Generalized Gaussian scale mixtures: A model for wavelet coefficients of natural images

Praful Gupta a,*, Anush Krishna Moorthy b, Rajiv Soundararajan c, Alan Conrad Bovik a,1

a Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, USA
b Netflix Inc., USA
c Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India

A R T I C L E   I N F O

Keywords:
Generalized Gaussian scale mixture model
Distorted image modeling
Distortion identification
No-reference image quality assessment

A B S T R A C T

We develop a Generalized Gaussian scale mixture (GGSM) model of the wavelet coefficients of natural and distorted images. The GGSM model, which is more general than and which subsumes the Gaussian scale mixture (GSM) model, is shown to be a better representation of the statistics of the wavelet coefficients of both natural as well as distorted images. We demonstrate the utility of the model by applying it to various image processing applications, including blind distortion identification and no reference image quality assessment (NR-IQA). Similar to the GSM model, the GGSM model is useful for motivating the use of local divisive energy normalization, especially when the wavelet coefficients are computed on distorted pictures. We show that the GGSM model can lead to improved performance in distortion-related applications, while providing a more principled approach to the statistical processing of distorted image signals. The software release of a GGSM-based NR-IQA approach called DIIVINE-GGSM is available online at http://live.ece.utexas.edu/research/quality/diivine-ggsm.zip for further experimentation.

1. Introduction

A popular theory in visual neuroscience is that because the human visual system has evolved in the natural environment, many of its properties have, over time, adapted to the statistical properties of the natural environment that the eyes are exposed to [1,2]. One could study the statistics of natural images and draw certain conclusions regarding the properties of the environment. Such conclusions could then be correlated with results from studies of the human visual system, and hypotheses regarding the function of cortical processing could be posited [2–4]. On the other hand, one could posit a hypothesis on the goal of sensory coding, and by quantifying the goal in information theoretic terms, attempt to describe the statistical properties of the natural environment, and analyze the obtained results [5]. While both approaches lead to interesting observations regarding human visual processing, such study of scene statistics has tremendous relevance for perceptual image processing. Such notions, for instance, have been successfully applied to the design of a number of perceptual image quality assessment (IQA) algorithms [6–10].

Many researchers have studied and modeled the scene statistics of natural images subjected to a scale-space-orientation decomposition (loosely, a bandpass or wavelet transform). It is a well known observation that the (marginal) coefficient distributions of wavelet filters tend to follow an approximate Laplacian distribution (i.e., more heavy tailed than a Gaussian) [11]. However, simple marginal statistics do not capture the statistical regularities that exist across intra and inter band wavelet coefficient neighbors. One model for wavelet coefficients that not only models the marginal distributions, but also the relationships between neighboring coefficients in the same subband, and those between adjacent subbands is the Gaussian scale mixture (GSM) model [2,11]. In the GSM model, a set of local wavelet coefficients are modeled using a scale mixture and in [11,12], the authors demonstrate that this semi-parametric model satisfies the dual requirement of being able to describe the heavy-tailed nature of wavelet coefficients as well as the multiplicative scaling between neighboring coefficients. The GSM model has been successfully used for image denoising [13], image restoration [14,15], full-reference (FR) IQA [16,17], reduced-reference (RR) IQA [18–20] and no-reference (NR) IQA [21–25]. It has also been applied to the video QA problem [26–30], the stereoscopic 3D QA...
Once this multiplier is estimated, one can perform a ‘divisive normalization’ of neighboring bands at the same scale. The mixing multiplier denotes equality in probability distribution, \( \mathbf{u} \equiv \sqrt{2} \cdot \mathbf{x} \), where \( \equiv \) denotes equality in probability distribution, \( \mathbf{u} \) is a zero-mean Gaussian random vector with covariance \( \mathbf{Z}_\nu \) and \( \mathbf{z} \) is a scalar random variable called a mixing multiplier. The density of \( \mathbf{x} \) is then given by

\[
p_x(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\mathbf{Z}_\nu|}|z|^{d/2}} \exp\left(-\frac{\mathbf{x}^\top \mathbf{Z}_\nu^{-1} \mathbf{x}}{2z}\right) \, p_z(z) \, dz,
\]

where \( p_z(z) \) is the density of \( z \). In the case of natural images, the vector \( \mathbf{x} \) is formed by clustering a set of neighboring wavelet coefficients within a subband, or across neighboring subbands in scale and orientation [11]. The GSM model of natural wavelet coefficients has been successfully applied to noise estimation [35], denoising [13] and image quality assessment [19,22,23,25,36–38]. IQA algorithms that utilize other bandpass decompositions have also been extensively studied [10,39–41]. While the GSM model is robust, it is only an approximate model of image wavelet coefficients. In order to demonstrate this, we analyze the GSM model as applied on pristine natural images from the LIVE image quality assessment (IQA) database [42], utilizing the steerable pyramid decomposition [43], which is an overcomplete bandpass wavelet transform that in the implementation we use, decomposes images over two scales and six orientations. In our analysis (as depicted in Fig. 1), the vector \( \mathbf{x} \) contains 15 coefficients, including 9 from the same subband (a 3 × 3 neighborhood around \( \mathbf{x}_c \) — the center coefficient), 1 from the parent band, and 5 from the same spatial location in the neighboring bands at the same scale. The mixing multiplier \( z \) may be estimated as \( \hat{z} = \mathbf{x}^\top \mathbf{Z}_\nu^{-1} \mathbf{x} / d \) [18,44]. Once this multiplier is estimated, one can perform a ‘divisive normalization transform’, whereby each center coefficient, \( \mathbf{x}_c \), is divided by the local energy estimate \( \hat{z} \). We performed such a divisive normalization process on the wavelet coefficients of each of the 29 LIVE [42] reference images, then estimated the shape and scale parameters of the resulting generalized Gaussian distribution to verify whether the divisively normalized coefficients could indeed be regarded as Gaussian with high (95%) confidence.

The results in Table 1 imply that, while the GSM model of wavelet coefficients of natural images is a fairly robust model, it is by no means comprehensive. Less than 60% of the images follow the wavelet GSM model over all scales. Even amongst un-distorted, pristine, natural images, the model fails to produce the predicted Gaussian (divisively-normalized) response distribution. Indeed, as the authors in [11] note, the divisive normalization process applied to a GSM only ensures that the resulting marginal coefficients are approximately Gaussian. This implies that there is a need for a more general class of models that would not only subsume the GSM model, but will also extend to those situations where the GSM model is inadequate — for example, in modeling the statistical properties of distorted images. In the next section, we describe such a general class of scale-mixtures — the Generalized Gaussian scale mixtures (GGSM) — which can not only be used to model the statistical relationships between the bandpass coefficients of natural images, but also those from distorted images.

### 2. Modeling wavelet coefficients

#### 2.1. Gaussian scale mixtures

A \( d \)-dimensional random vector \( \mathbf{x} \) is a GSM if \( \mathbf{x} \equiv \sqrt{2} \cdot \mathbf{u} \), where \( \equiv \) denotes equality in probability distribution, \( \mathbf{u} \) is a zero-mean Gaussian random vector with covariance \( \mathbf{Z}_\nu \) and \( z \) is a scalar random variable called a mixing multiplier. The density of \( \mathbf{x} \) is then given by

\[
p_x(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\mathbf{Z}_\nu|}|z|^{d/2}} \exp\left(-\frac{\mathbf{x}^\top \mathbf{Z}_\nu^{-1} \mathbf{x}}{2z}\right) \, p_z(z) \, dz,
\]

where \( p_z(z) \) is the density of \( z \). In the case of natural images, the vector \( \mathbf{x} \) is formed by clustering a set of neighboring wavelet coefficients within a subband, or across neighboring subbands in scale and orientation [11]. The GSM model of natural wavelet coefficients has been successfully applied to noise estimation [35], denoising [13] and image quality assessment [19,22,23,25,36–38]. IQA algorithms that utilize other bandpass decompositions have also been extensively studied [10,39–41]. While the GSM model is robust, it is only an approximate model of image wavelet coefficients. In order to demonstrate this, we analyze the GSM model as applied on pristine natural images from the LIVE image quality assessment (IQA) database [42], utilizing the steerable pyramid decomposition [43], which is an overcomplete bandpass wavelet transform that in the implementation we use, decomposes images over two scales and six orientations. In our analysis (as depicted in Fig. 1), the vector \( \mathbf{x} \) contains 15 coefficients, including 9 from the same subband (a 3 × 3 neighborhood around \( \mathbf{x}_c \) — the center coefficient), 1 from the parent band, and 5 from the same spatial location in the neighboring bands at the same scale. The mixing multiplier \( z \) may be estimated as \( \hat{z} = \mathbf{x}^\top \mathbf{Z}_\nu^{-1} \mathbf{x} / d \) [18,44]. Once this multiplier is estimated, one can perform a ‘divisive normalization transform’, whereby each center coefficient, \( \mathbf{x}_c \), is divided by the local energy estimate \( \hat{z} \). We performed such a divisive normalization process on the wavelet coefficients of each of the 29 LIVE [42] reference images, then estimated the shape and scale parameters of the resulting generalized Gaussian distribution to verify whether the divisively normalized coefficients could indeed be regarded as Gaussian with high (95%) confidence.

The results in Table 1 imply that, while the GSM model of wavelet coefficients of natural images is a fairly robust model, it is by no means comprehensive. Less than 60% of the images follow the wavelet GSM model over all scales. Even amongst un-distorted, pristine, natural images, the model fails to produce the predicted Gaussian (divisively-normalized) response distribution. Indeed, as the authors in [11] note, the divisive normalization process applied to a GSM only ensures that the resulting marginal coefficients are approximately Gaussian. This implies that there is a need for a more general class of models that would not only subsume the GSM model, but will also extend to those situations where the GSM model is inadequate — for example, in modeling the statistical properties of distorted images. In the next section, we describe such a general class of scale-mixtures — the Generalized Gaussian scale mixtures (GGSM) — which can not only be used to model the statistical relationships between the bandpass coefficients of natural images, but also those from distorted images.

#### 2.2. Generalized Gaussian scale mixtures

Before we describe the generalized Gaussian scale mixture (GGSM) model, we set up the multivariate generalized Gaussian (MVGG) distribution which we shall use with the GGSM model. We also detail the parameter estimation procedure which will be of use in later sections of this paper.

##### 2.2.1. The multivariate generalized Gaussian distribution

There exist multiple definitions of the Multivariate Generalized Gaussian (MVGG) distribution in the literature [45–47]. Here we consider a particular case of the Kotz-type distribution, a multivariate elliptical distribution [45] which has been used to model the statistics of wavelet coefficients in the past [48]. The zero-mean MVGG is defined as

\[
p_x(x) = \frac{1}{\pi^{d/2} F \left( \frac{d}{2}, \frac{d}{2} \right)} \exp\left(-\frac{1}{2} (\mathbf{x}^\top \mathbf{Z}^{-1} \mathbf{x}) \right),
\]

$$d$$ is the dimension of the random vector, \( \mathbf{Z} \) is the covariance matrix, \( F \) is the multivariate gamma function, and \( \mathbf{x} \) is a random vector of dimension \( d \).

### Table 1

Percentage of divisively normalized wavelet coefficients of natural image at different scales and orientations that were deemed to be Gaussian with high confidence.

<table>
<thead>
<tr>
<th>Scale</th>
<th>30°</th>
<th>60°</th>
<th>90°</th>
<th>120°</th>
<th>150°</th>
<th>180°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale 1</td>
<td>13.79%</td>
<td>51.72%</td>
<td>58.62%</td>
<td>0%</td>
<td>44.83%</td>
<td>44.83%</td>
</tr>
<tr>
<td>Scale 2</td>
<td>3.45%</td>
<td>34.48%</td>
<td>34.48%</td>
<td>0%</td>
<td>41.38%</td>
<td>34.48%</td>
</tr>
<tr>
<td>Scale 3</td>
<td>3.45%</td>
<td>27.59%</td>
<td>44.83%</td>
<td>10.34%</td>
<td>34.48%</td>
<td>31.03%</td>
</tr>
</tbody>
</table>

**Fig. 1.** The neighborhood structure used to construct GSM/GGSM vectors in the experiments. The GSM/GGSM vector \( \mathbf{Y} \) contains 15 coefficients, including 9 from the same subband (3 × 3 neighborhood around \( \mathbf{x}_c \) — the center coefficient), 1 from the parent band, and 5 from the same spatial location in the neighboring bands at the same scale.
where \(d\) is the dimension, \(s\) is a shape parameter (scalar), \(\Sigma\) is the scale parameter (matrix) and \(\Gamma()\) is the gamma function.

\[
\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt \quad \forall z \geq 0.
\]

Methods of estimating the parameters of MVGG distributions has been explored in the past \([45–47]\). We used the moment-matching technique to estimate the parameters of an MVGG distribution as described in \([45]\). Specifically, given a set of \(N\) i.i.d. MVGG vectors of dimension \(d\), \(x_1 \ldots x_N\), compute the sample version of Mardia’s multivariate kurtosis coefficient \(\gamma_2(x) = E[(x^T\Sigma^{-1}x)^2] - d(d + 2)\) as \([49]\)

\[
\hat{\gamma}_2(x_1 \ldots x_N) = \frac{1}{N} \sum_{i=1}^{N} (x_i^T S^{-1} x_i)^2 - d(d + 2), \quad (2)
\]

where \(S\) is the sample covariance. Fortunately, \(\gamma_2(x)\) has a closed form expression in the case of \((1)\) \([45]\) and is expressed as

\[
\gamma_2(x) = \frac{d^2 \Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{d+2}{2}\right)}{\Gamma\left(\frac{d+4}{2}\right)} - d(d + 2). \quad (3)
\]

By equating \((2)\) and \((3)\) we compute an estimate of the shape parameter \(s\). Once the shape-parameter is computed, the scale parameter is estimated from the expression for the covariance \(V(x)\) \([45]\) as

\[
V(x) = \frac{2^{d/2} \Gamma\left(\frac{d+2}{2}\right)}{d \Gamma\left(\frac{d}{2}\right)} \Sigma. \quad (4)
\]

where \(V(x)\) can be replaced by the sample covariance \(S\) \([45]\).

Thus, the parameters of the MVGG distribution may be estimated using a moment-matching approach. A similar approach was used by the authors in \([46]\) in order to construct an MVGG model of wavelet coefficients from the R, G and B planes of a color image.

In order to demonstrate that the above technique estimates the parameters reliably, we drew multiple samples from MVGG distributions with random values for the covariance matrix and shape parameter using the technique described in \([45]\). Multiple such draws (1000) were performed for each \(d = 1, 5, 10, 15, 20, 25\) and we computed the squared error between the actual value of the shape parameters and the covariance matrix and their estimated value. Fig. 3 plots the mean and standard error bars of these computed errors for each \(d\). To provide a visual illustration of the sampling and fitting procedure, Fig. 2 graphs a univariate and a bivariate generalized Gaussian and an overlay of the fits obtained for these two cases. We also list the Kullback Leibler divergence (KLD) \([50]\) between the empirical histogram and the estimated fit. As may be seen from these examples, the above procedure is capable of estimating the parameters of a MVGG with high accuracy.

Having described the MVGG distribution and its parameter estimation procedure, we now describe the generalized Gaussian scale mixture model.

2.2.2. Generalized Gaussian scale mixtures

A \(d\)-dimensional random vector \(x\) is a Generalized Gaussian scale mixture (GGSM) if \(x \equiv \sqrt{z} \cdot u\), where \(z\) denotes equality in probability distribution, \(u\) is a zero-mean Multivariate Generalized Gaussian (MVGG) random vector with scale parameter \(\Sigma\), and shape parameter \(s\), and \(z\) is a scalar random variable called a mixing multiplier. The MVGG reduces to the multivariate Gaussian when \(s = 1\), hence the GGSM model subsumes the GSM model. Both the GSM and the GGSM represent infinite (scale) mixtures of Multivariate Gaussian and MVGG vectors, respectively. The conditional density of the GGSM vector \(x\) given the variance field \(z\), is given by

\[
p(x | z) = \frac{\Gamma\left(\frac{d}{2}\right) \cdot s^{-d/2}}{\pi^{d/2} \Gamma\left(\frac{d+2}{2}\right) |\Sigma_u|^{1/2}} \exp\left\{-\frac{z^{-1}}{2 s} (x^T \Sigma_u^{-1} x)^\frac{d}{2}\right\}. \quad (5)
\]
(a) Pristine image.
(b)–(h) Histograms of raw wavelet coefficients with a Gaussian fit, GSM-based divisively normalized coefficients with a Gaussian fit, and GGSM-based divisively normalized coefficients with a generalized Gaussian fit. The signals being processed are the first subbands of pristine and JPEG compressed digital photographs from the LIVE corpus. Notice in (c) and (g), the prediction of the GSM model that the normalized responses follow a Gaussian distribution is less accurate for a distorted image, while it is accurate for a pristine image, whereas in (d) and (h), the GGSM model’s prediction that the normalized coefficients follow a generalized Gaussian is accurate for both pristine and its distorted counterpart.

Fig. 5. Box plot of the estimated shape parameter values of an MVGG distribution model of the first subband of the pristine and distorted images of the LIVE database. The implicit assumption of the shape parameter value \( s = 1 \) under the GSM model is highlighted in red.

In order to form the maximum likelihood (ML) estimate of the variance field \( z \), proceed as follows:

\[
\hat{z} = \arg \max_z (\log p(x|z))
\]
\[
= \arg \max_z \left( -\frac{d}{2} \log(z) - \frac{1}{2} z^{-s}(x^T \Sigma_u^{-1} x)^s \right)
\]
\[
= \arg \min_z \left( \frac{d}{2} \log(z) + \frac{1}{2} z^{-s}(x^T \Sigma_u^{-1} x)^s \right).
\]

Setting the derivative of the objective equal to zero (necessary condition of optimality) yields the variance field estimate

\[
\hat{z} = \left( \frac{s}{d} \right)^{1/s} (x^T \Sigma_u^{-1} x)^{s/2}.
\]

Note that the normalizer is the same as that for the GSM model when \( s = 1 \). We use an iterative approach to estimate the normalizer, \( \hat{z} \), while matching the statistics of the normalized coefficients to that of the underlying MVGG distribution. Specifically, at each iteration, we use the parameters obtained by fitting a MVGG distribution to the normalized subband coefficients to compute the variance field estimates. We outline the iterative estimation procedure for the variance field of each subband as follows:

Iterative Variance Field Estimation

1: Initialize:
\( x_0, \Sigma_0 \leftarrow f(x_0) \) ➞ Detailed in Section 2.2.1
2: while \( |s_{n+1} - s_n| \geq \epsilon \) do
3: \( \hat{z}_n \leftarrow \left( \frac{\omega_n}{d} \right)^{1/s_n} (x_0^T \Sigma_u^{-1} x_0)^{s_n/2} \) ➞ From Eq. 7
4: \( x_n \leftarrow x_0 / \sqrt{\hat{z}_n} \) ➞ Divisive Normalization
5: \( s_{n+1} \leftarrow f(s_n) \) ➞ Detailed in Section 2.2.1
6: end while
7: return \( \hat{z}_n \)

In the above-mentioned variance field estimation procedure, \( s, \Sigma \leftarrow f(\cdot) \) refers to the parameter estimation of MVGG described in detail in Section 2.2.1. We observed that the above estimation procedure converges to a stable value of \( s \) for each subband.

It is interesting to note that the MVGG belongs to a class of GSM, if and only if \( s \in (0, 1] \) [51]. This proposition, when integrated with the definition of GGSM, implies that the GGSM model could be considered to be a specific case of the GSM model, but only when \( s \in (0, 1] \). However, in practice the empirical values of \( s \) encountered do not generally fall within this range of shape parameter values (as shown in Fig. 5) when...
modeling both pristine and/or distorted images. Indeed, often the estimated value of $s$ exceeds 1, even approaching $s = 2$ on some distorted images. Thus, the GGSM model provides a natural way to model pristine and distorted images over all values of the shape parameter ($s \in (0, \infty)$), and is well-suited as a model of the bandpass statistics of pristine and distorted pictures.

As with the GSM, we model a set of neighboring wavelet coefficients using the GGSM model. In our simulations, we utilize the steerable pyramid decomposition over 2 scales and 6 orientations, and the GGSM vector is formed in the same way as the GSM vector [23] — given a center coefficient $x_i$ at each subband, $x$ contains 15 coefficients, including 9 from the same subband ($3 \times 3$ neighborhood around $x_i$), 1 from the parent band, and 5 from the same spatial location but from neighboring bands at the same scale. Fig. 4 plots the marginal statistics of the subband responses of a distorted and undistorted image, before and after divisive normalization under the GSM and GGSM models. The GGSM model predicts that the normalized responses should be distributed as generalized Gaussian which is exactly what Fig. 4 demonstrates.

3. Applications of the GGSM model

Having described the GGSM model of natural, photographic images, we next consider relevant applications of the GGSM model that demonstrate its efficacy for image processing applications. Specifically, we study three applications: (1) distorted image modeling in the wavelet domain, (2) blind distortion identification, and (3) no-reference image quality assessment (NR IQA).

We shall demonstrate that the GGSM is a useful model for modeling the statistics of distorted natural images. Statistics based on the GGSM model, as mentioned earlier, have been successfully used for no-reference image quality assessment (NR IQA) [21–25] and reduced-reference (RR) IQA [19]. Yet these approaches lack a consistent and coherent model for distorted image statistics, instead relying on the GGSM model of undistorted natural images. The GGSM model provides a natural way in which one may probe image distortions.

3.1. Modeling the statistics of distorted images

We have demonstrated that the GGSM model is better suited to modeling the statistics of natural images than the GSM model. However, the GGSM model has a far more important advantage — its ability to describe the statistics of distorted images. We propose that the GGSM models not only the local statistical properties of neighboring wavelet coefficients from un-distorted natural images, but provides an improved model for statistical characterizations of distorted natural images. Modeling distorted image statistics using GGSMs lends itself naturally to image quality assessment and distortion identification [23].

As we have noted before, the GGSM model predicts that the coefficient distribution after divisive normalization will be generalized Gaussian in nature, and hence, if distorted image coefficients follow the generalized Gaussian distribution after GGSM-based normalization, then the GGSM model is indeed appropriate for modeling distorted image statistics. The same implication does not hold under the GSM model.

3.2. Distortion-identification

Researchers have observed that natural image distortions (such as compression, blur etc.) follow certain characteristic statistics which can easily be parameterized, and such parameterizations can be used to identify a distortion that is present in an image [52].

In [23], we extracted a series of statistical features from the wavelet subband and demonstrated that these features are not only sufficient to identify the distortion present in the image, but to also perform no-reference image quality assessment. In [23], the features were extracted after divisive normalization using the GSM model. While the performance of the algorithm was good, the underlying model was not appropriate for distorted image modeling. Here we modify the approach by replacing the GSM-based divisive normalization step with GGSM-based normalization.

The approach in [23] uses a training-testing procedure, where a multi-class classifier is trained using the extracted statistical features and the known distortion-class labels; the performance is evaluated on the test set. We consider three leading IQA databases — LIVE [42], CSIQ [53] and TID13 [54] with diverse distortions to evaluate the distortion classification performance. LIVE IQA database consists of 29 reference images and 779 distorted images spanning five distortion categories (JPEG2000 compression, JPEG compression, white noise, Gaussian Blur, and wireless packet loss over a fading channel), CSIQ database comprises of 866 distorted images covering 6 distortion categories (JPEG2000 compression, JPEG compression, global contrast decrements, additive white and pink Gaussian noise, and Gaussian blur), and TID13 database contains 3000 distorted images with 25 unique contents encompassing 24 diverse distortion types. Since one content from the TID13 database does not belong to a natural image category,
we test our algorithm only on 24 unique contents. The LIVE in the Wild Challenge database [55] cannot be utilized for distortion identification tests, since unlike other synthetically distorted databases, the images it contains are afflicted by complex combinations of commingled authentic distortions, and as such, they are not annotated with distortion labels.

We split each of these databases into a training set consisting of 80% of the images and a test set consisting of the remaining 20%, such there was no content overlap between the training and the test sets. The classifier was trained on the training set and the accuracy such there was no content overlap between the training and the test sets. The classifier was trained on the training set and the accuracy was independent of the training set, we repeated this 80% train–20% test split over 100 iterations and report the median classification accuracy in Table 3. In order to provide a comparison, we also list the performance of the classifier from [23], which uses the GSM model. As Table 3 indicates, the GGSM-based model outperforms the GSM model in distortion classification for all databases. This is advantageous since the GGSM model based classifier provides a justification for the divisive normalization applied on distorted images.

### 3.3. No-reference image quality assessment

We earlier proposed a framework for no-reference image quality assessment (NR-IQA) called the Distortion Identification-based Image Verity and Inegrity Evaluation (DIVINE) index [23], where statistical features were extracted in the wavelet domain after divisive normalization under the GSM model. A brief description of these features, along with their method of computation is included in Table 2. Interested readers may refer to [23] for more details.

The use of divisive normalization for image quality assessment is well motivated by models of cortical neurons in area V1 of primary visual cortex [12]. Divisive normalization accounts for the non-linear behavior of cortical neurons and provides a method to model contrast masking [12,56]. While the use of divisive normalization is well motivated, the use of GSM model is not conceptually accurate. Since we have demonstrated that the GGSM is a more appropriate model of distorted wavelet coefficients, we conducted an experiment where we replaced the GSM model in [23] with the GGSM model. We call this NR-IQA model DIVINE-GGSM and the previous one DIVINE-GSM.

Many other NR IQA algorithms leverage distortion-dependent statistical models of natural images [39,57,58]. Other NR IQA methods that are data-driven [7,59,60], also achieve effective performance. We compared the performance of DIVINE-GGSM against several leading NR IQA models, including NIQE [21], which is ‘completely blind’, BRISQUE [22], BLIINDS [39], SSEQ [61], CORNIA [7], NFERM [62], and DIVINE-GSM [23]. Our comparison also includes two FR IQA models: PSNR and SSIM [63]. Five representative IQA databases were used: [42], CSIQ [53], TID13 [54], the LIVE in the Wild Challenge database [55].

### Table 2

<table>
<thead>
<tr>
<th>Feature number</th>
<th>Feature Description</th>
<th>Computation Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–24</td>
<td>Shape and variance of subband coefficients</td>
<td>Generalized Gaussian distribution (GGD) fit to subband coefficients</td>
</tr>
<tr>
<td>25–31</td>
<td>Shape parameter across subband coefficients</td>
<td>GGD fit to stacked subband coefficients at the same orientation but at different scales</td>
</tr>
<tr>
<td>32–43</td>
<td>Correlations across scales</td>
<td>Structural correlation between windowed highpass and bandpass filter responses</td>
</tr>
<tr>
<td>44–73</td>
<td>Spatial correlation across subbands</td>
<td>Error and coefficients of the 3rd order polynomial fit to the spatial correlation function</td>
</tr>
<tr>
<td>74–88</td>
<td>Across orientation statistics</td>
<td>Windowed structural correlation between adjacent orientations at same scale</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>LIVE</th>
<th>CSIQ</th>
<th>TID13</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM-Classifier (%)</td>
<td>83.38</td>
<td>64.94</td>
<td>43.83</td>
<td>54.68</td>
</tr>
<tr>
<td>GGSM-Classifier (%)</td>
<td>86.96</td>
<td>67.34</td>
<td>44.42</td>
<td>56.19</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th></th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>Glblur</th>
<th>FF</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM (SS)</td>
<td>0.942</td>
<td>0.954</td>
<td>0.965</td>
<td>0.929</td>
<td>0.949</td>
<td>0.919</td>
</tr>
<tr>
<td>DIIVINE-GSM</td>
<td>0.903</td>
<td>0.905</td>
<td>0.982</td>
<td>0.928</td>
<td>0.867</td>
<td>0.916</td>
</tr>
<tr>
<td>DIIVINE-GGSM</td>
<td>0.901</td>
<td>0.927</td>
<td>0.982</td>
<td>0.927</td>
<td>0.900</td>
<td>0.931</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Training DB</th>
<th>LIVE SROCC</th>
<th>PLCC</th>
<th>CSIQ SROCC</th>
<th>PLCC</th>
<th>TID13 SROCC</th>
<th>PLCC</th>
<th>Challenge SROCC</th>
<th>PLCC</th>
<th>Multiply SROCC</th>
<th>PLCC</th>
<th>Overall SROCC</th>
<th>PLCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>0.892</td>
<td>0.883</td>
<td>0.803</td>
<td>0.800</td>
<td>0.652</td>
<td>0.679</td>
<td>–</td>
<td>–</td>
<td>0.803</td>
<td>0.843</td>
<td>0.747</td>
<td>0.757</td>
</tr>
<tr>
<td>SSIM (SS)</td>
<td>0.919</td>
<td>0.906</td>
<td>0.841</td>
<td>0.823</td>
<td>0.639</td>
<td>0.695</td>
<td>–</td>
<td>–</td>
<td>0.698</td>
<td>0.791</td>
<td>0.752</td>
<td>0.775</td>
</tr>
<tr>
<td>NIQE</td>
<td><strong>0.948</strong></td>
<td><strong>0.949</strong></td>
<td><strong>0.794</strong></td>
<td><strong>0.817</strong></td>
<td><strong>0.634</strong></td>
<td><strong>0.676</strong></td>
<td><strong>0.582</strong></td>
<td><strong>0.605</strong></td>
<td><strong>0.869</strong></td>
<td><strong>0.896</strong></td>
<td><strong>0.742</strong></td>
<td><strong>0.769</strong></td>
</tr>
<tr>
<td>NFERM</td>
<td>0.912</td>
<td>0.907</td>
<td>0.632</td>
<td>0.721</td>
<td>0.327</td>
<td>0.430</td>
<td>0.458</td>
<td>0.502</td>
<td>0.787</td>
<td>0.844</td>
<td>0.560</td>
<td>0.623</td>
</tr>
<tr>
<td>SSEQ</td>
<td>0.895</td>
<td>0.902</td>
<td>0.699</td>
<td>0.734</td>
<td>0.568</td>
<td>0.624</td>
<td>0.475</td>
<td>0.506</td>
<td>0.822</td>
<td>0.852</td>
<td>0.659</td>
<td>0.697</td>
</tr>
<tr>
<td>CORNIA</td>
<td><strong>0.946</strong></td>
<td><strong>0.946</strong></td>
<td><strong>0.696</strong></td>
<td><strong>0.768</strong></td>
<td><strong>0.657</strong></td>
<td><strong>0.734</strong></td>
<td><strong>0.620</strong></td>
<td><strong>0.657</strong></td>
<td><strong>0.906</strong></td>
<td><strong>0.913</strong></td>
<td><strong>0.747</strong></td>
<td><strong>0.791</strong></td>
</tr>
<tr>
<td>BLINDS-II</td>
<td>0.931</td>
<td>0.952</td>
<td>0.675</td>
<td>0.692</td>
<td>0.598</td>
<td>0.667</td>
<td>0.499</td>
<td>0.529</td>
<td>0.848</td>
<td>0.866</td>
<td>0.689</td>
<td>0.726</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>0.937</td>
<td>0.938</td>
<td>0.695</td>
<td>0.698</td>
<td>0.524</td>
<td>0.549</td>
<td>0.608</td>
<td>0.637</td>
<td>0.893</td>
<td>0.915</td>
<td>0.691</td>
<td>0.709</td>
</tr>
<tr>
<td>DIVINE-GSM</td>
<td>0.916</td>
<td>0.913</td>
<td>0.733</td>
<td>0.756</td>
<td>0.655</td>
<td>0.700</td>
<td>0.600</td>
<td>0.623</td>
<td>0.855</td>
<td>0.877</td>
<td>0.727</td>
<td>0.753</td>
</tr>
<tr>
<td>DIVINE-GGSM</td>
<td>0.931</td>
<td>0.930</td>
<td>0.762</td>
<td>0.795</td>
<td><strong>0.664</strong></td>
<td><strong>0.702</strong></td>
<td><strong>0.600</strong></td>
<td><strong>0.610</strong></td>
<td><strong>0.870</strong></td>
<td><strong>0.873</strong></td>
<td><strong>0.743</strong></td>
<td><strong>0.770</strong></td>
</tr>
</tbody>
</table>
Table 6

<table>
<thead>
<tr>
<th>Training DB</th>
<th>LIVE</th>
<th>CSIQ</th>
<th>TID13</th>
<th>TID13</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIVINE-GSM</td>
<td>0.860</td>
<td>0.864</td>
<td>0.641</td>
<td>0.694</td>
</tr>
<tr>
<td>DIVINE-GGSM</td>
<td>0.871</td>
<td>0.872</td>
<td>0.859</td>
<td>0.791</td>
</tr>
</tbody>
</table>

Table 4 was computed by transforming each correlation coefficient using (RBF) kernel, whose parameters were estimated using cross-validation iterations. Support Vector Regressor [65] with radial basis function (RBF) kernel, whose parameters were estimated using cross-validation on the training set, was utilized for this regression task. The overall average correlation coefficient across the different databases listed in Table 4 was computed by transforming each correlation coefficient using Fisher’s z-transformation [66];

\[
z = \frac{1}{2} \ln \frac{1 + r}{1 - r}, \quad \text{where } r \text{ is SROCC or PLCC.} \tag{8}
\]

then calculating the mean of the z values, and finally back-transforming by inverting (8), yielding the overall correlation coefficient.

It is evident from Table 4 and Table 5 that DIVINE-GGSM performed better than its predecessor DIVINE-GSM, and competes quite well against the compared IQA algorithms on most of the databases used for comparison. An important exception was the CSIQ database, which contains global contrast changes as one of the distortions on which most of the NR IQA metrics fail, since NS-based IQA models which employ ‘divisive normalization’ are less sensitive to changes in image contrast, while they are more effective in capturing statistical changes due to structural distortions.

We also compared the generalization capability of DIVINE-GGSM against DIVINE-GSM by using an entire database for training and then testing on common distortions from a test database. The distortions used for testing were: JPEG2000 compression, JPEG, Additive white noise (WN) and Gaussian Blur (blur). As in the classification case, DIVINE-GGSM showed an overall improved performance over DIVINE-GSM, as indicated in Table 6. The philosophy behind the DIVINE-GSM approach is to model the distorted images using a model for the pristine images — the GSM. The GGSM on the other hand, more appropriately models natural images that are either pristine or distorted in a larger model-space, yielding better model fits and a reasonable improvement in performance.

4. Conclusion and future work

We have developed a new statistical model for the image wavelet coefficients by generalizing the GSM model for natural images. The GGSM model is suitable for the wavelet coefficients of both natural and distorted images. We also showed that the GGSM distribution better models the wavelet coefficients of distorted images as compared to the GSM distribution. Further, we demonstrated applications of the GGSM model in distortion identification, and NR-IQA.

There are a number of directions for possible future work. Recently, an interesting strategy was designed to evaluate the performance of NR IQA algorithms, which utilizes more than 100,000 distorted images to find a mapping from a feature space to quality scores [67,68]. An important ingredient of this strategy is the use of FR IQA algorithm scores as a proxy for human opinion scores. The authors argue that this technique can be used to more reliably verify the effectiveness of NR IQA features, as compared to approaches that involve a certain number of train–test splits of the dataset, which is more prone to overfitting.

While the estimation procedure for the GGSM model parameters uses a moment matching approach, an important extension would be to investigate maximum likelihood parameter estimation of the GGSM model parameters. Although we only sampled a few applications to demonstrate the utility of the GGSM model, there are a number of other applications, including image restoration, full reference and reduced reference quality assessment, compression, retrieval, and so on, where the model could be useful.

References


