C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes

Yi Zhang [a,*], Anush K. Moorthy [b], Damon M. Chandler [a], Alan C. Bovik [b]

[a] School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA
[b] Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

ABSTRACT

It is widely known that the wavelet coefficients of natural scenes possess certain statistical regularities which can be affected by the presence of distortions. The DIIVINE (Distortion Identification-based Image Verity and Integrity Evaluation) algorithm is a successful no-reference image quality assessment (NR IQA) algorithm, which estimates quality based on changes in these regularities. However, DIIVINE operates based on real-valued wavelet coefficients, whereas the visual appearance of an image can be strongly determined by both the magnitude and phase information.

In this paper, we present a complex extension of the DIIVINE algorithm (called C-DIIVINE), which blindly assesses image quality based on the complex Gaussian scale mixture model corresponding to the complex version of the steerable pyramid wavelet transform. Specifically, we applied three commonly used distribution models to fit the statistics of the wavelet coefficients: (1) the complex generalized Gaussian distribution is used to model the wavelet coefficient magnitudes, (2) the generalized Gaussian distribution is used to model the coefficients' relative magnitudes, and (3) the wrapped Cauchy distribution is used to model the coefficients' relative phases. All these distributions have characteristic shapes that are consistent across different natural images but change significantly in the presence of distortions. We also employ the complex wavelet structural similarity index to measure degradation of the correlations across image scales, which serves as an important indicator of the subbands' energy distribution and the loss of alignment of local spectral components contributing to image structure. Experimental results show that these complex extensions allow C-DIIVINE to yield a substantial improvement in predictive performance as compared to its predecessor, and highly competitive performance relative to other recent no-reference algorithms.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

A crucial task for any system that processes images for human viewing is the ability to assess the quality of each image in a manner that is consistent with human judgments of quality. To address this need, numerous algorithms for image quality assessment (IQA) have been developed and refined over the past several decades using a wide variety of image-modeling techniques. IQA algorithms have been successfully used in applications such as image and video coding (e.g., [1–4]); unequal error protection (e.g., [5]); image synthesis (e.g., [6,7]); and in numerous other areas (e.g., [8–11]).

The vast majority of IQA algorithms are so-called *full-reference* algorithms, which take as input both a distorted image and a reference image, and yield as output an estimate of the quality *difference* between the two images. The simplest approach to full-reference (FR) IQA is to measure local pixelwise differences, then collapse these

local measurements into a scalar which represents the overall quality difference; e.g., the mean-squared error (MSE) or peak signal-to-noise ratio (PSNR). More sophisticated FR IQA methods have employed a wide variety of approaches ranging from estimating quality differences based on weighted MSE/PSNR variants (often measured in different domains; e.g., [12]), to estimating quality differences based on models of the human visual system (e.g., [13–19,3,20–22,4,23–26]), and estimating quality based on various feature-extraction-based or information-theoretic-based approaches (e.g., [27–33]).

FR IQA provides a useful and effective way to evaluate quality differences; however, in many cases, the reference image or even partial information about the reference image is not available (partial information may be used for reduced-reference IQA; see, e.g., [34–36]). Although humans can often effortlessly judge the quality of a distorted image in the absence of a reference image, the no-reference QA task has proven to be quite challenging from a modeling perspective. No-reference IQA models attempt to perform this task, i.e., to estimate the quality of a distorted image without a corresponding reference image. The advantages of a no-reference (NR) approach are numerous: in an IP streaming application, for example, only the compressed (distorted) image is received, and thus quality judgments must be made without access to the reference. Similarly, in digital photography, it is often desirable to determine the quality of the captured image relative to the original scene; this scenario requires a NR approach because the original scene cannot be provided to the IQA method.

The vast majority of NR IQA algorithms attempt to detect specific types of distortion such as blurring, blocking, ringing, or various forms of noise (e.g., [37–42]). For example, algorithms for sharpness/blurriness estimation have been shown to perform well in NR IQA of blurred images. The vast majority of sharpness/blurriness estimators operate under the assumption that the appearance of edges is affected by blur, and accordingly these methods estimate sharpness/blurriness by using various edge-appearance models (e.g., [43–45]). Other methods have used spectral information to estimate sharpness (e.g., [46,47]), whereas more recent hybrid approaches employ a combination of edge-based and transform-based methods. NR IQA algorithms have also been designed specifically for JPEG or JPEG2000 compression artifacts (e.g., [37–39]). Such JPEG-specific algorithms generally employ detectors for blocking and blurring, which are often combined with measures of visual masking to estimate the visibility of each of these artifacts, and thereby estimate quality. Similar NR IQA algorithms have been designed specifically for JPEG2000 ringing, blurring, and aliasing artifacts (e.g., [40,37–39]). Some NR algorithms have employed combinations of these aforementioned measures, supplemented with noise measures and/or other measures for degradations of other visual features (e.g., [41,48,42]).

Other NR IQA algorithms take a more distortion-agnostic approach. For example, a training/learning-based approach has been recently developed which extracts Gabor-filter-based features from local image patches and then learns the mapping from the quantized feature space to image quality by using a visual-codebook-based method [49,50]. Other approaches estimate image quality based on the extent to which the statistical properties of the distorted image deviate from those of natural images.[1] Natural scenes have been studied extensively over the last two decades and these studies have revealed that such images have a large number of statistical regularities (e.g., [51–53]). Distortions can lead to deviations in these statistical regularities, and thus it is possible to estimate quality by quantifying these deviations. For example, in [54], the authors developed a NR IQA algorithm which estimates quality by using the Renyi entropy to measure deviations in anisotropy along various orientations. In [55], the authors developed a NR IQA algorithm (BLIINDS) which estimates quality based on deviations in DCT statistics (e.g., changes in the characteristic shape/symmetry of DCT coefficient histograms). Most recently, a spatial-domain-based NR IQA model was proposed in [56], which uses scene statistics of locally normalized luminance coefficients to quantify possible losses of 'naturalness' in the image due to the presence of distortion, leading to a holistic measure of quality.

One particular NR IQA algorithm, DIIVINE (Distortion Identification-based Image Verity and INtegrity Evaluation) [57], — the algorithm which the method proposed in this paper extends — employs a two-stage framework for estimating quality based on natural-scene statistics. DIIVINE estimates quality by using statistical features which are generally consistent across reference images, but which change in the presence of distortion. In this way, it is possible to compute the extent to which the statistical features in the distorted image deviate from these expected natural statistical features, and then to use these deviations as proxy measures of quality degradations.

In [57], a steerable pyramid decomposition is first applied to obtain a multi-scale, multi-orientation representation of the distorted image [58]. The real-valued coefficients are then processed through a divisive normalization operation. As demonstrated in [57], the histograms of these normalized coefficients exhibit Gaussian-like profiles which are generally consistent across natural images. The coefficients were also shown to exhibit strong correlations between spatially co-located/neighboring coefficients from different scales and orientations. From the distorted image, 88 statistical features are measured in [57]. These 88 statistical features are then used to estimate quality via the following two stages: The first stage performs distortion identification. In this stage, the statistical features extracted from the distorted image are fed to a classifier to estimate the probability that the image is afflicted by one of the multiple distortion types. The second stage performs distortion-specific quality assessment. In this stage, the same statistical features are used to estimate the distortion-specific quality of the image. Specifically, a regression model for each distortion

---

[1] In this paper, we consider natural images to be photographic images containing subject matter that may occur during normal photopic or scotopic viewing conditions.

type is used to map the statistical features to quality estimates based on the probabilities estimated in the first stage. DIIVINE has been shown to perform remarkably well in estimating quality. It is one of the best-performing NR IQA methods available, its ideas have given rise to a related pixel-based NR IQA method [56], and it has been shown to be competitive with top-performing FR IQA algorithms.

In this paper, we present an extended version of the DIIVINE algorithm which operates based on the local magnitude and phase statistics of complex wavelet image coefficients. Over the last three decades, considerable insights into the properties of visual systems have been gained by considering the statistics of natural scenes (e.g., [51,59,60,52,53,61]). These approaches have demonstrated that many basic properties of the early visual system (both selectivity and tiling of visual neurons) and properties of visual perception can be linked to the statistics of natural scenes. Natural scenes exhibit a characteristic magnitude spectrum which generally follows a $f^{-\alpha}$ trend, where f denotes radial spatial frequency [51]. Estimates of the parameter α for any given scene population typically vary from 0.7 to 1.5 with averages in the range of approximately 1.1 [59,62]. Natural scenes also posses a coherent phase structure which has been shown to be the primary contributor to an image's phenomenal appearance. Oppenheim and Lim [63] first demonstrated this fact by synthesizing an image from the magnitude spectrum of one image and phase spectrum of another; the resulting image appeared much more similar to the image whose phase structure was used. Thomson, Foster, and Summers [64] have demonstrated that randomization or quantization of this phase structure severely impacts the semblance of an image. Bex and Makous [65] have shown that randomizing a natural image's phase structure at a particular spatial scale decreases detection and contrast-matching performance by the same amount as removing the spatial scale altogether. In addition, Geisler et al. [66] have demonstrated that human performance in detecting contours can be predicted via a model based on the edge co-occurrence statistics of natural images.

Yet, although the phase spectrum of full-sized images has long been argued to be much more perceptually important than the magnitude spectrum, the individual contributions of magnitude and phase have been shown to vary according to scale. Morgan et al. [67] demonstrated that for larger image patches, the perceived image structure is well described by the phase spectrum, whereas for small image patches, the magnitude spectrum dominates. More recently, Field and Chandler [68] argued that the scale-dependent importance of magnitude vs. phase is due to the sparse structure of images and the nature of the information in small patches. Specifically, most small image patches contain blank regions, single edges, or bits of texture. The power spectra for these small patches can be quite informative regarding which of these classes are present. However, larger image patches will typically contain a significant number of edges as well as textures and blank regions. For these larger patches, the phase spectrum is determined by the relative combination and positions of these features, and thus the phase spectrum will begin to play a larger role in determining the image's appearance.

Here, we present a complex extension of DIIVINE, called *C-DIIVINE*, which extends its predecessor by estimating quality based on changes in local magnitude and phase statistics. To demonstrate the importance of phase statistics in predicting image quality, Fig. 1(a) shows two JPEG-compressed images from the TID database [69], which have different visual qualities. When a real-valued steerable pyramid wavelet decomposition is employed, the two images have almost the same subband coefficient histograms after divisive normalization. When using a complex version of the wavelet decomposition, the magnitude of the complex-valued subband coefficients still has similar distributions (see Fig. 1(b)). However, their phase information can be distinguishable. For some subbands (e.g., 90° orientation subband), the phase distributions are the same, but for others (e.g., 30° and 150° orientation subbands), they can be significantly different. Therefore, the proposed C-DIIVINE algorithm, which analyzes both the magnitude and the phase statistics, can predict the quality of these two images quite well, whereas DIIVINE cannot.

Numerous studies have shown that the neurons in primary visual cortex (V1) are organized in micro-columns of similar frequency and orientation, and approximately, divided into several central frequencies (on a log-scale) and multiple orientations (between 0° and 180°). Thus, the primary visual area V1 functions like filters/filter banks, and its response has been widely modeled by Gabor filters with specific central frequencies and orientations [70–74]. Here, to mimic the responses of V1 area, C-DIIVINE first applies a complex steerable pyramid decomposition to the distorted image to obtain a complex-valued multi-scale, multi-orientation representation. As in DIIVINE, the coefficients are then processed through a divisive normalization operation, and the normalized complex coefficients are modeled by using a complex Gaussian scale mixture [75]. C-DIIVINE then measures separate magnitude- and phase-specific versions of a subset of the statistical features used in DIIVINE, including the use of a complex version of SSIM (CW-SSIM) [76] to measure correlations across scales. These obtained statistical features are finally fed into combined frameworks to estimate the image's quality.

The main contributions of this work are as follows: (1) C-DIIVINE analyzes distorted images by using a complex version of the steerable pyramid wavelet transform, and the obtained complex wavelet coefficients are utilized to blindly assess image quality based on a complex Gaussian scale mixture (CGSM). We show in this work that distortions affect both the magnitude and phase statistics of the wavelet coefficients. (2) By using the CGSM, we propose to model the complex wavelet coefficient statistics in three ways: (a) the coefficient magnitudes are characterized by the complex generalized Gaussian distribution; (b) the coefficient relative magnitudes are characterized by the generalized Gaussian distribution; and (c) the coefficient relative phases are characterized by the wrapped Cauchy distribution. We also employ the CW-SSIM index to measure the coefficient correlation between different wavelet scales. (3) Based on (2), three types of quality-aware statistical features are extracted: magnitude-based features, phase-based features, and across-scale correlation features. These features are then
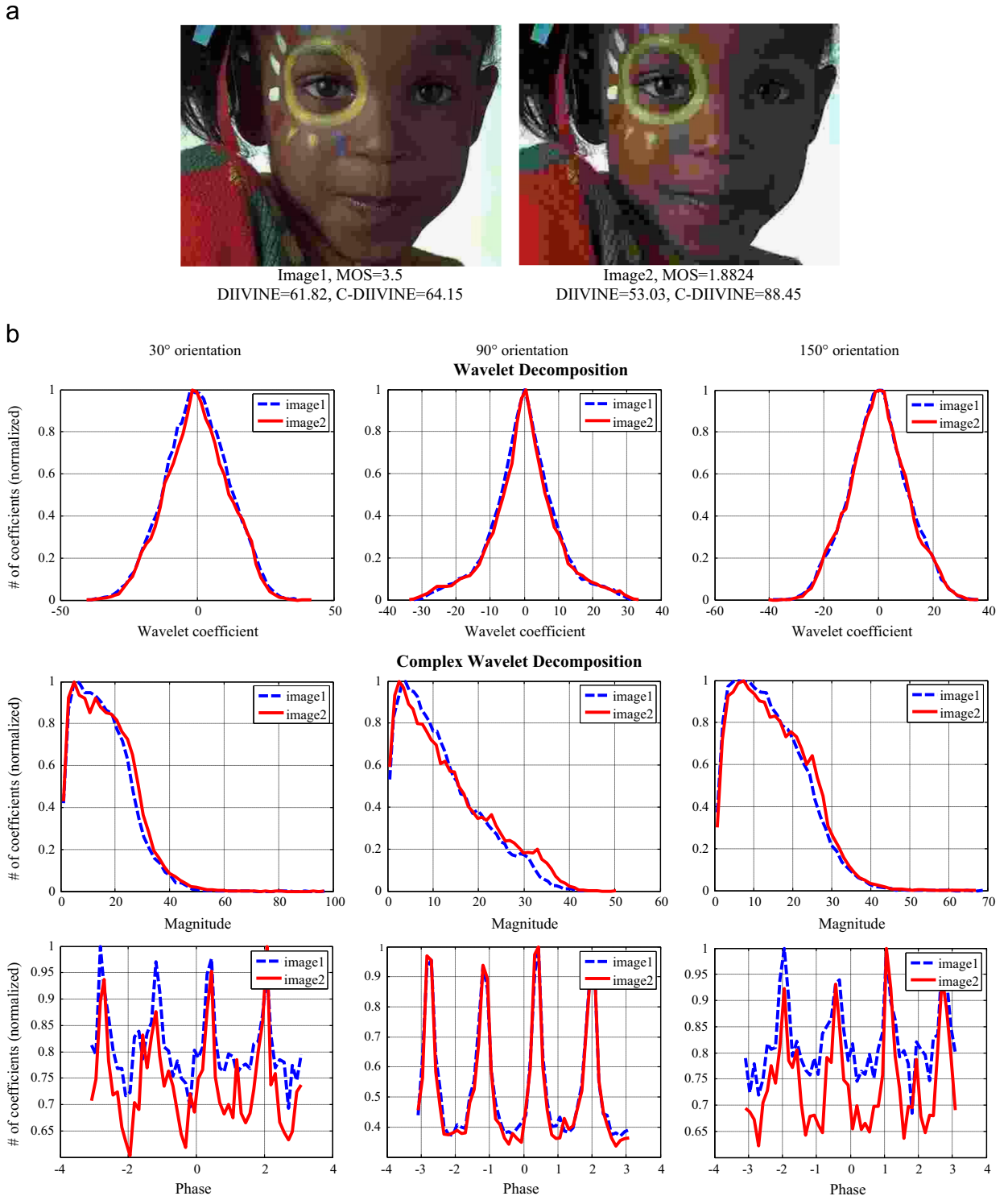
a

Image1, MOS=3.5
DIIVINE=61.82, C-DIIVINE=64.15

Image2, MOS=1.8824
DIIVINE=53.03, C-DIIVINE=88.45

b

Fig. 1. Two JPEG-compressed images have almost the same wavelet coefficient (magnitude) histograms, but different phase information. Notice that both images are decomposed at two scales and six orientations. The histograms plotted correspond to the coarser scale. Similar results are observed at the finer scale.

utilized to evaluate image quality based on a combined framework which consists of the one-stage framework used in [56] and the two-stage framework used in [57]. (4) We demonstrate in this work that these aforementioned

extensions allow C-DIIVINE not only to yield a substantial improvement in NR IQA performance as compared to its predecessor, but also to challenge some state-of-the-art IQA algorithms.
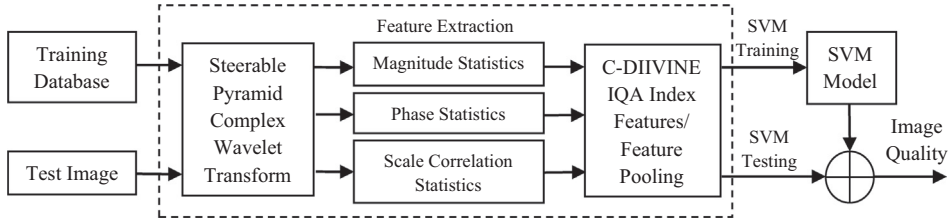
Fig. 2. A block diagram of the C-DIIVINE algorithm. Note that the algorithm contains two parts: training and testing. SVM models (classification and regression model) are obtained through the training process and these models are then used to evaluate qualities of the test images by performing a combined-stage image quality prediction task.

This paper is organized as follows: Section 2 provides details of the C-DIIVINE algorithm, including an introduction to the complex Gaussian scale mixture model, and details of the magnitude and phase statistics employed. In Section 3, we analyze and discuss the performance of C-DIIVINE on various image-quality databases; this section also includes a discussion of the individual contributions of the magnitude-, phase-, and CW-SSIM-based features to the overall performance. General conclusions are presented in Section 4.

2. Algorithm

The first step in C-DIIVINE is to apply a complex steerable pyramid decomposition to the distorted image to obtain complex-valued subband coefficients from which the statistical features are measured. Next, three types of features are measured: (1) magnitude-based features, (2) phase-based features, and (3) across-scale correlation features. Finally, these features are trained by using support vector machine (SVM) learning to estimate image quality. Fig. 2 shows a block diagram of the algorithm. In this section, we provide details of each of the steps.

2.1. Complex steerable pyramid coefficient magnitudes and phases

As shown in Fig. 2, the input (distorted) image is first decomposed via a wavelet-based decomposition. The original DIIVINE algorithm employed a steerable pyramid decomposition to obtain real-valued coefficients from which the statistical features were measured. In C-DIIVINE, a complex steerable pyramid is employed to obtain complex-valued coefficients from which the magnitude- and phase-based statistical features are measured.

Let $\mathbf{I}$ denote the input (distorted) grayscale image,[2] and let $\{\mathbf{Z}_{s,o}\}$ denote a set of complex steerable pyramid subbands indexed by scale s and orientation o. We apply the complex steerable pyramid decomposition using three scales $s \in [1, 3]$ and six orientations $o \in [0°, 30°, 60°, 90°, 120°, 150°]$, resulting in a total of 18 subbands. (The same six orientations and only two scales are used in [57].)

Let $z_{s,o}(x, y)$ denote a complex-valued coefficient located at spatial position (x,y) in subband $\mathbf{Z}_{s,o}$. The magnitude and phase of $z_{s,o}(x, y)$, denoted by $|z_{s,o}(x, y)|$ and $\angle z_{s,o}(x, y)$, respectively, are given by

$$|z_{s,o}(x,y)| = \sqrt{\Re\{z_{s,o}(x,y)\}^2 + \Im\{z_{s,o}(x,y)\}^2} \tag{1}$$

$$\angle z_{s,o}(x,y) = \tan^{-1}\left(\frac{\Im\{z_{s,o}(x,y)\}}{\Re\{z_{s,o}(x,y)\}}\right) \tag{2}$$

where $\Re\{z_{s,o}(x,y)\}$ and $\Im\{z_{s,o}(x,y)\}$ denote the real and imaginary components of $z_{s,o}(x,y)$, respectively.

The *relative magnitude* among four neighboring complex-valued coefficients $z_{1,1}, z_{1,2}, z_{2,1}$, and $z_{2,2}$, denoted by ψ, is given by $\psi = |z_{1,1}| + |z_{2,2}| - |z_{1,2}| - |z_{2,1}|$. The *relative phase* between two complex-valued coefficients z_1 and z_2, denoted by ϕ, is given by $\phi = \angle z_1 - \angle z_2$. Therefore, for each coefficient $z_{s,o}(x, y)$, we measure the relative magnitude which is given by

$$\psi_{s,o}(x,y) = |\hat{z}_{s,o}(x,y)| + |\hat{z}_{s,o}(x+1,y+1)| \\ - |\hat{z}_{s,o}(x+1,y)| - |\hat{z}_{s,o}(x,y+1)|, \tag{3}$$

where $\hat{z}_{s,o}(x,y)$ is the divisively normalized version of $z_{s,o}(x,y)$ (see Section 2.2). We also measure the relative phase between horizontal and vertical spatial neighbors, resulting in a horizontal relative phase $\phi_{s,o}^{horz}(x,y)$ and a vertical relative phase $\phi_{s,o}^{vert}(x,y)$, which are given by

$$\phi_{s,o}^{horz}(x,y) = \angle z_{s,o}(x,y) - \angle z_{s,o}(x+1,y) \tag{4}$$

$$\phi_{s,o}^{vert}(x,y) = \angle z_{s,o}(x,y) - \angle z_{s,o}(x,y+1). \tag{5}$$

The relative phases computed by Eqs. (4) and (5) for each coefficient fall in the range $[-2\pi, 2\pi]$, whereas the distribution we use to model the relative phase statistics requires $-\pi \leq \phi \leq \pi$ (see Eq. (12)). Thus, we map the relative phase as follows:

$$\hat{\phi}_{s,o}^{horz/vert} = \begin{cases} \phi_{s,o}^{horz/vert}, & -\pi \leq \phi_{s,o}^{horz/vert} \leq \pi \\ \phi_{s,o}^{horz/vert} + 2\pi, & -2\pi \leq \phi_{s,o}^{horz/vert} < -\pi. \\ \phi_{s,o}^{horz/vert} - 2\pi, & \pi < \phi_{s,o}^{horz/vert} \leq 2\pi \end{cases} \tag{6}$$

2.2. Divisive normalization

Divisive normalization was proposed in the literature of neural physiology to account for the nonlinear behavior of certain cortical cells and the contrast-gain-control mechanism observed in early visual processing (see [77,53,4]). For

---

[2] Color images are converted to grayscale via the pointwise transformation $\mathbf{I} = 0.2989 \times \mathbf{R} + 0.5870 \times \mathbf{G} + 0.1140 \times \mathbf{B}$, where $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$ denote, respectively, the red, green, and blue color components of the image. Other grayscale transformations such as just using the $\mathbf{G}$ plane or using $\mathbf{R} + \mathbf{G} + \mathbf{B}$ yield similar predictive performances.

example, it has been shown in [78] that an additional local gain-control divisive normalization process allows for a more complete explanation of the striate cell responses when interpreted by the conventional linear/energy model, which assumes that the simple cells in the striate cortex act like half-wave-rectified linear operators, and the complex cells act like energy mechanisms, constructed from linear subunits. In computational vision science, a long-standing view of the purpose of early visual sensory processing is the increased statistical independence between neuronal responses. The divisive normalization process has been shown to successfully reduce the statistical dependencies between subbands, producing approximately Gaussian marginal distributions for the wavelet coefficients [79], and this technique has been explicitly used for RR IQA in [80].

We applied the same technique to the complex wavelet coefficients (following [57]). As demonstrated in [75], the real and imaginary parts of the neighboring complex subband coefficients can have the same scalar multiplier if either one follows the Gaussian scale mixture, and by dividing an estimated scalar multiplier (i.e., divisive normalization), the wavelet coefficient magnitudes can be characterized by a complex generalized Gaussian distribution (see Fig. 5). We show that distortions affect the histogram shapes of these normalized magnitudes, whereas the natural images maintain almost the same magnitude histogram profiles.

The divisively normalized version of each coefficient, denoted by $\hat{z}_{s,o}$, is obtained via

$$\hat{z}_{s,o}(x,y) = \frac{z_{s,o}(x,y)}{\sqrt{\frac{1}{N}\mathbf{P}^T C_{\mathbf{P}_{all}}^{-1}\mathbf{P}}}, \tag{7}$$

where $\sqrt{(1/N)\mathbf{P}^T C_{\mathbf{P}_{all}}^{-1}\mathbf{P}}$ is a real-valued scalar representing the combined response of the normalization pool (neighboring coefficients). Note that, because $\sqrt{(1/N)\mathbf{P}^T C_{\mathbf{P}_{all}}^{-1}\mathbf{P}}$ is a real-valued scalar, both $\Re\{z_{s,o}(x,y)\}$ and $\Im\{z_{s,o}(x,y)\}$ are divided by this same value. Accordingly, the magnitude $|z_{s,o}(x,y)|$ is affected by the divisive normalization, whereas the phase $\angle z_{s,o}(x,y)$ (and thus the relative phase) is not.

The vector $\mathbf{P}$ in Eq. (7) is a 15-element vector consisting of the magnitudes of neighboring coefficients in space, scale, and orientation. Specifically, $\mathbf{P}$ contains nine coefficient magnitudes from a 3×3 spatial neighborhood around $z_{s,o}(x,y)$, one coefficient magnitude from the corresponding spatial location in the parent band, and five coefficient magnitudes from other orientations at the same spatial location and scale. This is the same neighborhood employed in [57], here using the coefficient magnitudes rather than the real values. The quantity $C_{\mathbf{P}_{all}} = E\{\mathbf{P}_{all}\mathbf{P}_{all}^T\}$ is the covariance matrix of the vector $\mathbf{P}_{all}$; this latter vector is composed of the magnitudes of all of the coefficients.

2.3. Statistical models of magnitude and phase

As mentioned in Section 1, the key idea employed by DIIVINE is to use statistical features which are generally consistent across reference images, but which change in the presence of distortion. In this way, it is possible to measure deviations in these expected (natural) statistical features as proxy measures of quality degradations.

In [57], it was shown that the real-valued subband coefficients of natural images (i.e., undistorted, reference images), following divisive normalization, exhibit consistent Gaussian-like histograms (marginal probability densities). For distorted images, the marginal densities of the real-valued subband coefficients were markedly and consistently more Laplacian-like. Thus, in [57], features corresponding to the shapes of the marginal densities of the subband coefficients were used to estimate quality.

Here, we extend this idea to the complex domain by examining changes in the marginal densities of the magnitudes, relative magnitudes, and relative phases of the complex-valued subband coefficients. Fig. 3 shows five images from the LIVE image database [81] which we will use to demonstrate consistency in the shapes of the marginal densities of the magnitudes and phases. In Sections 2.4 and 2.5, we will use four of them and their five distorted versions (a subset is shown in Fig. 4) as examples to illustrate behavior of the features used in the C-DIIVINE algorithm. Fig. 5(a) and (b) shows, for each reference image, the histograms of the magnitudes of the coefficients (before and after divisive normalization) from all orientations at the finest scale, i.e., $\{|z_{1,o}|\}$ and $\{|\hat{z}_{1,o}|\}$, $\forall o$. Fig. 6 shows the histograms of (a) relative magnitude, and (b) horizontal relative phase of the coefficients from the finest scale of three different orientations, i.e., $\{\psi_{1,30°}, \psi_{1,90°}, \psi_{1,150°}\}$ and $\{\hat{\phi}_{1,30°}^{horz}, \hat{\phi}_{1,90°}^{horz}, \hat{\phi}_{1,150°}^{horz}\}$.

As demonstrated in Figs. 5(a), (b) and Fig. 6, the histograms of the coefficients' magnitudes and phases generally exhibit consistent profiles that are largely independent of the particular reference image used to obtain the coefficients. Comparing the two magnitude histograms in Fig. 5(a) vs. (b), the divisive normalization yields profiles which, as we will demonstrate shortly, can be modeled by using the magnitude probability densities derived from a complex Gaussian scale mixture model.

To illustrate that the subband coefficients of distorted images yield marginal distributions which deviate from these characteristic profiles, Fig. 7 shows five distorted versions of one of the reference images (*sailing* 2). The corresponding histograms of the coefficients' magnitudes, relative magnitudes, and relative phases are shown in Figs. 5(c) and 8, respectively. For the magnitudes, the distortions tend to affect both the widths and the rates-of-decay of the profiles. For the relative magnitudes and relative phases, the distortions tend to affect the peakedness of the profiles. As we will discuss in the following subsections, these natural and distorted magnitude and relative phase histograms can be modeled by using a complex Gaussian scale mixture.

2.3.1. Complex Gaussian scale mixture

To exploit the advantages of complex wavelet transform and the usefulness of the magnitude and phase statistics in the IQA framework, an appropriate model that can handle complex random variables is required. Accordingly, the complex Gaussian scale mixture (CGSM), recently developed by Rakvongthai et al. [75], is an extension of Gaussian scale mixture (GSM) to efficiently model the complex wavelet coefficients, the latter of which has been used to model the marginal and joint
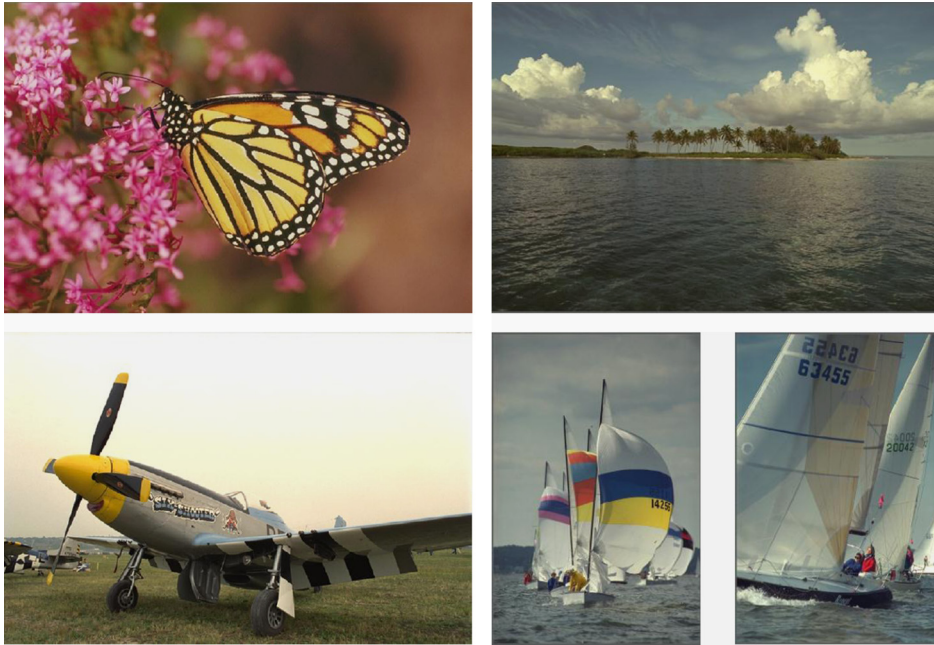
Fig. 3. Five reference images from the LIVE database [81] used to demonstrate the histogram consistency of the subband magnitudes, relative magnitudes, and relative phases across image content. *Top row*: Images *monarch*, *ocean*. *Bottom row*: Images *plane*, *sailing* 2, and *sailing* 3. Among them, four images (*monarch*, *ocean*, *plane*, and *sailing* 3) and their five distorted versions (a subset is shown in Fig. 4) will be used in Sections 2.4 and 2.5 as examples to illustrate feature behavior in the C-DIIVINE algorithm.
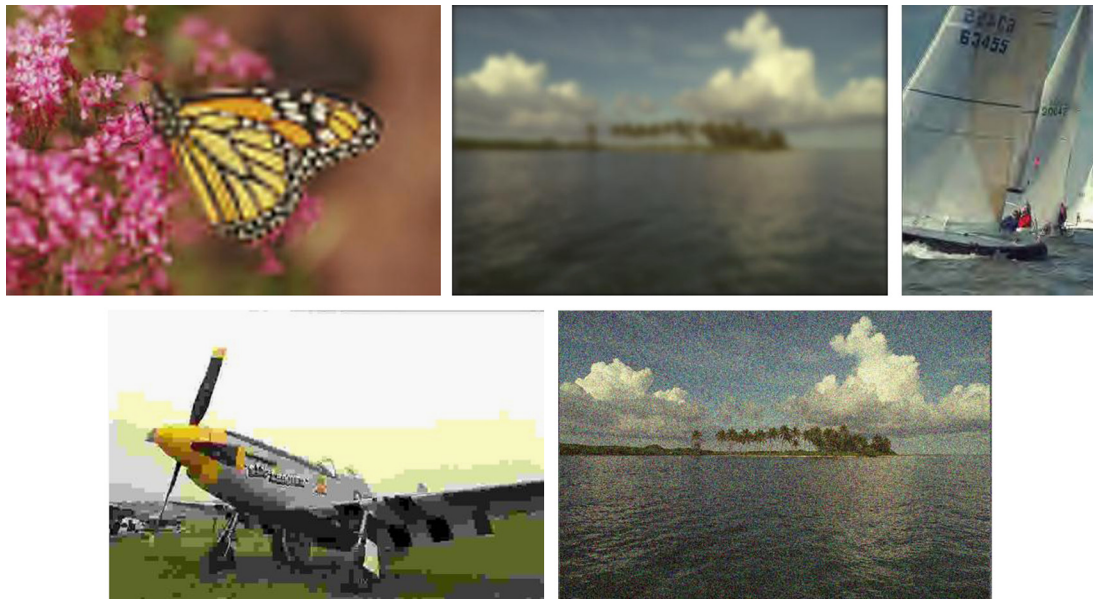

Fig. 4. A subset of the distorted versions of images in Fig. 3. *Top row*: Fast-fading, Gaussian blur, and JPEG2000 compression. *Bottom row*: JPEG compression and white noise.

statistics of the wavelet coefficients of natural images (e.g., [82]).

The GSM is used for real-valued random vector modeling. A real-valued random vector $\mathbf{X}$ with dimension N can be characterized by a GSM if it can be expressed as $\mathbf{X} = \sqrt{S}\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}$ is a zero-mean Gaussian random vector with covariance matrix $C_{\tilde{\mathbf{X}}}$. The multiplier S is an independent positive scalar random variable. If S has unit mean, then the

probability density function (pdf) of $\mathbf{X}$ can be given as

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \frac{\exp\left(-\frac{1}{2}\mathbf{x}^T(sC_{\tilde{X}})^{-1}\mathbf{x}\right)}{(2\pi)^{N/2}\sqrt{sC_{\tilde{X}}}} f_S(s)\, ds. \tag{8}$$

Now consider a complex-valued random vector $\mathbf{Z}$ with dimension N. This vector can be expressed as a CGSM if
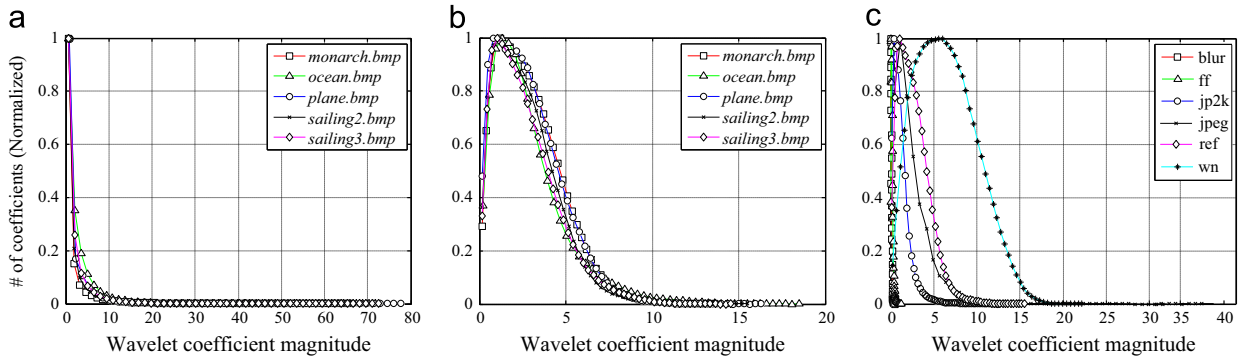
Fig. 5. Histograms of subband magnitudes for each of the five reference images in Fig. 3 and five distorted versions of image *sailing* 2 in Fig. 7. These magnitude histograms were generated by using all coefficients from all six subbands at the finest scale, without divisive normalization (a), and with divisive normalization (b), (c). Notice that the histograms exhibit consistent profiles that are largely independent of the particular reference image from which the magnitudes were computed, but vary significantly in the presence of distortions.
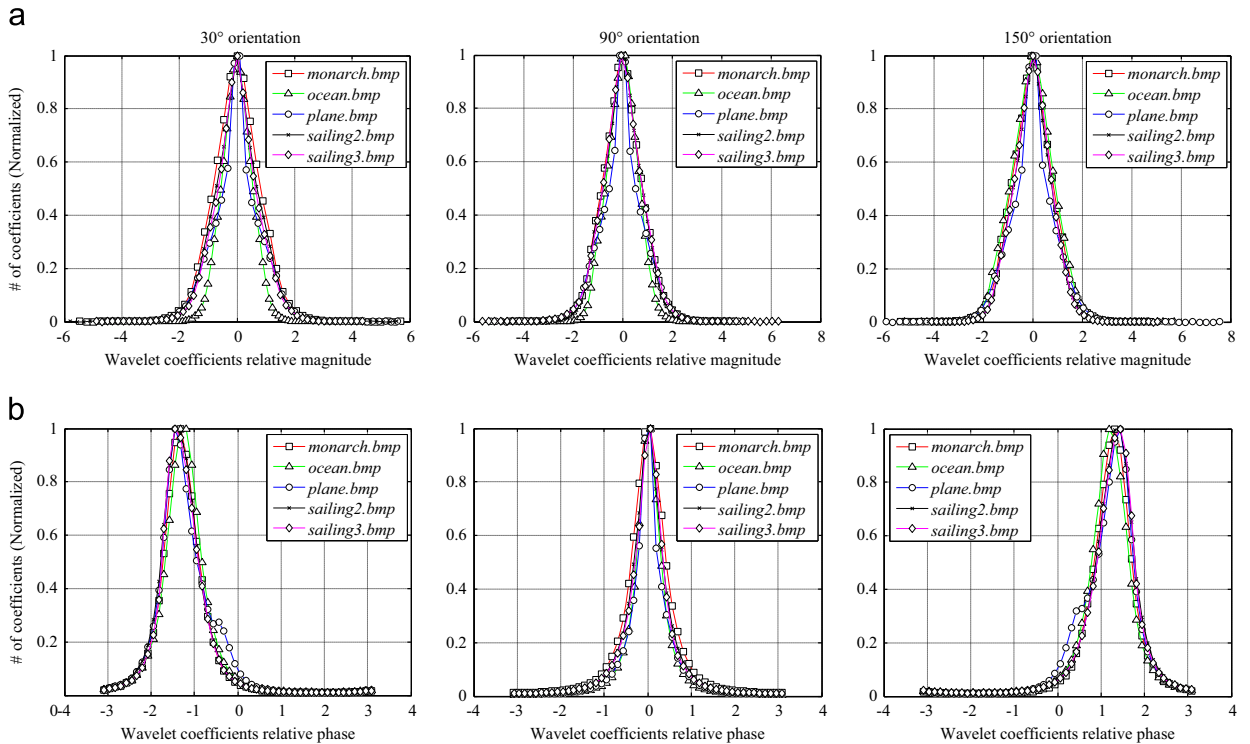

Fig. 6. Histograms of subband relative magnitude (a) and relative phase (b) for each of the five reference images in Fig. 3. These histograms were generated by using subbands of the indicated orientations at the finest scale. Notice that the histograms exhibit consistent profiles that are largely independent of the particular reference image from which the relative magnitudes and relative phases were computed. Also notice that the location of the relative phase histogram peak varies according to the subband's orientation.

$\mathbf{Z} = \Re\{Z\} + j\Im\{Z\} = \sqrt{S}\tilde{\mathbf{Z}}$, where $\tilde{\mathbf{Z}} = \Re\{\tilde{Z}\} + j\Im\{\tilde{Z}\}$ is a zero-mean complex Gaussian random vector, and where the multiplier S is an independent positive scalar random variable. If S has unit mean, then the pdf of $\mathbf{Z}$ can be given as

$$f_{\mathbf{Z}}(\mathbf{z}) = \int_0^\infty \frac{1}{\pi^N |sC_{\tilde{Z}}|} \exp\left(-\mathbf{z}^H (sC_{\tilde{Z}})^{-1}\mathbf{z}\right) f_S(s)\, ds. \qquad (9)$$

where $C_{\tilde{Z}} = E\{\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^H\}$ is the complex covariance matrix. The vector $\mathbf{Z}$ is called a CGSM because of its behavior as a complex Gaussian conditioned on S. Here, we see again the significance of divisive normalization: the definition of CGSM ($\mathbf{Z} = \sqrt{S}\tilde{\mathbf{Z}}$)

theoretically requires a normalization process (divided by $\sqrt{S}$) to make the output ($\tilde{\mathbf{Z}}$) to be a zero-mean complex Gaussian random vector, which can be characterized by the complex generalized Gaussian distribution.

2.3.2. Marginal density of magnitude

As shown by Rakvongthai et al. [75], when $N=1$ in Eq. (9) (i.e., both $\Re\{Z\}$ and $\Im\{Z\}$ are characterized by GSM model), Z is a complex random variable which can be characterized by a complex generalized Gaussian distribution. When Z is expressed in radial form as $Z = |Z|e^{j\angle Z}$, the

Fig. 7. Distorted versions of image *sailing* 2. *From left to right*: Fast-fading, Gaussian blur, JPEG2000 compression, JPEG compression, and white noise.
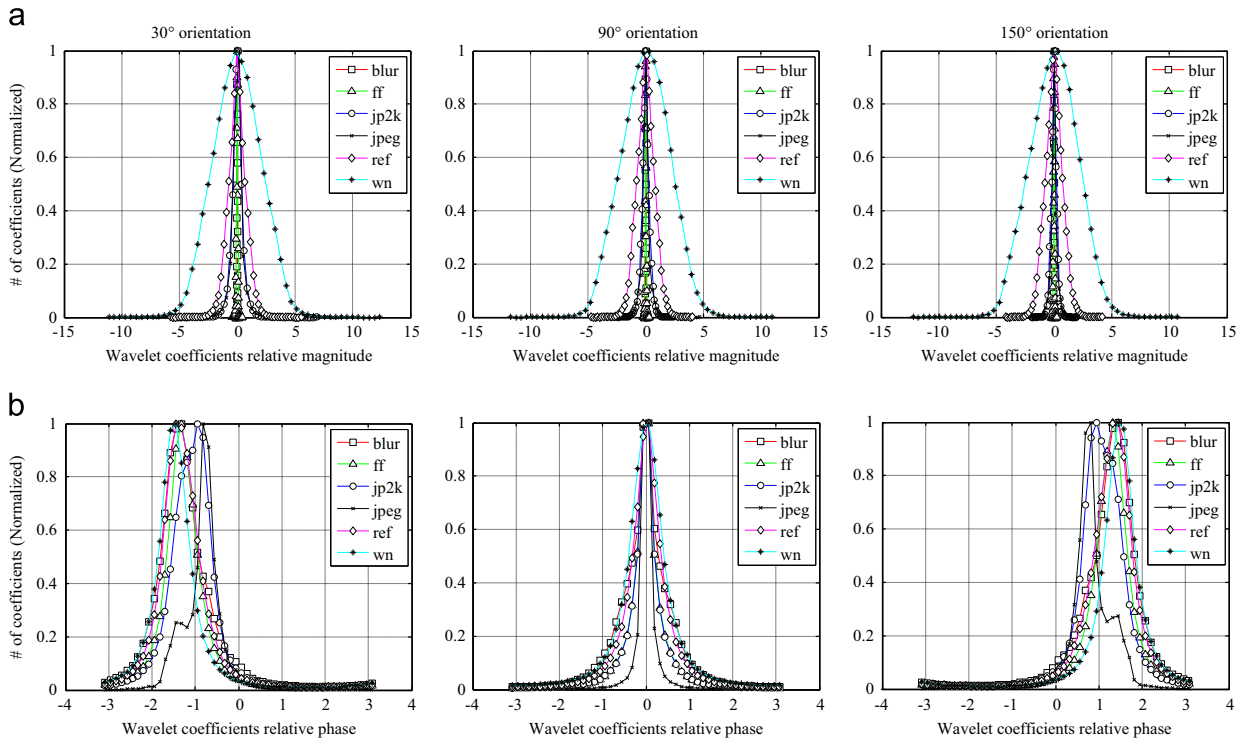

Fig. 8. Histograms of subband relative magnitude (a) and relative phase (b) for each of the five distorted versions of image *sailing* 2 in Fig. 7. The distortions tend to affect the peakedness of the characteristic profile observed for the reference images.

pdf of the magnitude $|Z|$ is given by [75]

$$f_{|Z|}(|z|) = \frac{\beta}{\alpha^2 \Gamma(2/\beta)} |z| \exp\left(-(|z|/\alpha)^\beta \right) \qquad (10)$$

where $\alpha = \sigma\sqrt{2\Gamma(1/\gamma)/\Gamma(2/\gamma)}$ is a scale parameter, and where $\beta = 2\gamma$ is a shape parameter. The parameters $\sigma > 0$ and $\gamma > 0$ are the parameters of the complex generalized Gaussian distribution from which $f_{|Z|}(|z|)$ is derived (see [75] for a derivation).

Fig. 9 shows how the parameters α and β affect the pdfs generated by Eq. (10). When β is fixed, the parameter α (the scale parameter) controls both the width and the location of the peak. When α is fixed, the parameter β

(the shape parameter) controls the rate at which the pdf decays from its peak. By properly selecting these parameters, it is possible to provide decent fits of Eq. (10) to the distorted image's magnitude histograms shown previously in Fig. 5(c).

2.3.3. Marginal density of relative magnitude

Distortions affect not only the coefficient magnitude distribution as a whole, but also the relationship among neighboring pixels. Therefore, we also model the marginal density of the relative magnitude, which was previously defined in Eq. (3). The pdf of the relative magnitude, denoted by random variable Ψ, can be modeled by a
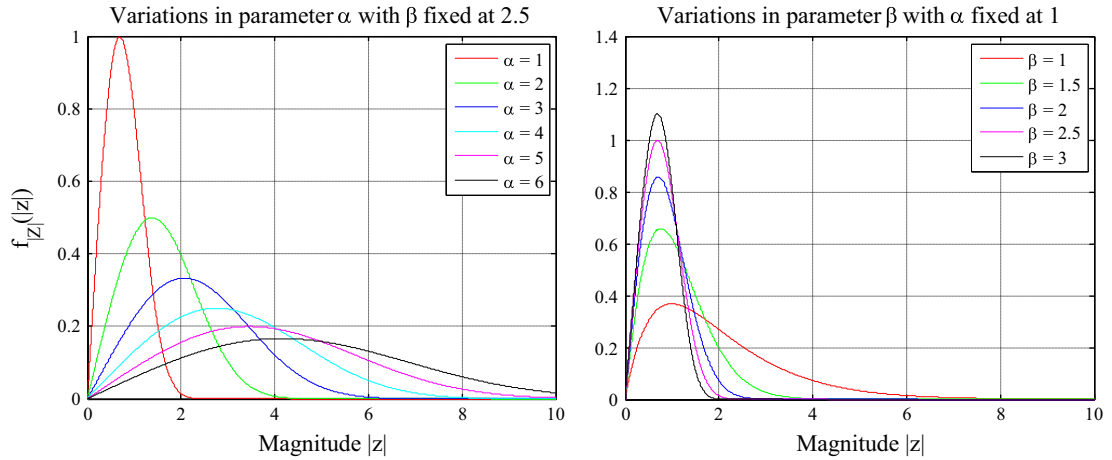
Fig. 9. Probability density functions of the complex generalized Gaussian distribution in Eq. (10). Proper selection of the parameters α and β allows this distribution to be fitted to the histograms shown previously in Fig. 5(c).
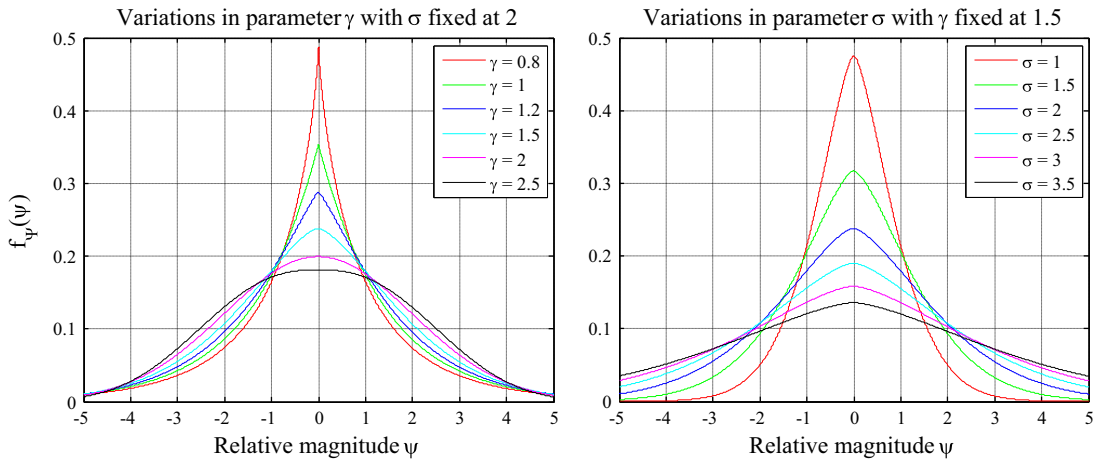
Fig. 10. Probability density functions of the generalized Gaussian distribution in Eq. (11). Proper selection of the parameters γ and σ allows this distribution to be fitted to the histograms shown previously in Fig. 8 (a).

zero-mean generalized Gaussian distribution given by

$$f_\Psi(\psi) = \frac{\gamma}{2\beta\Gamma(1/\gamma)}\exp\big(-(|\psi|/\beta)^\gamma\big) \qquad (11)$$

where $\beta = \sigma\sqrt{\Gamma(1/\gamma)/\Gamma(3/\gamma)}$. The parameter γ controls the general shape of the distribution and σ controls the variance. We estimate the two parameters (γ and σ) of the generalized Gaussian distribution model by using the moment-matching-based approach proposed in [83].

Fig. 10 shows the resulting pdfs obtained from Eq. (11) for various values of γ and σ. Proper selection of these parameters can provide decent fits of Eq. (11) to the distorted image's relative magnitude histograms shown previously in Fig. 8(a).

2.3.4. Marginal density of relative phase

In [75], the authors also provided an expression for the pdf of the relative phase which was derived by using $N=2$ in Eq. (9). Others have modeled the pdf of the relative phase by using a von Mises distribution [84]. Additional details on the pdf of the relative phase can be found in [85].

Here, to facilitate the parameter-fitting procedure (see [86]), we model the pdf of the relative phase, denoted by random variable Φ, using the following two-parameter wrapped Cauchy distribution:

$$f_\Phi(\phi) = \frac{1-\eta^2}{2\pi(1+\eta^2-2\eta\cos(\phi-\mu))}, \qquad -\pi \le \phi \le \pi, \qquad (12)$$

where $\eta = e^{-\gamma}$ is a scale parameter ($0 < \eta \le 1$) and μ is a location parameter.

Fig. 11 shows the resulting pdfs obtained from Eq. (12) for various values of η and μ. The parameter η controls the peakedness. The parameter μ is the mean (location of the peak). Again, by properly selecting these parameters, it is possible to fit Eq. (12) to the distorted image's relative phase histograms shown previously in Fig. 8(b).

2.4. Magnitude- and phase-based features

The statistical models of magnitude and phase described in the previous sections are used to compute the magnitude- and phase-based features that will ultimately be used

Fig. 11. Probability density functions of the two-parameter wrapped Cauchy distribution in Eq. (12). Proper selection of the parameters η and μ allows this distribution to be fitted to the histograms shown previously in Fig. 8 (b).
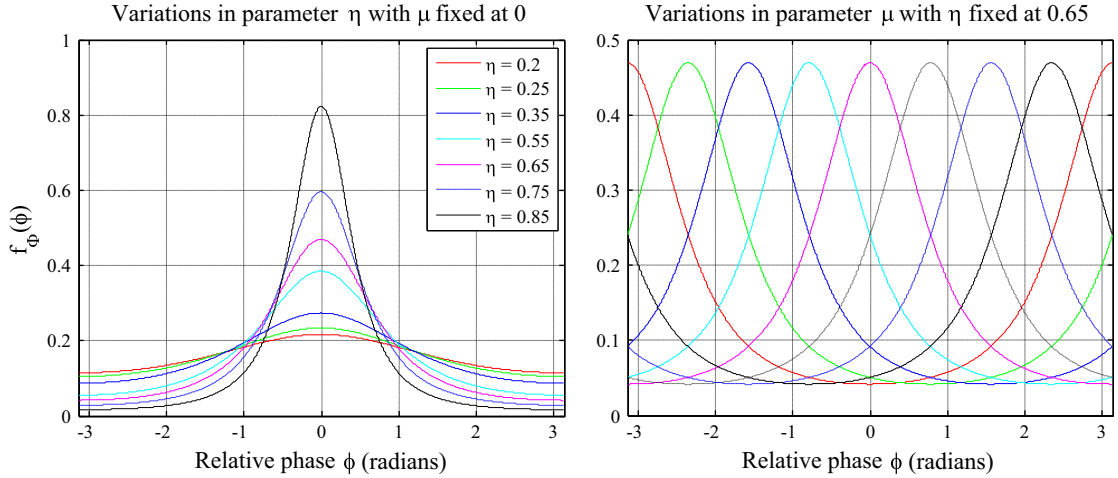
for quality assessment. As mentioned in Section 1, these features are computed at multiple scales and orientations to mimic the cortical decomposition in the human visual system. Also, note that these magnitude- and phase-based features are extracted only from the first and second scale of the wavelet subbands, since most distortion types have a pronounced effect on the higher-frequency components, resulting in considerable magnitude and phase distortions in the high-frequency band. The third scale wavelet subbands can indeed represent distortions to some extent. However, due to its relatively smaller coefficient number, such representations can be imprecise compared to the other two scales, and may even decrease the performance when employed.

Therefore, given an image decomposed by the complex wavelet transform at three scales and six orientations, resulting in a set of 18 subbands $\{Z_{s,o}\}$, we generate 12 vectors containing the coefficient magnitudes, six vectors containing the relative magnitudes, 12 vectors containing the horizontal relative phases, and 12 vectors containing the vertical relative phases for future analysis. The magnitude- and phase-based features are then measured as follows.

2.4.1. Magnitude-based features

Let $|Z_{s,o}|$ and $\Psi_{s,o}$ denote the vectors of magnitude and relative magnitude, respectively, generated by applying divisive normalization to and computing the magnitude/relative magnitude of each coefficient in subband $Z_{s,o}$. Then, we fit the histogram of each vector $|Z_{s,o}|$ and $\Psi_{s,o}$ at the finest scale with their corresponding pdf representatives.

Specifically, we fit $|Z_{s,o}|$ with Eq. (10) to determine the best-fitting parameters $\alpha_{s,o}$ and $\beta_{s,o}$. For the fitting procedure, we employ the maximum-likelihood estimation algorithm proposed in [87]. We fit $\Psi_{s,o}$ with Eq. (11) to determine the best-fitting parameters $\gamma_{s,o}$ and $\sigma_{s,o}$. For the fitting procedure, we employ the moment-matching-based approach proposed in [83]. Because there are six magnitude and relative magnitude vectors at the finest

scale, we obtain six values of $\alpha_{s,o}$, $\beta_{s,o}$, $\gamma_{s,o}$, and $\sigma_{s,o}$ respectively, where $s = 1$ and $o \in [0°, 30°, 60°, 90°, 120°, 150°]$.

We also generate two additional vectors of magnitudes, where each is assembled by grouping all six vectors corresponding to the same scale into a single vector:

$$|Z_s| = [|Z_{s,0°}|, |Z_{s,30°}|, |Z_{s,60°}|, |Z_{s,90°}|, |Z_{s,120°}|, |Z_{s,150°}|],$$

where $s \in [1, 2]$. In this case, we captured part of the information in the lower-frequency bands to fully characterize an image. For both vectors ($|Z_1|$ and $|Z_2|$), we apply the same histogram-fitting procedure to determine the best-fitting parameters α_s and β_s, $s \in [1, 2]$.

Thus, the four resulting magnitude-based feature vectors, $\mathbf{v}_\alpha$, $\mathbf{v}_\beta$, $\mathbf{v}_\gamma$, and $\mathbf{v}_\sigma$, contain the following 28 elements:

$$\mathbf{v}_\alpha = [\alpha_{1,0°}, \alpha_{1,30°}, \alpha_{1,60°}, \alpha_{1,90°}, \alpha_{1,120°}, \alpha_{1,150°}, \alpha_1, \alpha_2],$$
$$\mathbf{v}_\beta = [\beta_{1,0°}, \beta_{1,30°}, \beta_{1,60°}, \beta_{1,90°}, \beta_{1,120°}, \beta_{1,150°}, \beta_1, \beta_2],$$
$$\mathbf{v}_\gamma = [\gamma_{1,0°}, \gamma_{1,30°}, \gamma_{1,60°}, \gamma_{1,90°}, \gamma_{1,120°}, \gamma_{1,150°}],$$
$$\mathbf{v}_\sigma = [\sigma_{1,0°}, \sigma_{1,30°}, \sigma_{1,60°}, \sigma_{1,90°}, \sigma_{1,120°}, \sigma_{1,150°}].$$

Following [57], we use the logarithm of the scale parameter (here $\ln(\alpha)$; in [57] $\ln(\sigma^2)$) rather than the actual scale parameter.

Fig. 12 shows plots of the magnitude-based features for four reference images (shown in Fig. 3) and their distorted versions (shown in Fig. 4), in which the horizontal axis represents the feature index and the vertical axis represents the corresponding feature values. Notice that clustering of features across distortions is independent of image content.

2.4.2. Phase-based features

Let $\Phi_{s,o}^{horz}$ and $\Phi_{s,o}^{vert}$ denote vectors of horizontal and vertical relative phases generated by applying a mapping procedure to the original relative phases computed by Eq. (4) and (5) respectively, of each coefficient in subband $Z_{s,o}$, where $s \in [1, 2]$ and $o \in [0°, 30°, 60°, 90°, 120°, 150°]$. Then, for each of the vectors $\Phi_{s,o}^{horz}$ and $\Phi_{s,o}^{vert}$, we fit their histograms with Eq. (12) to determine the best-fitting parameters $\eta_{s,o}^{horz}$ and $\eta_{s,o}^{vert}$. For the fitting procedure, we
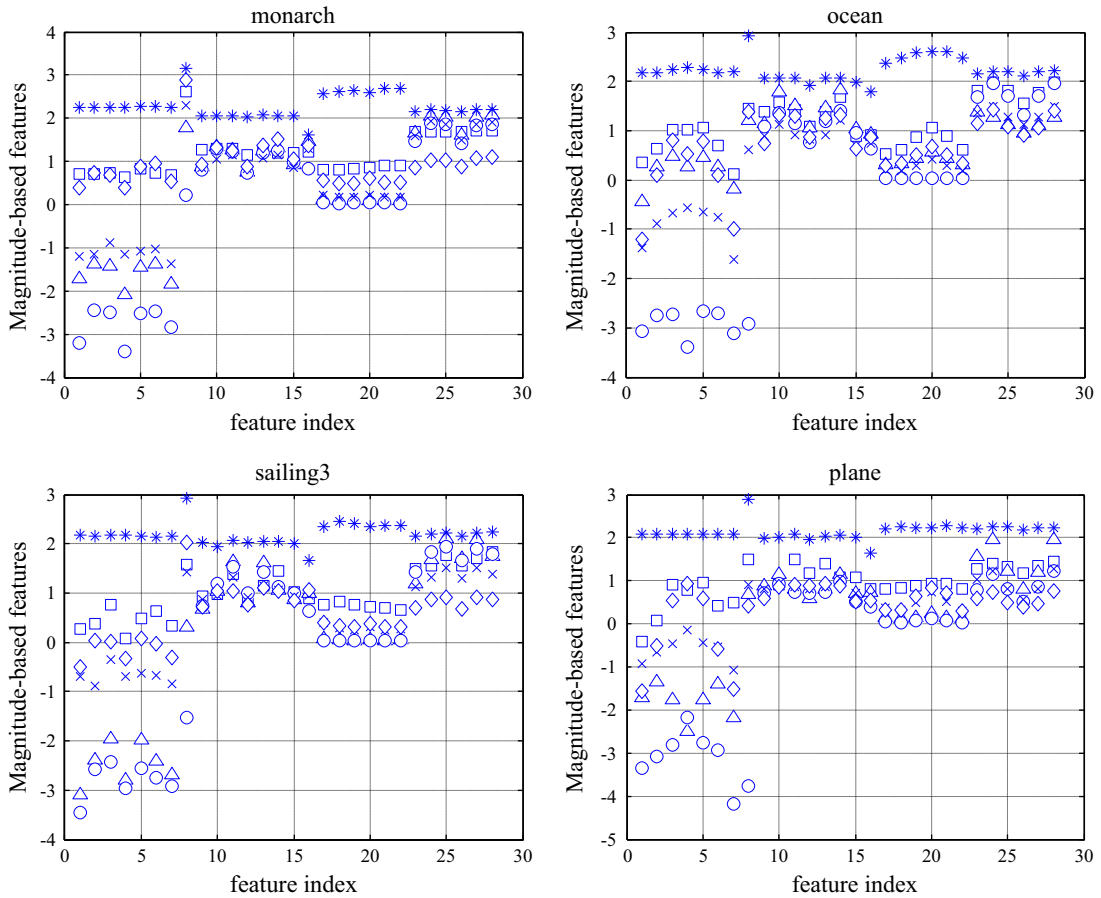
Fig. 12. Plots of the magnitude-based features for four reference images and their distorted versions from the LIVE database [81]. For each figure: reference image (□), fast fading (△), blur (○), JPEG2000 (×), JPEG (◇), and white noise (∗).

employ the technique described in [86]. (The parameter μ in Eq. (12) is determined largely by the orientation of the subband and therefore contributes negligibly toward the distortion identification and quality assessment processes; consequently, we do not store the parameter μ.)

We therefore obtain a single phase-based feature vector, $\mathbf{v}_\eta$, which contains the following 24 elements:

$$\mathbf{v}_\eta = [\eta_{1,0°}^{horz}, \eta_{1,30°}^{horz}, \eta_{1,60°}^{horz}, \eta_{1,90°}^{horz}, \eta_{1,120°}^{horz}, \eta_{1,150°}^{horz},$$
$$\eta_{2,0°}^{horz}, \eta_{2,30°}^{horz}, \eta_{2,60°}^{horz}, \eta_{2,90°}^{horz}, \eta_{2,120°}^{horz}, \eta_{2,150°}^{horz},$$
$$\eta_{1,0°}^{vert}, \eta_{1,30°}^{vert}, \eta_{1,60°}^{vert}, \eta_{1,90°}^{vert}, \eta_{1,120°}^{vert}, \eta_{1,150°}^{vert},$$
$$\eta_{2,0°}^{vert}, \eta_{2,30°}^{vert}, \eta_{2,60°}^{vert}, \eta_{2,90°}^{vert}, \eta_{2,120°}^{vert}, \eta_{2,150°}^{vert}].$$

Fig. 13 shows plots of the phase-based features for four reference images (shown in Fig. 3) and their distorted versions (shown in Fig. 4). The axes of each plot have similar meanings to those of Fig. 12. For the reference images, the values of η tend to be around 0.75. Again, distortion-specific clustering independent of content is observed.

2.5. Across-scale correlation feature

Another feature employed in DIIVINE, which we also use in C-DIIVINE, is a measure of the across-scale

correlations. As argued in [26], degradation of these across-scale correlations can lead to marked reductions in visual quality due to disruptions of the visual system's preference for integrating edges in a coarse-to-fine-scale fashion.

In the original DIIVINE algorithm, a windowed structural correlation (a component of SSIM [27]) was used to measure the across-scale correlations between each subband and the high-pass residual (HPR) band obtained from the steerable pyramid decomposition. In C-DIIVINE, the wavelet subbands contain complex-valued coefficients, which allow the use of a complex-valued structural correlation measure. Specifically, we employ the complex wavelet structural similarity index (CW-SSIM) [76].

The CW-SSIM index, denoted by ρ, between two bands $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is given by

$$\rho(\mathbf{Z}_1, \mathbf{Z}_2) = \left(\frac{2\sum_{i=1}^{N}|z_1(i)||z_2(i)| + K}{\sum_{i=1}^{N}|z_1(i)|^2 + \sum_{i=1}^{N}|z_2(i)|^2 + K} \right)$$
$$\times \left(\frac{2|\sum_{i=1}^{N}z_1(i)z_2(i)^*| + K}{2\sum_{i=1}^{N}|z_1(i)z_2(i)^*| + K} \right), \quad (13)$$

where $z_1(i)$ and $z_2(i)$ are elements of $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively.
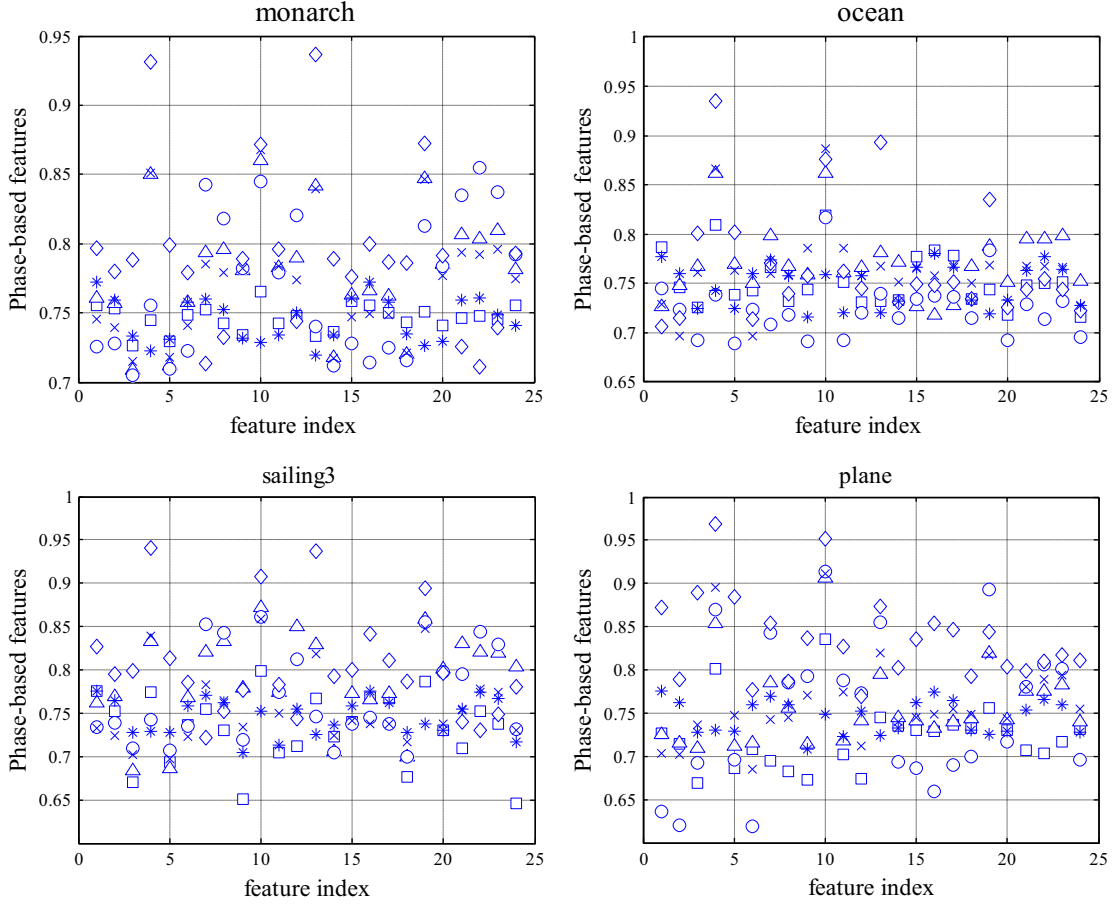
Fig. 13. Plots of the phase-based features for four reference images (shown in Fig. 3) and their distorted versions (shown in Fig. 4) from the LIVE database [81]. For each figure: Reference image (□), fast fading (△), blur (○), JPEG2000 (×), JPEG (◇), white noise (∗).

Eq. (13) consists of the product of two components. As reported in [76], the first component is completely determined by the magnitudes of the coefficients; it is equivalent to the SSIM index [27] applied to the magnitudes of the coefficients. The second component is completely determined by the consistency of the relative phases between $\mathbf{Z}_1$ and $\mathbf{Z}_2$; it achieves the maximum value of 1 when the phase difference between $z_1(i)$ and $z_2(i)$ is constant for all i.

We apply Eq. (13) to measure CW-SSIM between the HPR band and the first- and second-scale subbands. Again, such a choice is made due to the better distortion representations by the higher-frequency wavelet subbands. In addition, we measure CW-SSIM between pairs of subbands with the same orientation (18 pairs). Note that adding this type of features is largely based on the experiments, and is mainly to achieve better performance balance across different databases. The resulting across-scale-correlation-based feature vector, $\mathbf{v}_\rho$, contains the following 30 elements:

$$\mathbf{v}_\rho = [\rho(\mathbf{Z}_{1,0°}, \mathbf{Z}_{2,0°}), \rho(\mathbf{Z}_{1,30°}, \mathbf{Z}_{2,30°}), \rho(\mathbf{Z}_{1,60°}, \mathbf{Z}_{2,60°}),$$
$$\rho(\mathbf{Z}_{1,90°}, \mathbf{Z}_{2,90°}), \rho(\mathbf{Z}_{1,120°}, \mathbf{Z}_{2,120°}), \rho(\mathbf{Z}_{1,150°}, \mathbf{Z}_{2,150°}),$$
$$\rho(\mathbf{Z}_{1,0°}, \mathbf{Z}_{3,0°}), \rho(\mathbf{Z}_{1,30°}, \mathbf{Z}_{3,30°}), \rho(\mathbf{Z}_{1,60°}, \mathbf{Z}_{3,60°}),$$
$$\rho(\mathbf{Z}_{1,90°}, \mathbf{Z}_{3,90°}), \rho(\mathbf{Z}_{1,120°}, \mathbf{Z}_{3,120°}), \rho(\mathbf{Z}_{1,150°}, \mathbf{Z}_{3,150°}),$$
$$\rho(\mathbf{Z}_{2,0°}, \mathbf{Z}_{3,0°}), \rho(\mathbf{Z}_{2,30°}, \mathbf{Z}_{3,30°}), \rho(\mathbf{Z}_{2,60°}, \mathbf{Z}_{3,60°}),$$

$$\rho(\mathbf{Z}_{2,90°}, \mathbf{Z}_{3,90°}), \rho(\mathbf{Z}_{2,120°}, \mathbf{Z}_{3,120°}), \rho(\mathbf{Z}_{2,150°}, \mathbf{Z}_{3,150°}),$$
$$\rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,0°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,30°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,60°}),$$
$$\rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,90°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,120°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{1,150°}),$$
$$\rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,0°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,30°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,60°}),$$
$$\rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,90°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,120°}), \rho(\mathbf{Z}_{HPR}, \mathbf{Z}_{2,150°})].$$

Fig. 14 shows plots of the across-scale-correlation-based features for four reference images and their distorted versions. Again, observe that features extracted in this step are efficient in clustering images with different distortion types, independent of image content.

2.6. Combined-frameworks-based QA

Given the six feature vectors $\mathbf{v}_\alpha$, $\mathbf{v}_\beta$, $\mathbf{v}_\gamma$, $\mathbf{v}_\sigma$, $\mathbf{v}_\eta$, and $\mathbf{v}_\rho$, which contain 82 features in total, C-DIIVINE employs combined frameworks to measure image quality: (1) the one-stage framework used in BRISQUE, and (2) the two-stage framework used in DIIVINE.

As in [56], 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 C-DIIVINE-I. For the two-stage framework, C-DIIVINE features are used to perform (1) distortion
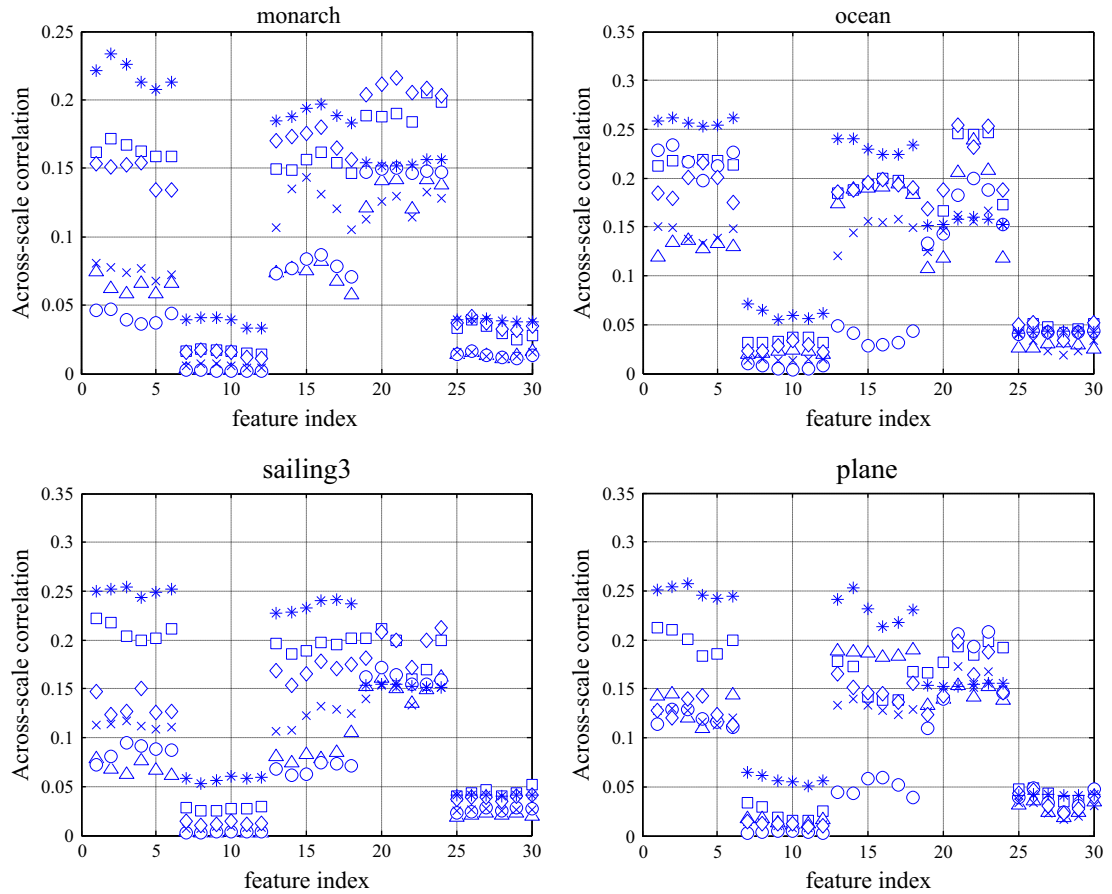
Fig. 14. Plots of the across-scale-correlation-based features for four reference images (shown in Fig. 3) and their distorted versions (shown in Fig. 4) from the LIVE database [81]. For each figure: Reference image (□), fast fading (▵), blur (○), JPEG2000 (×), and JPEG (◇), white noise (∗).

identification, and (2) distortion-specific quality assessment. As in [57], 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 Support Vector Regression (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 the n-dimensional vector of probabilities, and let **q** denote the n-dimensional vector of estimated qualities obtained from the n regression modules. The two-stage framework estimated quality, denoted by C-DIIVINE-II, is computed as follows:

$$\text{C-DIIVINE-II} = \sum_{i=1}^{n} p(i)q(i), \tag{14}$$

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

The advantages and disadvantages for both frameworks are observable. For the two-stage framework, the distortion identification stage can never be completely correct (due to the imperfectness of the extracted features and the complexity of various distortions), and thus it can always produce identification errors. The strategy of using a probability-weighted combination rule can help to compensate for these limitations. However, because of the identification error, it cannot achieve as high a performance as the one-stage framework on the cross-validation test (see Section 3.3). In comparison, the one-stage framework can help to avoid this potential error, but test results on the CSIQ and Toyama databases (see Section 3.4) show that this framework seems to be unreliable when the test images possess statistical properties not completely observed in the training database. Perhaps this is due to different ranges of feature values among the different databases.

Thus, to take advantages of both frameworks, the final estimate of quality combines the minimum and average values of the two frameworks' outputs. This approach, denoted by C-DIIVINE, is computed as follows:

$$\begin{aligned}
\text{C-DIIVINE} &= \tfrac{1}{2}[\min(\text{C-DIIVINE-I, C-DIIVINE-II}) \\
&\quad + \text{avg}(\text{C-DIIVINE-I, C-DIIVINE-II})] \\
&= \tfrac{1}{2}(\text{C-DIIVINE-I} + \text{C-DIIVINE-II}) \\
&\quad - \tfrac{1}{4}|\text{C-DIIVINE-I} - \text{C-DIIVINE-II}|.
\end{aligned} \tag{15}$$

Eq. (15) is supported by the following observations. First, the minimum operation assumes that the qualities of test images that were not 'seen' in the training database, may be predicted to have very different qualities. Since C-DIIVINE was trained on the LIVE database with subjective ratings of degradation expressed in terms of difference mean opinion scores (DMOS) which are positive, the predicted DMOS value exceed the normal range if the feature values extracted from a distorted image go beyond the range of the training database. Thus, the minimum operation acts to limit the predictions, making them more reliable. Second, the average of the two outputs is employed as the simplest linearly weighted prediction that combines the two outputs.

Note that both the SVC and SVR require training, and in our implementation, C-DIIVINE was trained on LIVE using the difference mean opinion scores (DMOS) (see Section 3.3). Thus, smaller values of C-DIIVINE denote predictions of better image quality. As we demonstrate next, training these stages on images from one database can yield excellent predictive performance on similarly distorted images from other databases.

3. Results and analysis

In this section, we analyze C-DIIVINE's ability to estimate quality. For this task, we evaluate the predictive performance of C-DIIVINE on various image quality databases. We also compare the performance of C-DIIVINE to other FR and NR IQA algorithms.

3.1. Image-Quality databases

We applied C-DIIVINE to four publicly available databases of subjective image quality: (1) the LIVE database [81]; (2) the Toyama database [88]; (3) the TID database [69]; and (4) the CSIQ database [89].

The LIVE database [81], from The University of Texas at Austin, USA, contains 29 original images, between 26 and 29 distorted versions of each original image, and subjective ratings of degradation for each distorted image. The database contains five types of distortions: Gaussian blurring (BLUR), additive white noise (WN), JPEG compression, JPEG2000 compression (JP2K), and simulated packet-loss of transmitted JPEG2000-compressed images, which is also known as fast fading (FF). The subjective ratings were collected using a double-stimulus scaling paradigm. LIVE contains a total of 779 distorted images and associated DMOS values.

The Toyama database [88], from the University of Toyama, Japan, contains 14 original images and 12 distorted versions of each original, and subjective ratings of quality for each distorted image (mean opinion scores, MOS values). The database contains two types of distortions: JPEG and JPEG2000 compressed images. The subjective ratings were collected using a single-stimulus scaling paradigm. Toyama contains a total of 168 distorted images and associated MOS values.

The TID database [69], from the Tampere University of Technology, Finland, contains 25 original images, 68 distorted versions of each original image, and subjective ratings of quality for each distorted image (MOS values). Seventeen types of distortions are present in the database, including JPEG and JPEG2000 compression, and Gaussian blurring (see [69] for a full list). The subjective ratings were collecting using a double-stimulus scaling paradigm. TID contains a total of 1700 distorted images and associated MOS values. However, no corresponding standard deviations are provided with these means, which precludes any outlier analysis (see Section 3.2).

The CSIQ database [89], from Oklahoma State University, USA, consists of 30 original images, between 24 and 30 distorted versions of each original, and subjective ratings of degradation for each distorted image (DMOS values). Six types of distortion are used in CSIQ: JPEG compression, JPEG2000 compression, global contrast decrements, additive pink Gaussian noise, additive white Gaussian noise, and Gaussian blurring. The subjective ratings were collected using a multiple-stimulus scaling paradigm. CSIQ contains a total of 866 distorted images and associated DMOS values.

3.2. Algorithms and performance measures

We compared C-DIIVINE with various FR and NR quality assessment methods for which code is publicly available. The three NR methods compared were BLIINDS-II [90], DIIVINE [57], and BRISQUE [56]), all of which operate based on natural-scene statistics and were trained on LIVE. For the cross-validation test on LIVE, we included results of three popular FR IQA algorithms (PSNR [91], SSIM [27], and MS-SSIM [28]) whose results are accessible in the corresponding papers. We also included C-DIIVINE results of using only the one-stage and two-stage frameworks denoted by *C-DIIVINE-I* and *C-DIIVINE-II*, respectively, for comparison. For the cross-database test, two FR IQA algorithms, VIF [29] and MAD [92], were used for comparison.

Three criteria were used to measure the prediction monotonicity and prediction accuracy of each algorithm: (1) the Spearman Rank-Order Correlation Coefficient (SROCC), (2) the Pearson Linear Correlation Coefficient (CC), and (3) the Root Mean Square Error (RMSE) after non-linear regression. As recommended in [93], the SROCC serves as a measure of prediction monotonicity, while the CC and RMSE serve as measures of prediction accuracy. Two additional criteria were used to measure the prediction consistency of each algorithm: (1) outlier ratio (OR), and (2) outlier distance (OD) [92].

A QA algorithm might yield predictions of quality that are nonlinearly related to the actual MOS/DMOS values. Such a nonlinear relationship can lead to a seemingly poor prediction accuracy/consistency when evaluated in terms of CC, RMSE, OR, and OD. As recommended in [93], we accounted for this fact by applying the following four-parameter logistic transform to each algorithm's raw predicted scores before computing the CC, RMSE, OR, and OD:

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp\left(-\frac{x - \tau_3}{|\tau_4|}\right)} + \tau_2, \tag{16}$$

where x denotes the raw predicted score, and where 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 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.

3.3. Training and cross-validation performance on LIVE

As mentioned in Section 2.6, C-DIIVINE employs combined frameworks to predict image quality and each framework requires training. In this paper, we trained C-DIIVINE on the LIVE database by using the realigned DMOS values recommended in [94]. The use of the LIVE database for training facilitates a fairer comparison with other FR and NR algorithms; all of these algorithms have either been trained on LIVE (for NR IQA) or have been extensively tested on LIVE (for FR IQA). As in [57], we used the libSVM package [95] to implement the training process. The parameters of the radial basis function kernel used for both the classification and regression were also optimized by the training process.

We performed two types of training on the LIVE database: (1) training on an 80% subset of the database for cross-validation and (2) training on the entire database to test performance on other databases (see Sections 3.4.2 and 3.4.3). For cross-validation, we randomly selected 80/20% training/testing splits of images from the LIVE database and repeated the train-test procedure 1000 times. We compared with three NR and three FR methods in terms of median SROCC and CC computed over 1000 trials and the values are shown in Table 1. In order to evaluate statistical significance, we also performed a one-sided t-test with a 95% confidence level between SROCC values generated by these algorithms across the 1000 train-test trials. The results are shown in Table 2, in which "1", "0", "-1" indicate that the mean correlation of the row (algorithm) is statistically superior, equivalent, or inferior to the mean correlation of the column (algorithm).

To demonstrate that C-DIIVINE features can be used for different distortion identification, Table 3 shows the median and mean classification accuracy of the classifier 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. 15 plots the mean and standard deviation of SROCC values across these 1000 trials for each algorithm.

According to the cross-validation test results, C-DIIVINE performs statistically the best among all NR IQA algorithms considered here. Not only does it improve upon its predecessor DIIVINE, but also outperforms BLIINDS-II and BRISQUE. Also observed are a slight dip in performance of C-DIIVINE-I, and even more of a dip for C-DIIVINE-II, as compared to C-DIIVINE. Such observations demonstrate that (1) the two-stage framework often produce more prediction errors than does the one-stage framework; and (2) the combined frameworks actually work better than using either of the two individual frameworks alone. Compared with FR IQA methods, C-DIIVINE is more consistent than PSNR and SSIM, but still remains inferior to MS-SSIM, indicating that there is room for further

Table 1
Median SROCC and CC values across 1000 train-test combinations on the LIVE database. Italicized entries denote NR IQA algorithms; others are FR IQA algorithms.

	JP2K	JPEG	WN	BLUR	FF	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*
C-DIIVINE-I	*0.9343*	*0.9398*	*0.9702*	*0.9275*	*0.9072*	*0.9416*
C-DIIVINE-II	*0.9102*	*0.9434*	*0.9778*	*0.9310*	*0.9021*	*0.9376*
C-DIIVINE	*0.9302*	*0.9444*	*0.9760*	*0.9386*	*0.9110*	*0.9444*
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*
C-DIIVINE-I	*0.9471*	*0.9598*	*0.9803*	*0.9329*	*0.9364*	*0.9468*
C-DIIVINE-II	*0.9240*	*0.9553*	*0.9858*	*0.9355*	*0.9237*	*0.9409*
C-DIIVINE	*0.9429*	*0.9593*	*0.9844*	*0.9412*	*0.9345*	*0.9474*

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

Classification accuracy (%)	JP2K	JPEG	WN	BLUR	FF	ALL
Median	88.9	91.7	100.0	93.3	73.3	89.4
Mean	88.8	90.6	99.5	91.3	72.9	88.7

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

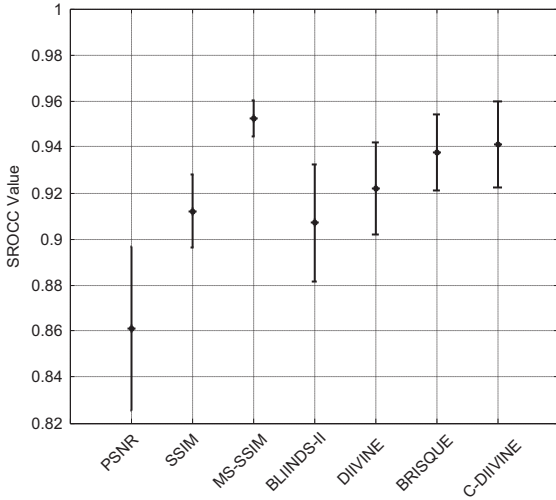|	PSNR	SSIM	MS-SSIM	BLIINDS-II	DIIVINE	BRISQUE	C-DIIVINE-I	C-DIIVINE-II	C-DIIVINE
PSNR	0	-1	-1	-1	-1	-1	-1	-1	-1
SSIM	1	0	-1	1	-1	-1	-1	-1	-1
MS-SSIM	1	1	0	1	1	1	1	1	1
BLIINDS-II	1	-1	-1	0	-1	-1	-1	-1	-1
DIIVINE	1	1	-1	1	0	-1	-1	-1	-1
BRISQUE	1	1	-1	1	1	0	0	1	-1
C-DIIVINE-I	1	1	-1	1	1	0	0	1	-1
C-DIIVINE-II	1	1	-1	1	1	-1	-1	0	-1
C-DIIVINE	1	1	-1	1	1	1	1	1	0

Fig. 15. Mean SROCC and standard error bars for various algorithms across the 1000 train-test trials on the LIVE database.

improvement. In the following sections, we will show more results of these NR IQA algorithms in assessing image quality on other databases. We will also further discuss the roles of the one- and two-stage frameworks in the IQA task.

3.4. Performance on other databases

In order to demonstrate that the performance of C-DIIVINE is not bounded by the database on which it is trained, we evaluated the performance of C-DIIVINE 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 (the distortion types on which C-DIIVINE was trained). In this case, we trained C-DIIVINE on the entire set of distorted images in LIVE, which contains 779 images with five distortion types (note that the fast fading distortion type was trained, but was not tested, because it was not found in CSIQ, TID, and Toyama). Before we evaluate and compare the performance of C-DIIVINE with other NR IQA approaches, we first investigate on the contributions of each individual feature type employed in the proposed algorithm.

3.4.1. Contributions of individual statistical features

To analyze the contributions of each of the three feature types (magnitude distribution parameters, phase distribution parameter, and cross-scale correlations) toward the overall performance, we created six abridged versions of C-DIIVINE in which each version used only one/two of the three feature types. Specifically, the following abridged versions were created:

- Magnitude Distribution Feature only (M);
- Phase Distribution Feature only (P);
- Scale Correlation only (SC);
- Magnitude Distribution Feature+Phase Distribution Feature (M+P);

- Magnitude Distribution Feature+Scale Correlation (M+SC);
- Phase Distribution Feature+Scale Correlation (P+SC).

Each of these versions was trained on the full LIVE database via the same technique used for training C-DIIVINE as described previously in Section 3.3. The testing was performed on the aforementioned subsets of images from CSIQ, TID, and Toyama. The results of this analysis are shown in Table 4 in terms of SROCC and CC. For reference, also shown in Table 4 are the SROCC and CC values of the full C-DIIVINE algorithm.

In general, for the CSIQ database, the best performance is obtained by using all three feature types. For the TID database, the magnitude and phase distribution features seem to be very important, for the overall performance of using both feature types (M+P) yields almost equivalent performance as that use all three of them (M+P+SC). More specifically, the phase-based statistic seems to be more sensitive to JPEG and blur distortions, whereas the magnitude-based statistic is more sensitive to the JPEG2000 and Gaussian white noise distortions. For the Toyama database, the largest contribution also arises from the combination of magnitude and phase.

With regard to the performance when all of the distortion types are evaluated together as a set (the row labeled "All" in Table 4), the best performance is achieved when using all three feature types. Overall, when looking at the entire database, the phase distribution feature seems to contribute the most, which also demonstrates the importance of phase information to image quality. Also, these results confirm that different distortion types manifest by modifying different subband statistics. Some distortions primarily affect the coefficient magnitudes, some primarily affect the coefficient phases and some primarily affect the across-scale correlation. Thus, all three feature types are required to properly estimate quality when all of the distortion types are evaluated together.

3.4.2. Overall performance

The overall testing results are shown in Table 5 in terms of SROCC, CC, RMSE, OR, and OD. Also included are C-DIIVINE-I and C-DIIVINE-II for comparison. Italicized entries denote NR algorithms. The results of the best-performing FR algorithm in each case are bolded, and results of the best-performing NR algorithm are italicized and bolded.

As shown in Table 5, C-DIIVINE performs quite well on the IQA task as compared with other NR methods. It improves upon DIIVINE and is superior to BLIINDS-II. Further, it even challenges BRISQUE, a spatial domain IQA algorithm, and some FR IQA methods. Specifically, in terms of prediction monotonicity (SROCC) and prediction accuracy (CC and RMSE), C-DIIVINE outperforms the other three NR IQA algorithms on all databases. In terms of prediction consistency (OR and OD), C-DIIVINE also performs better than DIIVINE and demonstrates across-the-board competitive predictive performance among the four NR IQA methods.

Table 4

SROCC and CC for abridged versions of C-DIIVINE using only one/two of the three feature types. M=Magnitude Distribution Feature only. SC=Scale Correlation only. P=Phase Distribution Feature only. M+P=Magnitude Distribution Feature+Phase Distribution Feature (no across-scale correlation). M+SC=Magnitude Distribution Feature+Scale Correlation (no phase). P+SC=Phase Distribution Feature+Scale Correlation (no magnitude). For reference, the SROCC and CC values of the full C-DIIVINE algorithm are also included (denoted by M+P+SC).

	M	P	SC	M+P	P+SC	M+SC	M+P+SC
CSIQ							
CC							
JPEG2000	0.902	0.847	0.889	0.891	0.903	0.905	**0.920**
JPEG	0.934	0.893	0.772	0.947	0.926	0.933	**0.961**
BLUR	0.889	0.761	0.680	0.911	0.799	0.918	**0.932**
WN	0.846	0.639	0.723	0.897	0.788	0.880	**0.907**
All	0.900	0.813	0.783	0.914	0.868	0.914	**0.935**
SROCC							
JPEG2000	0.873	0.818	0.859	0.860	0.877	0.882	**0.893**
JPEG	0.894	0.841	0.761	0.899	0.879	0.899	**0.916**
BLUR	0.860	0.782	0.712	0.886	0.823	0.903	**0.907**
WN	0.831	0.627	0.715	0.884	0.780	0.878	**0.897**
All	0.867	0.792	0.769	0.886	0.859	0.897	**0.910**
TID							
CC							
JPEG2000	0.904	0.825	0.874	0.923	0.910	0.910	**0.936**
JPEG	0.913	0.931	0.824	0.914	0.932	0.905	**0.953**
BLUR	0.846	**0.924**	0.865	0.894	0.901	0.894	0.891
WN	0.704	0.621	0.584	**0.827**	0.619	0.813	0.795
All	0.730	0.872	0.834	0.921	0.894	0.901	**0.925**
SROCC							
JPEG2000	0.900	0.825	0.880	0.919	0.913	0.907	**0.937**
JPEG	0.893	0.900	0.823	0.875	0.902	0.883	**0.924**
BLUR	0.849	**0.928**	0.877	0.898	0.914	0.899	0.900
WN	0.717	0.601	0.578	**0.842**	0.633	0.825	0.809
All	0.792	0.879	0.834	0.919	0.891	0.904	**0.921**
Toyama							
CC							
JPEG2000	0.875	0.879	0.751	**0.903**	0.828	0.785	0.870
JPEG	**0.905**	0.900	0.701	0.869	0.841	0.752	0.884
All	0.887	**0.889**	0.679	0.886	0.830	0.814	0.876
SROCC							
JPEG2000	0.880	0.876	0.723	**0.903**	0.797	0.786	0.874
JPEG	**0.887**	**0.887**	0.698	0.869	0.832	0.828	0.882
All	0.883	0.884	0.709	**0.885**	0.810	0.815	0.877

Compared with C-DIIVINE-I and C-DIIVINE-II, we see that the proposed combined frameworks achieves a performance balance across the different databases. It can be clearly observed that the one-stage framework (C-DII-VINE-I) performs quite well when training on LIVE and testing on TID. This might be attributed to the considerable similarity between the two databases. In fact, the cross-validation test results shown in Section 3.3 also demonstrate that the one-stage framework can be a very good choice for decision-making when the training data and testing data share some common properties, considering that the two-stage framework often produces classification errors in its identification stage. However, it is also noteworthy that C-DIIVINE-I performs slightly worse than C-DIIVINE-II when testing on CSIQ and Toyama, which demonstrates that the two-stage framework is perhaps more robust. Therefore, combining the two frameworks is a simple, yet effective technique for achieving a reasonable balance in IQA performance across different databases.

The last rows of the SROCC, CC, and RMSE results in Table 5 show the average SROCC, CC, and RMSE, where the averages are weighted by the number of distorted images tested in each database. On an average, C-DIIVINE and BRISQUE demonstrate the best NR QA performance. Note that OD is dependent on the dynamic range of the database, and therefore cannot be compared across databases, only within.

Fig. 16 shows scatter-plots of logistic-transformed C-DIIVINE quality predictions vs. 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, overall, the proposed C-DIIVINE algorithm can predict quality well across the range of MOS/DMOS.

3.4.3. Performance on individual distortion types

In order to demonstrate that C-DIIVINE can achieve fairly good quality evaluation across different distortion types as long as it has been trained on those distortion types, we also report the performance of C-DIIVINE on subsets of the three previously mentioned databases corresponding to four individual distortion types: JPEG, JPEG2000, Gaussian blurring,

Table 5

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

	MS-SSIM	VIF	MAD	*DIIVINE*	*BLIINDS-II*	*BRISQUE*	*C-DIIVINE-I*	*C-DIIVINE-II*	*C-DIIVINE*
CC									
CSIQ	0.950	0.967	**0.974**	*0.854*	*0.901*	*0.924*	*0.914*	***0.935***	***0.935***
TID	0.912	**0.950**	0.947	*0.877*	*0.864*	*0.907*	***0.931***	*0.907*	*0.925*
Toyama	0.893	0.914	**0.941**	*0.634*	*0.754*	*0.850*	*0.851*	*0.870*	***0.876***
Average	0.929	0.954	**0.960**	*0.830*	*0.867*	*0.907*	*0.910*	*0.916*	***0.923***
SROCC									
CSIQ	0.953	0.959	**0.967**	*0.828*	*0.873*	*0.900*	*0.879*	***0.913***	*0.910*
TID	0.897	**0.940**	0.935	*0.891*	*0.840*	*0.898*	***0.922***	*0.905*	*0.921*
Toyama	0.887	0.909	**0.936**	*0.642*	*0.724*	*0.848*	*0.851*	*0.871*	***0.877***
Average	*0.924*	0.945	**0.952**	*0.822*	*0.840*	*0.892*	*0.889*	*0.904*	***0.909***
RMSE									
CSIQ	0.089	0.072	**0.064**	*0.147*	*0.123*	*0.108*	*0.115*	***0.100***	***0.100***
TID	0.650	**0.495**	0.509	*0.760*	*0.798*	*0.668*	***0.577***	*0.668*	*0.602*
Toyama	0.564	0.507	**0.425**	*0.968*	*0.864*	*0.660*	*0.656*	*0.616*	***0.604***
Average	0.345	0.277	**0.265**	*0.468*	*0.456*	*0.375*	*0.348*	*0.365*	***0.341***
OR									
CSIQ	0.220	0.187	**0.157**	*0.365*	*0.320*	*0.262*	*0.297*	***0.242***	*0.247*
TID	0.734	0.630	**0.622**	*0.706*	*0.766*	*0.693*	***0.641***	*0.672*	*0.651*
Toyama	0.066	0.030	**0.024**	*0.214*	*0.155*	*0.066*	*0.083*	*0.071*	***0.065***
Average	0.369	0.312	**0.293**	*0.457*	*0.444*	*0.377*	*0.380*	*0.360*	***0.355***
OD									
CSIQ	6.473	5.013	**3.243**	*24.561*	*16.318*	*12.697*	*14.498*	*11.810*	***11.600***
TID	120.957	**78.159**	79.194	*137.472*	*162.692*	*121.259*	***96.667***	*113.943*	*99.832*
Toyama	3.172	**1.397**	1.673	*16.517*	*12.685*	***4.158***	*5.786*	*5.380*	*5.418*

and additive Gaussian white noise. The same logistic transform was used and the performance for each distortion type was evaluated based on extracting the corresponding transformed scores previously computed when all four distortion types were considered. Table 6 shows the results in terms of SROCC, CC, and RMSE values. Also included are results of three other NR and FR IQA algorithms for comparison. From the table, we see that C-DIIVINE provides better predictions on most distortion types in comparison to BLIINDS-II, DIIVINE, and BRISQUE. When looking at the performance of C-DIIVINE-I and C-DIIVINE-II, we conclude again that the proposed combined framework actually takes advantages of both individual frameworks, and achieves a performance balance across different databases on most distortion types.

3.4.4. Performance on reference images

We also evaluated the performance of C-DIIVINE on reference images. For this evaluation, we tested C-DIIVINE on undistorted versions of natural images contained in the three aforementioned databases (note that only 24 undistorted natural images in the TID database was tested). Since there are no valid human opinion scores for reference images in the CSIQ and TID databases, and the true reference image MOS values contained in the Toyama database are too close to indicate any useful linear/non-linear relationships, we employed here a different approach for evaluation. Specifically, we first predicted the original quality score for each reference image using our algorithm, and then applied the same logistic transform (the sigmoid function parameters were estimated by using the distorted images in the same database) defined in Eq. (16) to bring the predicted values on the same scales as the true MOS/DMOS values for the distorted images. Finally, for each database, we compared

these linearized estimated DMOS values to the corresponding scatter plots in Fig. 16 to see if they are all located in a region representing the highest quality. The results are shown in Fig. 17.

By referring to both figures (Figs. 16 and 17), we see that C-DIIVINE works well on most reference images. Specifically, 28 out of 30 reference images in the CSIQ database are given quality scores below 0.2, which indicates that only two of them were predicted a bit far away from their true DMOS values. Note in Fig. 16 that the reference images in the CSIQ database should have DMOS values close to zero. For the TID database, qualities of all 24 reference images seem to be well predicted. It also seems that only one reference image in the Toyama database was improperly estimated. Note again in Fig. 16 that the undistorted natural images in the TID and Toyama databases should have MOS values around/above 5.5 and 4.5, respectively. All three results demonstrate that the proposed C-DIIVINE algorithm performs well on most of these reference images.

3.5. Computational complexity

Having evaluated the performance of C-DIIVINE and other NR algorithms on various databases, we now analyze their computational complexity. Although C-DIIVINE extends its predecessor DIIVINE to the complex domain, it is still computationally efficient. To evaluate the computational complexity of each of the three feature types, we measured the relative percentage of time required to compute the quality of a 512×512 image. We also compared the overall computational complexity of C-DIIVINE with other NR IQA algorithms (BLIINDS-II, DIIVINE, and BRISQUE) on an image of resolution 512×640. Note that all of these tests were

Table 6
SROCC, CC and RMSE of C-DIIVINE 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 bolded, and results of the best-performing NR algorithm are italicized and bolded.

|	MS-SSIM	VIF	MAD	*DIIVINE*	*BLIINDS-II*	*BRISQUE*	*C-DIIVINE-I*	*C-DIIVINE-II*	*C-DIIVINE*
CC									
CSIQ									
JPEG2000	0.977	0.978	**0.983**	*0.893*	*0.912*	*0.896*	*0.913*	*0.916*	***0.920***
JPEG	0.981	**0.988**	0.983	*0.697*	*0.912*	*0.946*	*0.942*	***0.966***	*0.961*
BLUR	0.959	0.974	**0.976**	*0.898*	*0.897*	*0.928*	*0.912*	*0.928*	***0.932***
WN	0.947	**0.961**	0.956	*0.787*	*0.897*	***0.938***	*0.809*	*0.921*	*0.907*
TID									
JPEG2000	0.975	0.971	**0.982**	*0.896*	*0.919*	*0.906*	***0.947***	*0.902*	*0.936*
JPEG	0.966	**0.973**	0.961	*0.899*	*0.889*	*0.950*	***0.956***	*0.943*	*0.953*
BLUR	**0.951**	0.942	0.801	*0.844*	*0.825*	*0.873*	***0.902***	*0.872*	*0.891*
WN	0.810	**0.907**	0.819	***0.828***	*0.714*	*0.810*	*0.791*	*0.779*	*0.795*
Toyama									
JPEG2000	0.949	**0.962**	0.961	*0.603*	*0.686*	*0.869*	*0.851*	***0.871***	*0.870*
JPEG	0.787	0.900	**0.919**	*0.709*	*0.826*	*0.865*	*0.855*	*0.873*	***0.884***
SROCC									
CSIQ									
JPEG2000	0.969	0.967	**0.975**	*0.830*	*0.884*	*0.867*	*0.880*	*0.892*	***0.893***
JPEG	0.962	**0.970**	0.962	*0.704*	*0.881*	*0.909*	*0.890*	***0.921***	*0.916*
BLUR	0.972	**0.975**	0.968	*0.871*	*0.870*	*0.903*	*0.881*	*0.911*	*0.907*
WN	0.947	**0.957**	0.954	*0.797*	*0.886*	***0.925***	*0.799*	*0.911*	*0.897*
TID									
JPEG2000	0.973	0.970	**0.975**	*0.907*	*0.911*	*0.904*	***0.947***	*0.906*	*0.937*
JPEG	**0.940**	0.931	0.925	*0.871*	*0.838*	*0.911*	*0.923*	*0.913*	***0.924***
BLUR	**0.963**	0.958	0.847	*0.859*	*0.826*	*0.874*	***0.908***	*0.887*	*0.900*
WN	0.818	**0.913**	0.833	***0.834***	*0.715*	*0.823*	*0.803*	*0.788*	*0.809*
Toyama									
JPEG2000	0.945	**0.956**	0.955	*0.612*	*0.627*	*0.867*	*0.852*	***0.875***	*0.874*
JPEG	0.835	0.907	**0.917**	*0.702*	*0.820*	*0.857*	*0.855*	*0.865*	***0.882***
RMSE									
CSIQ									
JPEG2000	0.067	0.066	**0.058**	*0.130*	*0.142*	*0.140*	*0.129*	*0.127*	***0.124***
JPEG	0.059	**0.047**	0.057	*0.125*	*0.219*	*0.099*	*0.102*	***0.079***	*0.084*
BLUR	0.081	0.065	**0.062**	*0.127*	*0.126*	*0.107*	*0.117*	*0.107*	***0.104***
WN	0.054	**0.047**	0.050	*0.074*	*0.104*	***0.058***	*0.099*	*0.065*	*0.071*
TID									
JPEG2000	0.431	0.459	**0.366**	*1.178*	*0.756*	*0.810*	***0.617***	*0.827*	*0.675*
JPEG	0.435	**0.388**	0.468	*0.741*	*0.774*	*0.528*	***0.497***	*0.561*	*0.513*
BLUR	**0.359**	0.390	0.694	*0.621*	*0.655*	*0.566*	***0.501***	*0.568*	*0.527*
WN	0.359	**0.258**	0.351	***0.343***	*0.429*	*0.359*	*0.375*	*0.383*	*0.371*
Toyama									
JPEG2000	0.398	**0.344**	0.348	*1.007*	*0.983*	*0.625*	*0.662*	***0.620***	*0.622*
JPEG	1.237	0.540	**0.488**	*0.872*	*0.709*	*0.622*	*0.641*	*0.604*	***0.579***
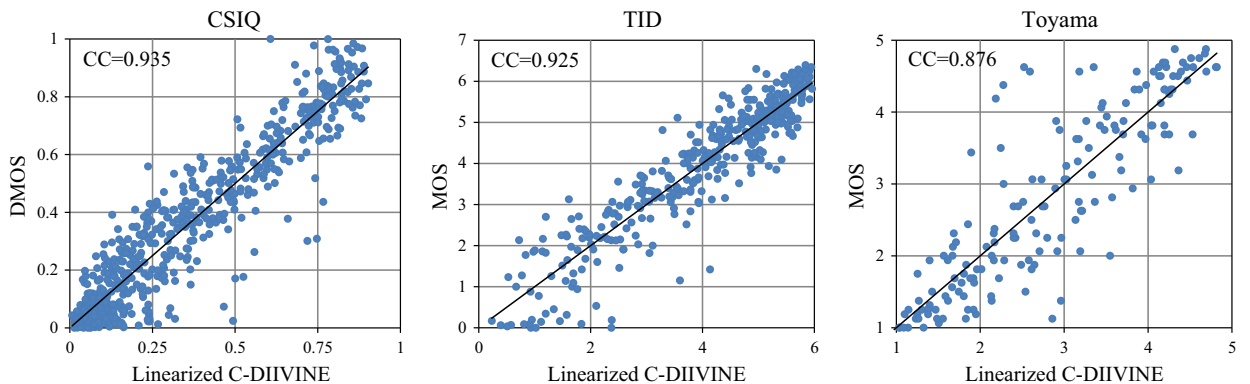
Fig. 16. Scatter plots of objective scores predicted by the C-DIIVINE algorithm after logistic transform versus subjective scores on different image databases. Note that the *x*-axis across all three figures represents the predicted value transformed via Eq. (16); the *y*-axis represents the true DMOS value for the CSIQ database, true MOS value for the TID and Toyama databases.
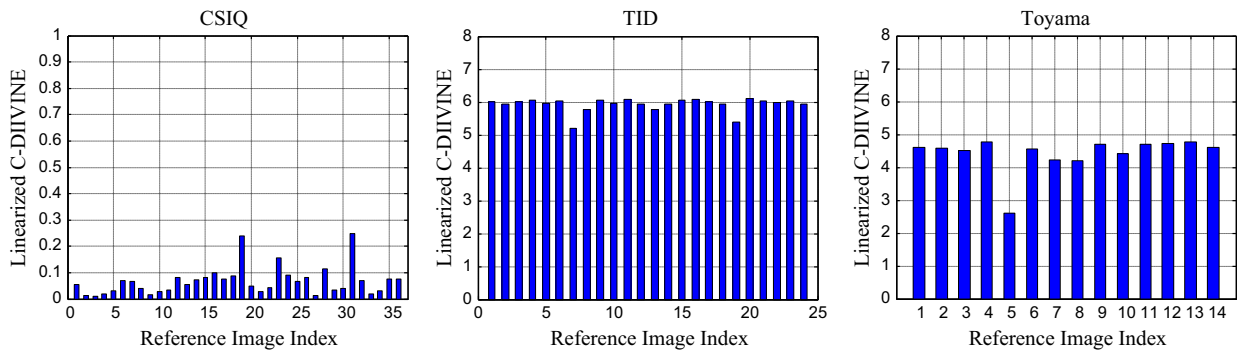
Fig. 17. Objective quality scores predicted by the C-DIIVINE algorithm after logistic transform on reference images. Note that the *x*-axis across all three figures represents the index of reference images in each database; the *y*-axis represents the predicted value transformed via Eq. (16).

Table 7
Informal complexity analysis of C-DIIVINE. Tabulated values reflect the percentage of the time devoted to each of the three feature types in C-DIIVINE.

Feature type	Percentage of time
Magnitude-based features	51.05
Phase-based features	38.03
Across-scale correlation features	10.92

Table 8
A comparison of the runtime requirements (seconds) for four NR IQA algorithms on a 512×640 image.

Algorithm	Time (s)
BLIINDS-II	50.67
DIIVINE	28.77
BRISQUE	0.13
C-DIIVINE	12.80

performed by running unoptimized MATLAB code on a modern desktop computer (AMD Phenom II X4 965 Processor at 3.39 GHz, 4.00 GB RAM, Windows 7 Pro 64-bit). The results are shown in Tables 7 and 8, respectively.

As shown in Table 8, C-DIIVINE runs faster than DIIVINE and BLIINDS-II. This is due to the fact that DIIVINE computes the spatial correlation statistics of an image, and BLIINDS-II extracts the DCT-based features from non-overlapping image blocks, both of which occupy a considerable chunk of the processing time. However, C-DIIVINE runs slower than BRISQUE. This is mainly because of the divisive normalization, a relatively time-consuming process required for computing the magnitude-based features.

4. Conclusion

We presented an extension of DIIVINE to the complex domain that, for the first time, incorporates the CGSM model to assess image quality without need of a reference across a variety of distortion categories. The proposed C-DIIVINE algorithm employs combined frameworks and complex statistical features to assess image quality by utilizing both magnitude and phase information of the wavelet coefficients. Specifically, we use the complex generalized Gaussian distribution to model the distribution of wavelet coefficient magnitude, the generalized Gaussian distribution to model the distribution of relative magnitude, and the wrapped Cauchy distribution to model the distribution of relative phase. We also use the CW-SSIM index to capture the correlation between different image scales. Results show that C-DIIVINE performs well on most distorted images from the LIVE, CSIQ, TID and Toyama databases, predicting image qualities in a meaningful agreement with the subjective human being assessment. It achieves higher CC and SROCC compared with DIIVINE and BLIINDS-II on most distortion types from the testing databases, and even challenges BRISQUE and some of the well-known FR IQA algorithms.

Acknowledgments

References

[1] W. Zeng, S. Daly, S. Lei, Point-wise extended visual masking for JPEG2000 image compression, in: Proceedings on the International Conference on Image Processing, vol. 1, 2000, pp. 657–660.

[2] I. Hontsch, L. Karam, Locally adaptive perceptual image coding, IEEE Trans. Image Process. 9 (2000) 1472–1483.

[3] C.J. van den Branden Lambrecht, A working spatio-temporal model of the human visual system for image representation and quality assessment applications, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996, pp. 2291–2294.

[4] S. Winkler, Visual quality assessment using a contrast gain control model, in: IEEE Signal Processing Society Workshop on Multimedia Signal Processing, 1999, pp. 527–532.

[5] T. Lam, L.J. Karam, G.P. Abousleman, Robust image coding using perceptually-tuned channel- optimized trellis-coded quantization, 42nd Midwest Symposium on Circuits and Systems 2 (1999) 1131–1134.

[6] J.A. Ferwerda, Fundamentals of spatial vision, Appl. Vis. Percept. Comput. Graphics (1998) 1–27.

[7] B. Walter, S.N. Pattanaik, D.P. Greenberg, Using perceptual texture masking for efficient image synthesis, Comput. Graph. Forum 21 (3) (2002) 393–399.

[8] D.I. Neuhoff, T.N. Pappas, Perceptual coding of images for halftone display, IEEE Trans. Image Process. 3 (4) (1994) 341–354.

[9] Q. Yu, K.J. Parker, Quality issues in blue noise halftoning, in: Proceeding of SPIE Color Imaging: Device Independent Color, Color Hardcopy, and Graphic Arts III: Device, vol. 3300, 1998, pp. 376–385.

[10] S. Lu, H. Pu, C. Lin, A HVS-directed neural-network-based approach for impulse-noise removal from highly corrupted images, in: Proceedings of the IEEE International Conference on Man and Cybernetic Systems, vol. 1, 2003, pp. 72–77.

[11] Z. Wang, A.C. Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (1) (2009) 98–117.

[12] B. Moulden, F.A.A. Kingdom, L.F. Gatley, The standard deviation of luminance as a metric for contrast in random-dot images, Perception 19 (1990) 79–101.

[13] F. Lukas, Z. Budrikis, Picture quality prediction based on a visual model, IEEE Trans. Commun. 30 (7) (1982) 1679–1692.

[14] J. Lubin, A visual discrimination model for imaging system design and evaluation, Vis. Models Target Detect. Recognit. (1995) 245–283.

[15] Y. Lai, C.J. Kuo, Image quality measurement using the haar wavelet, Wavelet Appl. Signal Image Process. V Proc. SPIE (1997) 127–138.

[16] S. Daly, The visible differences predictor: an algorithm for the assessment of image fidelity, Digit. Images Hum. Vis. 20 (1993) 179–206.

[17] N. Nill, A visual model weighted cosine transform for image compression and quality assessment, IEEE Trans. Commun. 33 (6) (1985) 551–557.

[18] P.C. Teo, D.J. Heeger, Perceptual image distortion, Proc. IEEE Conf. Image Process. 2 (1994) 982–986.

[19] S.J.P. Westen, R.L. Lagendijk, J. Biemond, Perceptual image quality based on a multiple channel HVS model, in: International Conference on Acoustics, Speech, and Signal Processing, vol. 4, 1995, pp. 2351–2354.

[20] A.B. Watson, M. Taylor, R. Borthwick, Image quality and entropy masking, Hum. Vis. Electron. Imaging II Proc. SPIE 3016 (1997) 2–12.

[21] S. Winkler, A perceptual distortion metric for digital color images, in: Proceedings of IEEE International Conference on Image Processing, vol. 3, 1998, pp. 399–403.

[22] A. Bradley, A wavelet visible difference predictor, IEEE Trans. Image Process. 8 (1999) 717–730.

[23] P. LeCallet, A. Saadane, D. Barba, Frequency and spatial pooling of visual differences for still image quality assessment, Proc. SPIE Hum. Vis. Electr. Imaging V 3959 (2000) 595–603.

[24] N. Damera-Venkata, T.D. Kite, W. Geisler, B.L. Evans, A.C. Bovik, Image quality assessment based on a degradation model, IEEE Trans. Image Process. 9 (2000) 636–650.

[25] M. Carnec, P.L. Callet, D. Barba, An image quality assessment method based on perception of structural information, in: Proceedings of International Conference on Image Processing, vol. 3, 2003, pp. 185–188.

[26] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.

[27] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[28] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, vol. 2, 2003, pp. 1398–1402.

[29] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.

[30] G. Zhai, W. Zhang, X. Yang, Image quality assessment metrics based on multi-scale edge presentation, in: IEEE Workshop Signal Processing Systems Design and Implementation, 2005, pp. 331–336.

[31] C.-L. Yang, W.-R. Gao, L.-M. Po, Discrete wavelet transform-based structural similarity for image quality assessment, in: 15th IEEE International Conference on Image Processing, 2008, pp. 377–380.

[32] M. Zhang, X. Mou, A psychovisual image quality metric based on multi-scale structure similarity, in: IEEE International Conference on Image Processing, 2008, pp. 381–384.

[33] A. Shnayderman, A. Gusev, A.M. Eskicioglu, An SVD-based grayscale image quality measure for local and global assessment, IEEE Trans. Image Process. 15 (2) (2006) 422–429.

[34] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, Proc. SPIE Hum. Vis. Electron. Imaging X 5666 (2005).

[35] K. Chono, Y.-C. Lin, D.P. Varodayan, Y. Miyamoto, B. Girod, Reduced-reference image quality assessment using distributed source coding,

[36] W. Xue, X. Mou, Reduced-reference image quality assessment based on weibull statistics, in: International Workshop on Quality of Multimedia Experience, 2010, pp. 1–6.

[37] Z. Wang, H.R. Sheikh, A.C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proceedings of IEEE International Conference on Image Processing, vol. 1, 2002, pp. 477–480.

[38] L. Meesters, J.B. Martens, A single-ended blockiness measure for JPEG-coded images, Signal Process. 82 (3) (2002) 369–387.

[39] Y. Horita, S. Arata, T. Murai, No-reference image quality assessment for JPEG/JPEG2000 coding, in: European Signal Processing Conference, 2004, pp. I301–I304.

[40] H. Tong, M. Li, H. Zhang, C. Zhang, No-reference quality assessment for JPEG2000 compressed images, in: International Conference on Image Processing, vol. 5, 2004, pp. 3539–3542.

[41] M.G. Choi, J.H. Jung, J.W. Jeon, No-reference image quality assessment using blur and noise, World Acad. Sci. Eng. Technol. (2009) 163–167.

[42] C. Mingjun, A.C. Bovik, No-reference image blur assessment using multiscale gradient, in: International Workshop on Quality of Multimedia Experience, 2009, pp. 70–74.

[43] Y.-C. Chung, J.-M. Wang, R.R. Bailey, S.-W. Chen, S.-L. Chang, A non-parametric blur measure based on edge analysis for image processing applications, in: Proceedings of IEEE Conference on Cybernetics and Intelligent Systems, vol. 1, 2004, pp. 356–360.

[44] Y.-C. Chung, S.-L. Chang, J.-M. Wang, S.-W. Chen, An edge analysis based blur measure for image processing applications, J. Taiwan Normal Univ. Math. Sci. Technol. 51 (1) (2006) 21–31.

[45] L. Liang, J. Chen, S. Ma, D. Zhao, W. Gao, A no-reference perceptual blur metric using histogram of gradient profile sharpness, in: 16th IEEE Conference on Image Processing, 2009, pp. 4369–4372.

[46] C.T. Vu, D.M. Chandler, S3: A spectral and spatial sharpness measure of local perceived sharpness in natural images, IEEE Trans. Image Process. 21 (3) (2012) 934–945.

[47] D. Shaked, I. Tastl, Sharpness measure: towards automatic image enhancement, in: IEEE Conference on Image Processing, vol. 1, 2005, pp. 937–940.

[48] A. Ciancio, A.L.N.T. da Costa, E.A.B. da Silva, A. Said, R. Samadani, P. Obrador, No-reference blur assessment of digital pictures based on multifeature classifiers, IEEE Trans. Image Process. 20 (1) (2011) 64–75.

[49] P. Ye, D. Doermann, No-reference image quality assessment based on visual codebook, in: 18th IEEE International Conference on Image Processing, 2011, pp. 3089–3092.

[50] P. Ye, D. Doermann, No-reference image quality assessment using visual codebook, IEEE Trans. Image Process. 21 (7) (2012) 3129–3138.

[51] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. A 4 (1987) 2379–2394.

[52] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? Vis. Res. 37 (1996) 3311–3325.

[53] E.P. Simoncelli, B.A. Olshausen, Natural image statistics and neural representation, Annu. Rev. Neurosci. 24 (2001) 1193–1216.

[54] S. Gabarda, G. Cristobal, Blind image quality assessment through anisotropy, J. Opt. Soc. Am. A 24 (2007) B42–B51.

[55] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, IEEE Signal Process. Lett. 17 (2010) 583–586.

[56] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.

[57] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.

[58] E.P. Simoncelli, W.T. Freeman, The steerable pyramid: a flexible architecture for multi-scale derivative computation, in: Proceedings of International Conference on Image Processing, vol. 3, 1995, pp. 444–447.

[59] D.J. Field, Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems, in: Wavelets, Fractals and Fourier Transforms: New Development and New Applications, 1993, pp. 151–193.

[60] D.J. Field, What is the goal of sensory coding? Neural Comput. 6 (1994) 559–601.

[61] W.E. Vinje, J.L. Gallant, Sparse coding and decorrelation in primary visual cortex during natural vision, Science 287 (5456) (2000) 1273–1276.

[62] D.J. Tolhurst, Y. Tadmor, T. Chao, Amplitude spectra of natural images, Ophthal. Phys. Opt. 12 (2) (1992) 229–232.

[63] A.V. Oppenheim, J.S. Lim, The importance of phase in signals, Proc. IEEE 69 (1981) 529–541.

[64] M.G.A. Thomson, D.H. Foster, R.J. Summers, Human sensitivity to phase perturbations in natural images: a statistical framework, Perception 29 (2000) 1057–1069.

[65] P.J. Bex, W. Makous, Spatial frequency, phase, and the contrast of natural images, J. Opt. Soc. Am. A 19 (2002) 1096–1106.

[66] W.S. Geisler, J.S. Perry, B.J. Super, D.P. Gallogly, Edge co-occurrence in natural images predicts contour grouping performance, Vis. Res. 41 (2001) 711–724.

[67] M.J. Morgan, J. Ross, A. Hayes, The relative importance of local phase and local amplitude in patchwise image reconstruction, Biol. Cybern. 65 (2) (1991) 113–119.

[68] D.J. Field, D.M. Chandler, Method for estimating the relative contribution of phase and power spectra to the total information in natural-scene patches, J. Opt. Soc. Am. A 29 (1) (2012) 55–67.

[69] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008 – a database for evaluation of full-reference visual quality assessment metrics, Adv. Mod. Radioelectron. 10 (2009) 30–45.

[70] J.G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles, Vis. Res. 20 (10) (1980) 847–856.

[71] J.G. Daugman, Uncertainty relation for resolution in space, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Am. 2 (27) (1985) 1160–1169.

[72] S. Marcelja, Mathematical description of the response of simple cortical cells, J. Opt. Soc. Am. 70 (11) (1980) 1297–1300.

[73] J.P. Jones, L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, J. Neurophysiol. 58 (6) (1987) 1233–1258.

[74] G. Wallis, Linear models of simple cells: correspondence to real cell responses and space spanning properties, Spat. Vis. 14 (3–4) (2001) 237–260.

[75] Y. Rakvongthai, A.P.N. Vo, S. Orantara, Complex Gaussian scale mixtures of complex wavelet coefficients, IEEE Trans. Signal Process. 58 (7) (2010) 3545–3556.

[76] Z. Wang, E.P. Simoncelli, Translation insensitive image similarity in the complex wavelet domain, in: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing, 2005, pp. 573–576.

[77] A.B. Watson, J.A. Solomon, A model of visual contrast gain control and pattern masking, J. Opt. Soc. Am. A 14 (1997) 2378–2390.

[78] D.J. Heeger, Normalization of cell responses in cat striate cortex, Vis. Neurosci. 9 (1992) 181–197.

[79] M.J. Wainwright, O. Schwartz, E.P. Simoncelli, Natural image statistics and divisive normalization: modeling nonlinearity and adaptation in cortical neurons, In: Probabilistic Models of the Brain: Perception and Neural Function, 2002, pp. 203–222.

[80] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Topics Signal Process. 3 (2) (2009) 202–211.

[81] H.R. Sheikh, Z. Wang, A.C. Bovik, L. K. Cormack, Image and video quality assessment research at LIVE, ⟨http://live.ece.utexas.edu/research/quality/⟩.

[82] M.J. Wainwright, E.P. Simoncelli, Scale mixture of Gaussians and the statistics of natural images, Adv. Neural Inf. Process. Syst. 12 (2000) 855–861.

[83] K. Sharifi, A. Leon-Garcia, Estimation of shape parameter for generalized Gaussian distribution in subband decompositions of video, IEEE Trans. Circuits Syst. Video Technol. 5 (1) (1995) 52–56.

[84] S.O.A. Vo, T. Nguyen, Using phase and magnitude information of the complex directional filter bank for texture image retrieval, in: Proceedings of International Conference on Image Processing, vol. 4, 2007, pp. IV–61–IV–64.

[85] A. Vo, S. Oriaintara, N. Nguyen, Vonn distribution of relative phase for statistical image modeling in complex wavelet domain, Signal Process. 91 (1) (2011) 114–125.

[86] A. Vo, S. Oraintara, A study of relative phase in complex wavelet domain: property, statistics and applications in texture image retrieval and segmentation, Signal Process. Image Commun. 25 (1) (2010) 28–46.

[87] Y. Rakvongthai, S. Oriaintara, Statistical image modeling with the magnitude probability density function of complex wavelet coefficients, in: IEEE International Symposium on Circuits and Systems, 2009, pp. 1879–1882.

[88] Y. Horita, K. Shibata, Y. Kawayoke, Z.M.P. Sazzad, Subjective quality assessment Toyama database, ⟨http://mict.eng.u-toyama.ac.jp/mictdb.html⟩.

[89] See ⟨http://vision.okstate.edu/csiq/⟩.

[90] M.A. Saad, A.C. Bovik, Blind image quality assessment: a natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (2012) 3339–3352.

[91] A. T1.TR.74-2001, Objective Video Quality Measurement Using a Peak-Signal-to-Noise-Ratio (PSNR) Full Reference Technique (2001).

[92] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 011006.

[93] VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II, ⟨http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx⟩(2003).

[94] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.

[95] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27:1–27:27.