Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation

Qiang Li, *Student Member, IEEE*, and Zhou Wang, *Member, IEEE*

***Abstract*—Reduced-reference image quality assessment (RRIQA) methods estimate image quality degradations with partial information about the "perfect-quality" reference image. In this paper, we propose an RRIQA algorithm based on a divisive normalization image representation. Divisive normalization has been recognized as a successful approach to model the perceptual sensitivity of biological vision. It also provides a useful image representation that significantly improves statistical independence for natural images. By using a Gaussian scale mixture statistical model of image wavelet coefficients, we compute a divisive normalization transformation (DNT) for images and evaluate the quality of a distorted image by comparing a set of reduced-reference statistical features extracted from DNT-domain representations of the reference and distorted images, respectively. This leads to a generic or general-purpose RRIQA method, in which no assumption is made about the types of distortions occurring in the image being evaluated. The proposed algorithm is cross-validated using two publicly-accessible subject-rated image databases (the UT-Austin LIVE database and the Cornell-VCL A57 database) and demonstrates good performance across a wide range of image distortions.**

***Index Terms*—Divisive normalization, image quality assessment, reduced-reference image quality assessment (RRIQA), perceptual image representation, statistical image modeling.**

I. Introduction

IN RECENT years, there has been an increasing need of accurate and easy-to-use image quality assessment (IQA) algorithms in a variety of real world applications, including image compression, communication, printing, display, restoration, segmentation, and fusion [1]. Most existing IQA methods require full access to an original reference image that is assumed to have perfect quality. Without the reference image, the IQA task becomes very difficult, and almost all existing no-reference IQA metrics were designed for one or a set of predefined specific distortion types (such as blocking [2]–[5] and blurring [5] in JPEG; and ringing [6], blurring [6] and wavelet quantization effect [7], [8] in JPEG2000). They are unlikely to generalize for evaluating images degraded with other types of distortions. In practice, these no-reference methods are useful only when the types of distortions between the reference and distorted images are fixed and known.

Reduced-reference IQA (RRIQA) methods provide an interesting tradeoff. They predict the quality degradation of an image with only partial information about the reference image, in the form of a set of RR features [1]. RRIQA measures supply a practically useful and convenient tool in applications such as real-time visual information communications over wired or wireless networks, where they can be employed to monitor image quality degradations or control the network streaming resources on the fly. Fig. 1 illustrates how an RRIQA system may be deployed. The system includes a feature extraction process at the sender side and a feature extraction/quality analysis process at the receiver side. The extracted RR features, or the side information, usually have a much lower data rate than the image data and are typically transmitted to the receiver through an ancillary channel [1]. It is often assumed that the ancillary channel is error-free. However, this is not an absolutely necessary requirement since even partly decoded RR features may still be helpful in evaluating the quality of the distorted image, though the accuracy may be affected. The ancillary channel may also be merged with the distortion channel, in which the RR features would need to receive stronger protection (e.g., by error control coding) than the image data during the transmission. Such examples include the "quality-aware image" system proposed in [9]. At the receiver side, the difference between the features extracted from the reference and distorted images is used to evaluate image quality degradation. The feature extraction process at the receiver side may also be adapted according to the information obtained from the RR features received from the ancillary channel.

The general RRIQA framework described in Fig. 1 leaves flexibilities on the selection of RR features. This is indeed the major challenge in the design of RRIQA algorithms, where the appropriate RR features are desirable to:

1) provide an efficient summary of the reference image;
2) be sensitive to a variety of image distortions;
3) be relevant to the visual perception of image quality.

Another important aspect that has to be kept in mind in the selection of RR features is to maintain a good balance between the data rate of RR features and the accuracy of image quality prediction. With a high data rate, one can include a large amount of information about the reference image, leading to more accurate estimation of image quality degradations, but it also becomes a heavy burden to transmit the RR features to the receiver. On the other hand, a lower data rate makes it easier to transmit the RR information, but more difficult for accurate quality estimation. In practical implementation and deployment, the maximal allowed RR data rate is often given and must be observed. Overall,

Manuscript received May 15, 2008; revised December 09, 2008. Current version published March 11, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lina Karam.

Q. Li is with the Department of Electrical Engineering, The University of Texas at Arlington, Arlington, TX 76019 USA.

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: zhouwang@ieee.org).

Digital Object Identifier 10.1109/JSTSP.2009.2014497

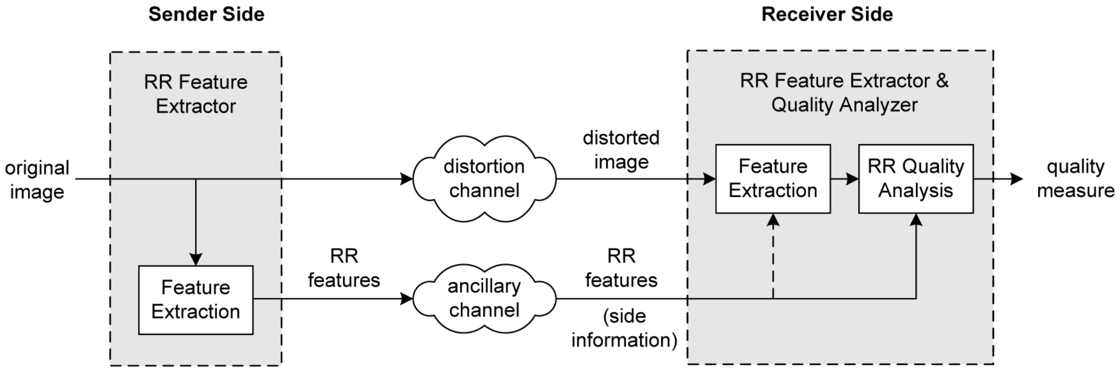


Fig. 1. General framework for the deployment of RRIQA systems.

the merits of an RRIQA system should not be gauged only by the quality prediction accuracy, but by a tradeoff between the accuracy and the RR data rate.

Three different but related types of approaches have been employed in existing RRIQA methods [9]–[16]. The first type of approaches are based on *modeling image distortions* and are mostly developed for specific application environments [10]–[14]. For example, when the distortion type is known to be standard image/video compression, a set of typical distortion artifacts such as blurring, blocking and ringing may be identified, and image features may be defined that are particularly useful to quantify these artifacts [11], [12]. For another example, in [10], [13], a set of spatial and temporal features have been found to be effective in measuring the distortions occurring in standard video compression and communication environment. The second type of approaches are based on *modeling the human visual system* [15], [16], where perceptual features motivated from computational models of low level vision were extracted to provide a reduced description of the image. One advantage of these approaches is that the perceptual features being employed are not directly related to any specific distortion system. As a result, RRIQA methods built upon them could potentially be extended for general purpose. They may also be trained on different types of distortions and produce a variety of distortion-specific RRIQA algorithms under the same general framework. However, no study has been reported so far that applies these methods to the images with generic distortions except for JPEG and JPEG2000 compression [15], [16]. The third type of approaches are based on *modeling natural image statistics* [9]. The basic assumption behind these approaches is that most real-world image distortions disturb image statistics and make the distorted image "unnatural." The unnaturalness measured based on models of natural image statistics can then be used to quantify image quality degradation. In [9], a generalized Gaussian density function is used to model the marginal statistics of the linear coefficients in wavelet subbands, and the parameters of the fitting model are employed as RR features. This general-purpose approach has achieved somewhat surprising success, as it does not require any training, and has a rather low RR data rate, but still supplies reasonable performance when tested with a wide range of image distortion types [9].

Although the method introduced in [9] achieved notable success, our further investigation has revealed some important limitations. First, although the method performed quite well when tested with individual distortion types (e.g., JPEG or JPEG 2000 compression, blurring, or noise contamination), its performance degrades significantly when images with different types of distortions are tested together, as will be shown later in this paper. Second, it uses a rather weak model of natural image statistics, as only marginal distributions of wavelet coefficients are considered. It has been widely noticed that there exist strong dependencies between neighboring wavelet coefficients, which has been completely ignored by this method. Third, it also uses a rather weak model for perceptual image representation, as wavelet decomposition is linear and cannot reflect the nonlinear mechanisms used by the biological visual systems.

In this paper, we propose a new RRIQA method that is inspired by the recent success of the divisive normalization transform (DNT) as a perceptually and statistically motivated image representation [17], [18]. In computational vision science, it has long been hypothesized that the purpose of early visual sensory processing is to increase the statistical independence between neuronal responses [19], [20]. However, linear decompositions, such as Fourier- and wavelet-types of transformations, only reduces the first-order correlation, but cannot reduce the higher order statistical dependencies [21]. In the literature of neural physiology, it has been shown that a local gain-control divisive normalization model is powerful in accounting for the neuronal responses in biological visual systems [22], [23]. This nonlinear gain-control mechanism is built upon linear transform models, where each neuronal response (or linear transform coefficient) is normalized (divided) by the energy of a cluster of neighboring neuronal responses (neighboring coefficients). This process has been shown to significantly reduce the statistical dependencies of the original linear representation [21] and produce approximately Gaussian marginal distributions [24]. Similar models has also been employed in real world image processing applications, including image compression [25] and image enhancement [18]. The strong perceptual and statistical relevance of divisive normalization representation (as compared to linear decompositions) motivated us to switch from the linear wavelet transform domain (as in [9]) to DNT domain in the design of our RRIQA method.

II. Divisive Normalization-Based Image Representation

A. Computation of Divisive Normalization Transformation

A divisive normalization transform (DNT) is built upon a linear image decomposition, followed by a divisive normalization stage. The linear transformations may be discrete cosine transform (DCT) (as in [25]) or wavelet-type of transforms (as in [17], [18], [21]). Here, we assume a wavelet image decomposition, which provides a convenient framework for localized representation of images simultaneously in space, frequency (scale) and orientation. Let y represent a wavelet coefficient, then a normalized coefficient is computed as $\tilde{y}=y/p$, where p is a positive divisive normalization factor that is calculated as the energy of a cluster of coefficients that are close to the coefficient y in space, scale, and orientation.

Several different approaches have been used to compute the normalization factor p [17], [18], [21], [25]. Most of them use a weighted sum of the squared neighboring coefficients plus a positive constant [18], [21], [25]. This involves several parameters (the weights and the constant) that are sometimes difficult to determine. They may be hand-picked (as in [25]) or chosen to maximize the independence of the normalized response to an ensemble of natural images [21]. In [18], a global model of Markov random field over the wavelet coefficients is assumed and the parameters were derived by learning the model parameters using natural images. A more convenient approach is to derive the factor p through a local statistical image model. In particular, the Gaussian scale mixtures (GSM) model has found to be very useful in this context [17]. A length-N random vector Y is a GSM if it can be expressed as the product of two independent components: $Y \doteq zU$, where $\doteq$ denotes equality in probability distribution, U is a zero-mean Gaussian random vector with covariance C_U, and z is a scalar random variable called a mixing multiplier. In other words, the GSM model expresses the density of a random vector as a mixture of Gaussians with the same covariance structure (C_U) but scaled differently (by z). Suppose that the mixing density is $p_z(z)$, then the density of Y can be written as

$$p_Y(Y)=\int \frac{1}{[2\pi]^{\frac{N}{2}}|z^2C_U|^{1/2}}\exp\left(-\frac{Y^TC_U^{-1}Y}{2z^2}\right)p_z(z)dz. \tag{1}$$

This GSM model has shown to be very useful to account for both the marginal and joint statistics of the wavelet coefficients of natural images [17], where the vector Y is formed by clustering a set of neighboring wavelet coefficients within a subband, or across neighboring subbands in scale and orientation. The GSM model has also found successful applications such as image desnoing [26], image restoration [27], and image quality assessment [28].

The general form of the GSM model allows for the mixing multiplier z to be a continuous random variable at each location of the wavelet subbands. To simplify the model, we assume that z only takes a fixed value at each location (but varies over space and subbands). The benefit of this simplification is that when z is fixed, Y is simply a zero-mean Gaussian vector with covariance z^2C_U. As a result, it becomes natural to define the normalization factor p in the DNT representation as an estimate of the multiplier z from the neighboring coefficient vector Y. The coefficient cluster Y moves step by step as a sliding window across a wavelet subband, resulting in a spatially varying normalization factor p. In our implementation, the normalization factor computed at each step is only applied to the center coefficient y_c of the vector Y, and the normalized new coefficient becomes $\tilde{y}_c=y_c/\hat{z}$, where $\hat{z}$ is the estimate of z. A convenient method to obtain $\hat{z}$ is by a maximum-likelihood estimation [17] given by

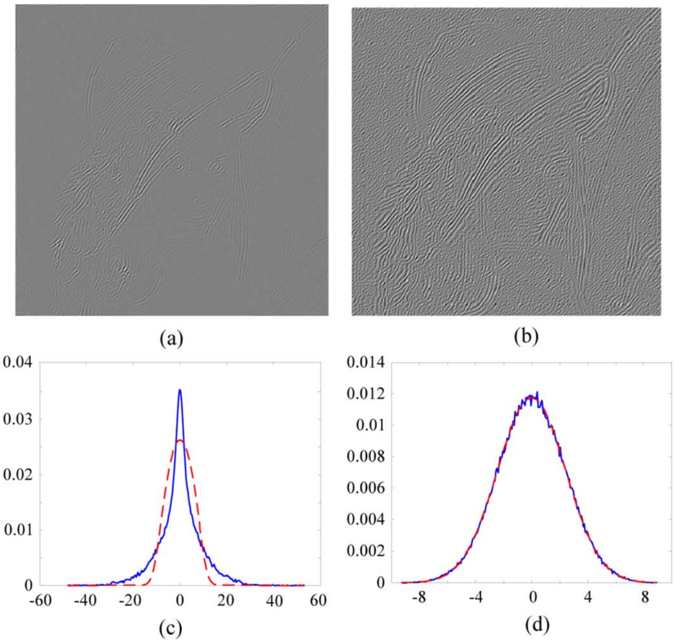


Fig. 2. (a) Original wavelet coefficients. (b) DNT coefficients. (c) Histogram of original coefficients (solid curve) and a Gaussian curve with the same standard deviation (dashed curve). (d) Histogram of DNT coefficients (solid) fitted with a Gaussian model (dashed).

$$\begin{aligned}\hat{z} &= \arg\max_{z}\{\log p(Y \mid z)\}\\ &= \arg\min_{z}\left\{N\log z + Y^TC_U^{-1}Y/2z^2\right\}\\ &= \sqrt{Y^TC_U^{-1}Y/N}\end{aligned} \tag{2}$$

where the covariance matrix $C_U=E[UU^T]$ is estimated from the entire wavelet subband before the estimation of local z, and N is the length of vector Y, or the number of neighboring wavelet coefficients.

B. Image Statistics in Divisive Normalization Transform Domain

As will be shown in the next section, our RRIQA approach is essentially based on the statistics of the transform coefficients in DNT domain and how they vary with image distortions. Before the development of the specific RRIQA algorithm, it is useful to observe variations of image statistics before and after the DNT is applied. In Fig. 2, we compare the marginal distributions of an original wavelet subband computed from a steerable pyramid decomposition [29] [Fig. 2(a)] and the same subband after DNT [Fig. 2(b)]. In Fig. 2(c), the original wavelet coefficient histogram is compared with a Gaussian shape that has the

TABLE I
KLD BETWEEN THE MARGINAL DISTRIBUTIONS OF WAVELET/DNT COEFFICIENTS AND GAUSSIAN FIT

	Lena	Barbara	Barco	Boat	House	Peppers	Fingerprint	Flintstones
Wavelet domain	0.4143	0.4301	0.5226	0.3848	0.4084	0.4722	0.0123	0.2436
DNT domain	0.0009	0.0125	0.0058	0.0098	0.0106	0.0082	0.0029	0.0034

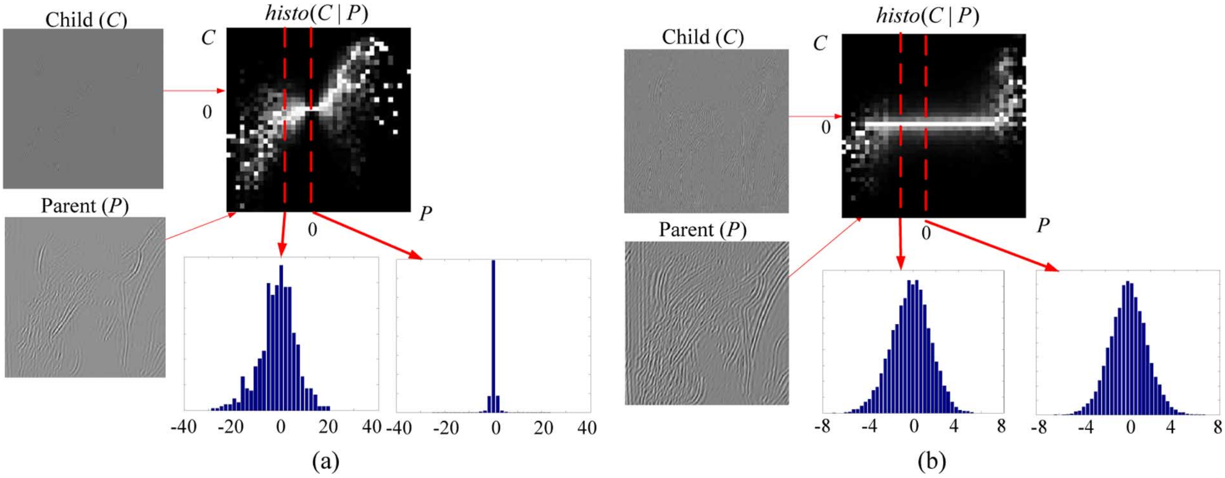


Fig. 3. (a) Conditional histograms between a parent and a child coefficients extracted from the original wavelet representation and (b) the corresponding DNT representation.

same standard deviation. The significant gap between the two curves indicates that the original wavelet coefficients are highly non-Gaussian. It has been shown that such histograms can be well-fitted with a generalized Gaussian density function (GGD) given by [30]

$$p_{\mathrm{GGD}}(x) = \frac{\lambda}{2\mu\Gamma(1/\lambda)} e^{-(|x|/\mu)^{\lambda}} \tag{3}$$

where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ (for $a > 0$) is the Gamma function, and λ and μ are called the scale and power factors, respectively. The Gaussian density is a special case of GGD when λ is fixed to be 2. However, for the histograms of the wavelet coefficients of natural images, the best fitting value of λ typically lies between 0.5 and 1.0 [31]. By contrast, the histogram of the coefficients after DNT can be well-fitted with a Gaussian, as demonstrated in Fig. 2(d). Similar behavior is observed for other natural images. To provide a quantitative measure, we compute the Kullback–Leibler distance (KLD) [32] between the histogram and the best-fitting Gaussian curve before and after DNT for a set of natural images. The results are shown in Table I, where we can see that Gaussian fit is consistently better in DNT domain for all test images.

Fig. 3 demonstrates the impact of DNT on the joint statistics of wavelet coefficients. In Fig. 3(a) and (b), we show the conditional histograms of the coefficients extracted from two neighboring subbands (a parent band and a child band) in the original wavelet decomposition and in the DNT representation, respectively. It can be observed that in the conditional histogram [$\mathrm{histo}(C|P)$ in Fig. 3(a)], the variance of a child coefficient (vertical axis) is highly dependent on the magnitude of its parent coefficient (horizontal axis). Such strong second-order variance dependency is confirmed by the significant difference between the widths of two cross-sections of the conditional histogram. By contrast, in the DNT representation, the histogram of the child coefficients makes little difference when conditioned on the magnitudes of the parent coefficients, as can be seen in Fig. 3(b). This demonstration clearly shows that the DNT representations can significantly reduce the second-order dependencies between the transform coefficients.

C. *Perceptual Relevance of Divisive Normalization Representation*

The DNT image representation is not only an effective way to reduce the statistical redundancies between wavelet coefficients, it is also highly relevant to biological vision. First, based on the widely accepted hypothesis that the early visual sensory processing is optimized to increase the statistical independence between neuronal responses (subject to certain physical limitations such as power consumption) through the evolution and development processes, the modeling of the biological visual system and the modeling of natural scene statistics are dual problems [19]–[21]. Second, in the context of neural physiology, it has been found that divisive normalization provides an effective model to account for many recorded data of cell responses in the visual cortex [22], [23]. It is also a useful framework in explaining the adaptations of neural responses with respect to the variations of the visual environment [33]. Third, in psychophysical vision, it has been shown that the divisive normalization procedure can well explain the visual masking effect [34], [35], where the visibility of an image component (e.g., a wavelet coefficient) is reduced in the presence of large neighboring components (e.g., the wavelet coefficients close in space, scale, and orientation). Furthermore, the perceptual relevance of DNT image representation has also been demonstrated by testing its resilience to noise contamination as well as its effectiveness in image compression and image contrast enhancement [18].

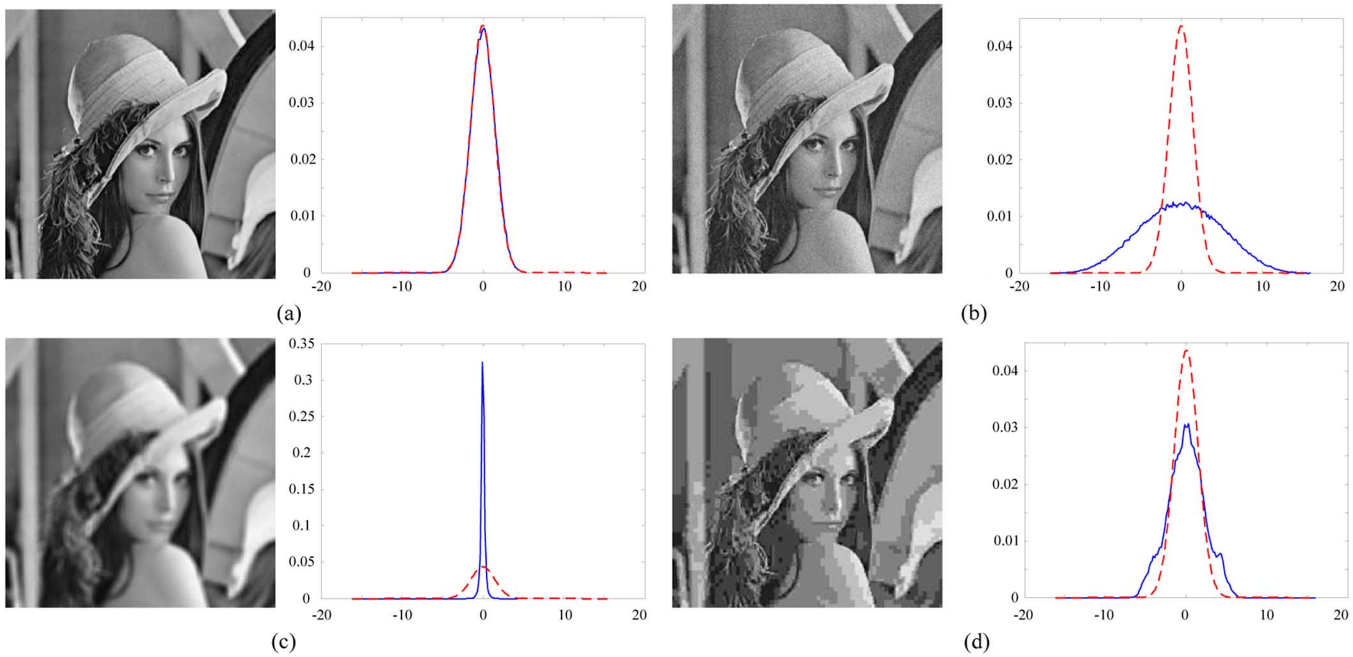


Fig. 4. Histograms of DNT coefficients in a wavelet subband under different types of image distortions. (a) Original "Lena" image. (b) Gaussian noise contaminated image. (c) Gaussain blurred image. (d) JPEG compressed image. Solid curves: histograms of DNT coefficients. Dashed curves: the Gaussian model fitted to the histogram of DNT coefficients in the original image. Significant departures from the Gaussian model is observed in the distorted images (b), (c), and (d).

III. Reduced-Reference Image Quality Assessment

A. DNT-Domain Statistics of Distorted Images

The strong perceptual and statistical relevance of DNT image representation provides good justifications for the use of DNT for RRIQA. In addition to that, we must also show that the statistics of DNT coefficients are sensitive to various image distortions. To study this, we apply DNT to a set of images with different types of distortions and observe how these distortions alter the statistics of the coefficients in DNT domain. This is demonstrated in Fig. 4, where the histogram of the DNT coefficients of a wavelet subband can be well-fitted with a Gaussian model [Fig. 4(a)]. However, when we draw the same Gaussian model together with the histogram of the DNT coefficients computed from Gaussian noise contaminated image [Fig. 4(b)], Gaussian blurred image [Fig. 4(c)], or JPEG compressed image [Fig. 4(d)], significant changes are observed. It is also interesting to see that the way the distribution changes varies with the distortion type. For example, Gaussian noise contamination increases the width of the histogram, but maintains the shape of Gaussian. By contrast, Gaussian blur reduces the width of the histogram and creates a much peakier distribution than Gaussian. These observations are important because our RRIQA algorithm is based on quantifying the variations of DNT-domain image statistics as a measure of image quality degradation.

B. Reduced-Reference Image Quality Assessment Algorithm

We propose an RRIQA algorithm by working with the marginal distributions of DNT coefficients. Although this algorithm still works with marginal distributions only (no explicit joint statistical model is employed, as in [9]), it does take into account the dependencies between the original neighboring wavelet coefficients because of the involvement of the divisive normalization process. We consider this as a major advantage of the proposed approach (as compared to [9]) in capturing the joint statistics of wavelet coefficients while maintaining the simplicity of the algorithm. Moreover, the algorithm has a low data rate, as only a small set of RR features are extracted from the reference image and are employed in quality evaluation of the distorted image.

A convenient approach to measure the variations of the marginal probability distributions of the DNT coefficients between the original and distorted images (as being observed in Fig. 4) is to compute the KLD between them

$$d(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{4}$$

where $p(x)$ and $q(x)$ are the probability density functions of the DNT coefficients in the same subband of the original and distorted images, respectively. To accomplish this, the DNT coefficient histograms of both the reference and distorted images must be available. The latter can be easily computed from the distorted image, which is always available. The difficulty is in obtaining the DNT coefficient histogram of the original image. Using all the histogram bins as RR features would result in either a heavy RR data rate (when the bin size is fine) or a poor approximation accuracy (when the bin size is coarse). To overcome this problem, we make use of the important property that the probability density function $p(x)$ of the original DNT coefficients can be well approximated with a zero-mean Gaussian model [as has been observed in Figs. 2(d) and 4(a)]

$$p_m(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{5}$$

This model provides a very efficient means to summarize the DNT coefficient histogram of the original image, such that only one parameter σ is needed to describe it (as opposed to all the histogram bins). Furthermore, to account for the variations between the model and the true distribution, we compute the KLD between $p_m(x)$ and $p(x)$ as

$$d(p_m||p) = \int p_m(x) \log \frac{p_m(x)}{P(x)} dx \tag{6}$$

and use it as an additional RR feature. This is computed for each subband independently, resulting in two parameters (σ and $d(p_m||p)$) for each subband.

In order to evaluate the quality of a distorted image, we estimate the KLD between the probability density function $q(x)$ of the DNT coefficients computed from the distorted image and the model $p_m(x)$ estimated from the original image

$$d(p_m||q) = \int p_m(x) \log \frac{p_m(x)}{q(x)} dx. \tag{7}$$

Combining this with the available RR feature $d(p_m||p)$, we obtain an estimate of the KLD between $p(x)$ and $q(x)$

$$\hat{d}(p||q) = d(p_m||q) - d(p_m||p). \tag{8}$$

It can be easily shown that

$$\hat{d}(p||q) = \int p_m(x) \log \frac{p_(x)}{q(x)} dx. \tag{9}$$

The estimation error can then be calculated as

$$d(p||q) - \hat{d}(p||q) = \int [p(x) - p_m(x)] \log \frac{p_(x)}{q(x)} dx. \tag{10}$$

This error is small when $p_m(x)$ and $p(x)$ are close, which is true for typical natural images. With the additional cost of adding one more RR parameter $d(p_m||p)$, (9) not only delivers a more accurate estimate of $d(p||q)$ than (7), but also provides a useful feature that when there is no distortion between the original and distorted images (which implies that $p(x) = q(x)$ for all x), then both the targeted distortion measure $d(p||q)$ and estimated distortion measure $\hat{d}(p||q)$ are exactly zero.

In addition to $\hat{d}(p||q)$, we also found the following measures useful in improving the accuracy of image quality evaluation:

$$d_\sigma = |\sigma_o - \sigma_d| \tag{11}$$

$$d_k = |k_o - k_d| \tag{12}$$

$$d_s = |s_o - s_d| \tag{13}$$

where σ_o, k_o, s_o, and σ_d, k_d, s_d are the standard deviation, the kurtosis (the fourth-order central moment divided by the fourth power of the standard deviation and then minus 3), and the skewness (the third-order central moment divided by the third power of the standard deviation) of the DNT coefficients computed from the original and distorted images, respectively. These measures provide further information about the shape changes of the probability density functions. In particular, two images with the same KLD with respect to the original image may have different types of distortions, and visual quality assessment varies across distortion types. Adding these features not only provides new means to quantify the amount of distortions, but also supplies new information that helps the algorithm differentiate distortion types. We have also carried out experiments to compare our IQA algorithm with and without these features, and we found that adding these features lead to significant improvement in terms of the performance of image quality prediction. Since σ_d, k_d, s_d can be computed from the available distorted image and σ is already acquired when fitting the Gaussian model of (5), only two new RR features, k_o and s_o, are added. Indeed, both of them are close to zero because the probability distribution of DNT coefficients of the original image is approximately Gaussian, which has zero skewness and kurtosis.

At each subband, we define the overall image distortion measure as a linear combination of $\hat{d}(p||q)$, d_σ, d_k and d_s in the logarithmic domain

$$\begin{aligned} D_{\text{band}} &= \alpha \log(\hat{d}(p||q)) + \beta \log(d_\sigma) \\ &\quad + \gamma \log d_k + \delta \log d_s \\ &= \log((\hat{d}(p||q))^\alpha (d_\sigma)^\beta (d_k)^\gamma (d_s)^\delta) \end{aligned} \tag{14}$$

where α, β, γ, and δ are weighting parameters. Finally, the overall distortion of the distorted image is computed as the sum of the distortion measures of all subbands

$$D = \sum_{\text{all subbands}} D_{\text{band}}. \tag{15}$$

C. Implementation Issues

To compute the DNT representation of an image, we first decompose the image using a steerable pyramid [29] with three scales and four orientations, as shown in Fig. 5. For each center coefficient y_c at each subband, we define a DNT neighboring vector Y that contains 13 coefficients, including nine from the same subband (including the center coefficient itself), one from the parent band, and three from the same spatial location in the other orientation bands at the same scale. An illustration is given in Fig. 5. These coefficients are selected from the direct neighbors of the center coefficient because the magnitudes of clusters of wavelet coefficients tend to scale together [20] and thus are more likely to share the same scale factor z in the GSM model described earlier. Increasing the size of the neighborhood will increase the computational complexity of DNT calculation [specifically, the estimation of $\hat{z}$ in (2)], but will not add extra RR features (because it only affects the DNT computation and all other processes after DNT remain unaltered). In our experiments, we did not observe significant variations of the overall performance of the algorithm under slight changes of the neighborhood, but more careful study on this issue remains future work. After the DNT computation, four RR features are extracted from each subband of the original image, including σ, $d(p_m||p)$, k_o and s_o. This results in a total of 48 scalar RR features for each original image.

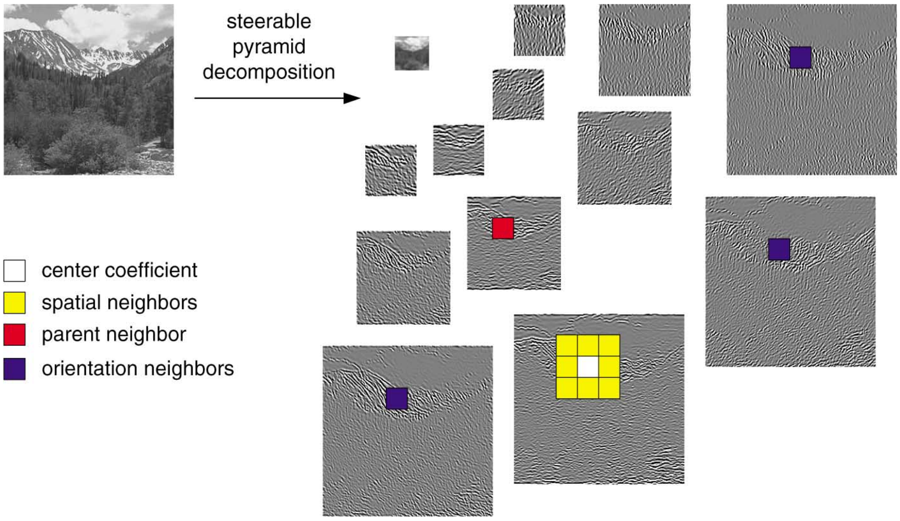


Fig. 5. Illustration of steerable pyramid decomposition and the selection of DNT neighbors. The neighboring coefficients include the 3×3 spatial neighbors within the same subband, one parent neighboring coefficient and three orientation neighboring coefficients.

The evaluation of the KLD between probability density functions needs to be done numerically using histograms. For example, for (6), we compute

$$d(p_m \| p) = \sum_{i=1}^{L} P_m(i) \log \frac{P_m(i)}{P(i)} \tag{16}$$

where $P(i)$ and $P_m(i)$ are the normalized heights of the ith histogram bins, and L is the number of bins in the histograms.

One problem with the subband quality measure of (14) is that when $\hat{d}(p\|q)$, d_σ, d_k, or d_s is close to zero, the measure becomes unstable. In our implementation, to avoid such instability, we compute

$$D_{\text{band}} = \log\left(1 + \frac{(\hat{d}(p\|q))^{\alpha}(d_\sigma)^{\beta}(d_k)^{\gamma}(d_s)^{\delta}}{D_0}\right) \tag{17}$$

where D_0 is a positive constant. Another useful property of this formulation is that the resulting distortion measure is always non-negative, and is zero when the original and distorted images are exactly the same.

Before applying the proposed algorithm for image quality assessment, five parameters, $\alpha, \beta, \gamma, \delta$, and D_0, need to be learned from the data. It is important to cross-validate these parameters with different selections of the training and testing data. Details will be given in the next section. For a given set of training images and the associated subjective scores, we use the Matlab nonlinear optimization routine *fminsearch* in the optimization toolbox to find the optimal parameters.

IV. Validation

To validate the proposed RRIQA algorithm, two publicly-accessible subject-rated image databases are used, which are the LIVE database [36] developed at Laboratory for Image and Video Engineering at The University of Texas at Austin and the Cornell-VCL A57 database [37] developed at the Visual Communications Laboratory at Cornell University. The LIVE database contains seven datasets of 982 subject-rated images created from 29 original images with five types of distortions at different distortion levels. The distortion types include 1) JP2: JPEG2000 compression (2 sets), 2) JPG: JPEG compression (2 sets), 3) Noise: white noise contamination, 4) Blur: Gaussian blur, and 5) FF: fast fading channel distortion of JPEG2000 compressed bitstream. The subjective test was carried out with each of the seven data sets individually. A cross-comparison set that mixes images from all distortion types is then used to help align the subject scores across different data sets. The subjective scores of all images are then adjusted according to this alignment. The alignment process is rather crude. However, the aligned subjective scores (all data) are still very useful references, which are particularly important for testing general-purpose IQA algorithms, for which cross-distortion comparisons are highly desirable. In the Cornell-VCL database, there are totally 60 distorted images generated from three original images. Six different types of distortions are included, which are 1) FLT: quantization of the LH subbands of a five-level DWT of the image using the 9/7 filters, where the bands were quantized via uniform scalar quantization with step sizes chosen such that the RMS contrast of the distortions was equal, 2) NOZ: additive Gaussian white noise, 3) JPG: baseline JPEG compression, 4) JP2: JPEG2000 compression using the 9/7 filters and no visual frequency weighting; 5) DCQ: JPEG2000 compression using the 9/7 filters with the dynamic contrast-based quantization algorithm, which applies greater quantization to the fine spatial scales relative to the coarse scales in an attempt to preserve global precedence, and 6) BLR: blurring by using a Gaussian filter.

Three criteria are used to evaluate how well the objective scores predict the subjective scores: 1) Correlation coefficient (CC) between the subjective/objective scores after a nonlinear mapping is computed to evaluate prediction accuracy, 2) Spearman rank-order correlation coefficient (ROCC) is calculated to evaluate prediction monotonicity, 3) Outlier ratio is used to evaluate prediction consistency, which is defined as

TABLE II
WAVELET AND DNT DOMAIN COMPARISON OF THE PROPOSED METHODS USING THE LIVE DATABASE

LIVE data set		JP2(1)	JP2(2)	JPG(1)	JPG(1)	Noise	Blur	FF	All data
		Correlation Coefficient (prediction accuracy)							
Proposed	wavelet + GGD	0.9115	0.9422	0.8501	0.9354	0.9401	0.8773	0.9243	0.8930
Proposed	DNT + Gaussian	0.9485	0.9655	0.8203	0.9579	0.9654	0.9562	0.9464	0.9173
		Rank-Order Correlation Coefficient (prediction monotonicity)							
Proposed	wavelet + GGD	0.9081	0.9239	0.8389	0.8734	0.9316	0.8608	0.9237	0.9093
Proposed	DNT + Gaussian	0.9478	0.9610	0.8143	0.8937	0.9559	0.9584	0.9443	0.9287
		Outlier Ratio (prediction consistency)							
Proposed	wavelet + GGD	0.0230	0.0122	0.0805	0.1250	0.0345	0.0483	0.0345	0.1853
Proposed	DNT + Gaussian	0.0115	0.0122	0.1149	0.0341	0.0000	0.0000	0.0207	0.1069

TABLE III
WAVELET AND DNT DOMAIN COMPARISON OF THE PROPOSED METHODS USING THE CORNELL-VCL DATABASE

Cornell-VCL data set		FLT	JPG	JP2	DCQ	BLR	NOZ	All data
		Correlation Coefficient (prediction accuracy)						
Proposed	wavelet + GGD	0.4592	0.8303	0.7802	0.8808	0.9270	0.7748	0.5125
Proposed	DNT + Gaussian	0.7630	0.9108	0.8185	0.9095	0.9340	0.9900	0.6635
		Rank-Order Correlation Coefficient (prediction monotonicity)						
Proposed	wavelet + GGD	0.4167	0.7833	0.8333	0.8833	0.7500	0.7333	0.5134
Proposed	DNT + Gaussian	0.5000	0.7667	0.8000	0.6667	0.8000	0.9833	0.7018

TABLE IV
PERFORMANCE COMPARISON OF IQA ALGORITHMS USING THE LIVE DATABASE

LIVE data set	JP2(1)	JP2(2)	JPG(1)	JPG(1)	Noise	Blur	FF	All data
	Correlation Coefficient (prediction accuracy)							
PSNR	0.9337	0.8948	0.9015	0.9136	0.9866	0.7742	0.8811	0.8709
Wang *et al.* [9]	0.9353	0.9490	0.8452	0.9695	0.8889	0.8872	0.9175	0.8226
Proposed (training: Cornell-VCL)	0.9115	0.9422	0.8501	0.9354	0.9401	0.8773	0.9243	0.8930
Proposed (training: LIVE)	0.9485	0.9655	0.8203	0.9579	0.9654	0.9562	0.9464	0.9173
	Rank-Order Correlation Coefficient (prediction monotonicity)							
PSNR	0.9231	0.8816	0.8907	0.8077	0.9855	0.7729	0.8785	0.8755
Wang *et al.* [9]	0.9298	0.9470	0.8332	0.8908	0.8639	0.9145	0.9162	0.8437
Proposed (training: Cornell-VCL)	0.9081	0.9239	0.8389	0.8734	0.9316	0.8608	0.9237	0.9093
Proposed (training: LIVE)	0.9478	0.9610	0.8143	0.8937	0.9559	0.9584	0.9443	0.9287
	Outlier Ratio (prediction consistency)							
PSNR	0.0805	0.0976	0.092	0.1818	0.0000	0.2069	0.1517	0.2373
Wang *et al.* [9]	0.0690	0.0366	0.1839	0.0341	0.1793	0.1172	0.0621	0.2311
Proposed (training: Cornell-VCL)	0.0230	0.0122	0.0805	0.1250	0.0345	0.0483	0.0345	0.1853
Proposed (training: LIVE)	0.0115	0.0122	0.1149	0.0341	0.0000	0.0000	0.0207	0.1069

the percentage of predictions outside the range of ±2 standard deviations between subjective scores. These criteria had been used in the previous tests conducted by the video quality expert group [38]. Since we do not have access to the raw subjective scores of the Cornell-VCL database, the standard deviations between subjective scores for each test image cannot be computed. Therefore, only CC and ROCC comparisons are included for the Cornell-VCL database.

Our validation work has two major purposes. The first is to verify that using DNT image representation is beneficiary for the improvement of IQA algorithms. The second is to compare the performance of the proposed method with existing IQA algorithms.

To show the impact of DNT representation, we compare the performance of the proposed RRIQA algorithm implemented in the wavelet domain (linear steerable pyramid decomposition) and in the DNT domain (linear steerable pyramid decomposition, followed by the nonlinear DNT process). Specifically, GGD is used to model the marginal distribution of wavelet coefficients and Gaussian density is employed to model that of DNT coefficients. All other aspects of the algorithm, including the standard deviation, skewness and kurtosis features, the KLD measure, the subband and overall quality measurement approach, and the training data and process, are exactly the same. The test results on the LIVE database and the Cornell-VCL database are shown in Tables II and III, respectively, where the training data are the full LIVE database and the full Cornell-VCL database, respectively. It can be concluded from these tables that the overall performance is clearly improved from wavelet-domain to DNT-domain implementations.

The performance comparison with other IQA algorithms is shown in Tables IV and V. To the best of our knowledge, the only other RRIQA algorithm that has a comparable low RR data rate and is designed for general-purpose is the method proposed in [9]. In addition to this method, we have also included peak signal-to-noise-ratio (PSNR), which is still the most widely used full-reference IQA measure. Although such comparison is highly unfair to the proposed method and the method in [9] (PSNR requires full access to the original image, as opposed to the 48 scalar features in the proposed method), it provides a useful indication of the relative performance of the proposed algorithm. For any IQA algorithm that involves a training process of the parameters, it is important to verify that the model is not overtrained. In other words, the performance of the algorithm should not change dramatically with different training data set. Therefore, in both Tables IV and V, we have

TABLE V
PERFORMANCE COMPARISON OF IQA ALGORITHMS USING THE CORNELL-VCL DATABASE

Cornell-VCL data set	FLT	JPG	JPG2	DCQ	BLR	NOZ	All data
	Correlation Coefficient (prediction accuracy)						
PSNR	0.9100	0.7008	0.7957	0.5637	0.5904	0.9340	0.6347
Wang *et al.* [9]	0.4939	0.8575	0.7880	0.9357	0.7687	0.6252	0.3166
Proposed (training: LIVE)	0.4864	0.9183	0.8813	0.8847	0.8602	0.9489	0.5385
Proposed (training: Cornell-VCL)	0.7630	0.9108	0.8185	0.9095	0.9340	0.9900	0.6635
	Rank-Order Correlation Coefficient (prediction monotonicity)						
PSNR	0.9000	0.6333	0.8000	0.5000	0.4667	0.9500	0.6205
Wang *et al.* [9]	0.1000	0.7667	0.5333	0.8000	0.6667	0.7333	0.2948
Proposed (training: LIVE)	0.1500	0.7833	0.8667	0.7333	0.7500	0.9500	0.5110
Proposed (training: Cornell-VCL)	0.5000	0.7667	0.8000	0.6667	0.8000	0.9833	0.7018

included two versions of the proposed DNT-domain algorithm, where the only difference between them is that their model parameters (α, β, γ, δ, and D_0) are trained with the LIVE database or the Cornell-VCL database (using all images in both cases). Such a cross-validation process is useful to test the robustness of the model. Not surprisingly, the test results are better when the parameters are trained with the same database than the results obtained by cross-training the parameters (Note that some image distortion types included in one database may not be included in the other). However, in both cases and for both databases, the proposed algorithm performs better than the method in [9]. In particular, it can be seen from both Tables IV and V that for the all-data cases, where all the images with different distortion types are mixed together, the method in [9] does not perform well, and the improvement of the proposed method is quite significant. Indeed, its CC and ROCC values (for all-data cases) are comparable or even higher than the full-reference PSNR measure.

V. CONCLUSION AND DISCUSSION

We proposed an RRIQA algorithm using statistical features extracted from a divisive normalization-based image representation. We demonstrate that such a DNT image representation has simultaneous perceptual and statistical relevance and its statistical properties are significantly changed under different types of image distortions. These properties make it well-suited for the development of RRIQA algorithms. Experimental verifications with publicly-accessible subject-rated image databases suggest that this new image representation leads to improved performance in the evaluation of image quality. The proposed algorithm has a relatively low data rate for RR features. It does not make any assumption about the image distortion types, thus has the potential to be used for general-purpose in a wide range of applications.

Several further questions may be asked from this work. First, while the statistical features used in the proposed algorithm seem to be perceptually relevant and useful for IQA, is there any better means to combine them into a single scalar quality measure of the distorted image? Second, other than the variance dependency that are well-captured by DNT, there are many other types of dependencies between neighboring wavelet coefficients that are still missing, for example, local phase coherence [39]. Is there any efficient way to incorporate these dependencies as well? Third, using the proposed RRIQA measure, together with the statistical properties (RR features) about the "perfect-quality" original image, can we design image quality enhancement method that can correct or improve the quality of the distorted image being evaluated? Finally, since the proposed RRIQA method is relevant to the quantification of the naturalness of images and does not use any knowledge about image distortion types, would it be possible to further develop it into a general-purpose no-reference image quality assessment method?

REFERENCES

[1] Z. Wang and A. C. Bovik*, Modern Image Quality Assessment*. San Rafael, CA: Morgan & Claypool, Mar. 2006.

[2] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 11, pp. 317–320, Nov. 1997.

[3] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2000, vol. 3, pp. 981–984.

[4] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifact in digital video," *Proc. IEEE*, vol. 90, pp. 154–169, Jan. 2002.

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

[6] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process.: Image Commun.*, vol. 19, pp. 163–172, Feb. 2004.

[7] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Blind quality assessment for JPEG2000 compressed images," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2002, pp. 1403–1407.

[8] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

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

[10] S. Wolf and M. H. Pinson, "Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system," *Proc. SPIE*, vol. 3845, pp. 266–277, 1999.

[11] I. P. Gunawan and M. Ghanbari, "Reduced reference picture quality estimation by using local harmonic amplitude information," in *Proc. London Commun. Symp.*, Sep. 2003, pp. 137–140.

[12] T. M. Kusuma and H.-J. Zepernick, "A reduced-reference perceptual quality metric for in-service image quality assessment," in *Proc. Joint 1st Workshop Mobile Future and Symp. Trends Commun.*, Oct. 2003, pp. 71–74.

[13] S. Wolf and M. Pinson, "Low bandwidth reduced reference video quality monitoring system," in *Proc. Int. Workshop Video Process. Quality Metrics for Consumer Electron.*, Scottsdale, AZ, Jan. 2005, CD-ROM.

[14] P. Le Callet, C. Viard-Gaudin, and D. Barba, "Continuous quality assessment of MPEG2 video with reduced reference," in *Proc. Int. Workshop Video Process. Quality Metrics for Consumer Electron.*, Scottsdale, AZ, Jan. 2005.

[15] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, vol. 3, pp. 185–188.

[16] M. Carnec, P. Le Callet, and D. Barba, “Visual features for image quality assessment with reduced reference,” in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, vol. 1, pp. 421–424.

[17] M. J. Wainwright and E. P. Simoncelli, “Scale mixtures of gaussians and the statistics of natural images,” *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 855–861, 2000.

[18] S. Lyu and E. P. Simoncelli, “Statistically and perceptually motivated nonlinear image representation,” in *Proc. SPIE Conf. Human Vision Electron. Imaging XII*, Jan. 2007, vol. 6492, pp. 649207–1–649207–15.

[19] H. B. Barlow, , W. A. Rosenblith, Ed., “Possible principles underlying the transformation of sensory messages,” in *Sensory Commun.*. Cambridge, MA: MIT Press, 1961, pp. 217–234.

[20] E. P. Simoncelli and B. Olshausen, “Natural image statistics and neural representation,” *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.

[21] O. Schwartz and E. P. Simoncelli, “Natural signal statistics and sensory gain control,” *Nature: Neurosci.*, vol. 4, pp. 819–825, Aug. 2001.

[22] D. J. Heeger, “Normalization of cell responses in cat striate cortex,” *Vis. Neural Sci.*, vol. 9, pp. 181–198, 1992.

[23] E. P. Simoncelli and D. J. Heeger, “A model of neuronal responses in visual area MT,” *Vis. Res.*, vol. 38, pp. 743–761, Mar. 1998.

[24] D. L. Ruderman, “The statistics of natural images,” *Network: Comput. Neural Syst.*, vol. 5, pp. 517–548, 1996.

[25] J. Malo, I. Epifanio, R. Navarro, and E. P. Simoncelli, “Non-linear image representation for efficient perceptual coding,” *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 68–80, Jan. 2006.

[26] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[27] J. Portilla and E. P. Simoncelli, “Image restoration using Gaussian scale mixtures in the wavelet domain,” in *Proc. IEEE Int. Conf. Image Process.*, Barcelona, Spain, Sep. 2003, vol. 2, pp. 965–968.

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

[29] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multi-scale transforms,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[30] S. G. Mallat, “Multifrequency channel decomposition of images and wavelet models,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2091–2110, Dec. 1989.

[31] E. P. Simoncelli and E. H. Adelson, “Noise removal via Bayesian wavelet coring,” in *Third Int. Conf. Image Process.*, Lausanne, Switzerland, Sep. 1996, vol. I, pp. 379–382, IEEE Signal Process. Soc..

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[33] M. J. Wainwright, “Visual adaptation as optimal information transmission,” *Vis. Res.*, vol. 39, pp. 3960–3974, 1999.

[34] J. Foley, “Human luminance pattern mechanisms: Masking experiments require a new model,” *J. Opt. Soc. Amer.*, vol. 11, no. 6, pp. 1710–1719, 1994.

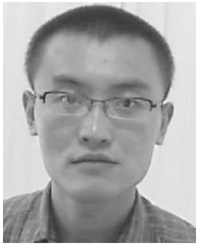
[35] A. B. Watson and J. A. Solomon, “Model of visual contrast gain control and pattern masking,” *J. Opt. Soc. Amer.*, vol. 14, no. 9, pp. 2379–2391, 1997.

[36] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, “Image and Video Quality Assessment Research at LIVE.” [Online]. Available: http://live.ece. utexas.edu/research/quality/

[37] D. M. Chandler and S. S. Hemami, “VSNR: A wavelet-based visual signal-to-noise ratio for natural images,” [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[38] P. Corriveau *et al.*, “Video quality experts group: Current results and future directions,” in *Proc. SPIE Visual Commun. Image Process.*, Jun. 2000, vol. 4067, pp. 742–753.

[39] Z. Wang and E. P. Simoncelli, “Local phase coherence and the perception of blur,” in *Adv. Neural Inf. Process. Syst. (NIPS03)*. Cambridge, MA: MIT Press, May 2004, vol. 16.



Qiang Li (S’06) received the B.S. and M.S. degrees from the Beijing Institute of Technology, Beijing, China, in 2000 and 2003, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at The University of Texas, Arlington.

His research interests include full-reference and reduced-reference quality assessment and statistical models of the natural scene image and their application to image and video processing problems.

Mr. Li is a recipient of the IBM Student Paper Award at the 2008 IEEE International Conference on Image Processing.

Zhou Wang (S’99–A’01–M’02) received the Ph.D. degree from The University of Texas at Austin in 2001.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Before that, he was an Assistant Professor in the Department of Electrical Engineering, The University of Texas at Arlington, a Research Associate at Howard Hughes Medical Institute and New York University, and a Research Engineer at AutoQuant Imaging, Inc. His research interests include image processing, coding, communication, and quality assessment; computational vision and pattern analysis; multimedia coding and communications; and biomedical signal processing. He has more than 60 publications and one U.S. patent in these fields and is an author of *Modern Image Quality Assessment* (Morgan & Claypool, 2006).

Prof. Wang is an Associate Editor of IEEE Signal Processing Letters and *Pattern Recognition*, and a Guest Editor of IEEE Journal of Selected Topics in Signal Processing: Special Issue on Visual Media Quality Assessment.