi Zhang received his BS and MS degrees in electrical engineering from Northwestern Polytechnical University, Xi'an, China, in 2008 and 2011, respectively. He is currently working toward his PhD degree in the School of Electrical and Computer Engineering at Oklahoma State University, Stillwater, Oklahoma. His research areas include image and video processing, machine learning, pattern recognition, and computer vision.

Damon M. Chandler received his BS and MS degrees in biomedical engineering from Johns Hopkins University, Baltimore, Maryland, in 1998, and his MEng, MS, and PhD degrees in electrical engineering from Cornell University, Ithaca, New York, in 2000, 2003, and 2005, respectively. From 2005 to 2006, he was a postdoctoral research associate in the Department of Psychology at Cornell. He is currently an associate professor in the School of Electrical and Computer Engineering at Oklahoma State University, where he heads the image coding and analysis laboratory. His research interests include image processing, data compression, computational vision, natural scene statistics, and visual perception.