Signal Processing: Image Communication 66 (2018) 87–94

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

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

ARTICLE INFO

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

ABSTRACT

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

* Corresponding author.
E-mail addresses: praful_gupta@utexas.edu (P. Gupta), amoorthy@netflix.com (A.K. Moorthy), rajivs@ece.iisc.ernet.in (R. Soundararajan), bovik@ece.utexas.edu (A.C. Bovik).
[1] Fellow, IEEE.

https://doi.org/10.1016/j.image.2018.05.009
Received 15 October 2017; Received in revised form 12 May 2018; Accepted 12 May 2018
Available online 16 May 2018

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.

	30°	60°	90°	120°	150°	180°
Scale 1	13.79%	51.72%	58.62%	0%	44.83%	44.83%
Scale 2	3.45%	34.48%	34.48%	0%	41.38%	34.48%
Scale 3	3.45%	27.59%	44.83%	10.34%	34.48%	31.03%

problem [31,32], and to devise a simple but effective way to compute monocular 3D depths [33,34].

While the GSM model is robust, it is only an approximate model of un-distorted natural bandpass images. Further, the GSM model does not generalize to *distorted* natural images, and hence does not lend itself naturally to problems in which one requires a model for distorted images. Here we propose a generalization of the GSM model that addresses these two drawbacks. The generalized Gaussian scale mixture model (GGSM) which is described in the remainder of this article, models the statistics of neighboring bandpass or image wavelet coefficients as obeying a scale mixture of multivariate generalized Gaussian distributions. The GGSM accurately models the statistics of natural un-distorted images and is also capable of describing the scene statistics of distorted natural images. The GGSM is a generalization of the GSM, and therefore subsumes the GSM, and models a larger space of natural images.

In this article, we first discuss the limitations of the GSM model and then describe the GGSM model. The ability of the GGSM to model bandpass natural images is evaluated, and properties similar to GSM modeling are studied (Section 2). Section 3 describes three applications of the GGSM model: modeling the statistics of distorted images, distortion identification and no-reference quality assessment. In Section 4, we describe possible future extensions of the GGSM model for perceptual image processing.

2. Modeling wavelet coefficients

2.1. Gaussian scale mixtures

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

$$p_{\mathbf{x}}(\mathbf{x}) = \int \frac{1}{(2\pi)^{d/2}|z\boldsymbol{\Sigma}_{\mathbf{u}}|^{1/2}} \exp\left(\frac{-\mathbf{x}^T\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{x}}{2z}\right) p_z(z)dz,$$

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 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 $\hat{z} = \mathbf{x}^T\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{x}/d$ [18,44]. Once this multiplier is estimated, one can perform a 'divisive normalization transform', whereby each center coefficient, 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.

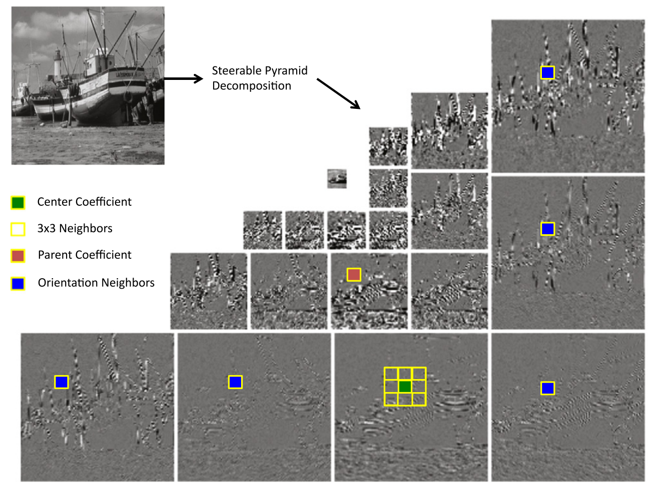


Fig. 1. The neighborhood structure used to construct GSM/GGSM vectors in the experiments. The GSM/GGSM vector Y contains 15 coefficients, including 9 from the same subband (3×3 neighborhood around x_c — the center coefficient), 1 from the parent band, and 5 from the same spatial location in the neighboring bands at the same scale.

The 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 statistics 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

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma\left(\frac{d}{2}\right)\cdot s}{\pi^{d/2}\Gamma\left(\frac{d}{2s}\right)2^{d/2s}|\boldsymbol{\Sigma}|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x})^s\right\}, \tag{1}$$

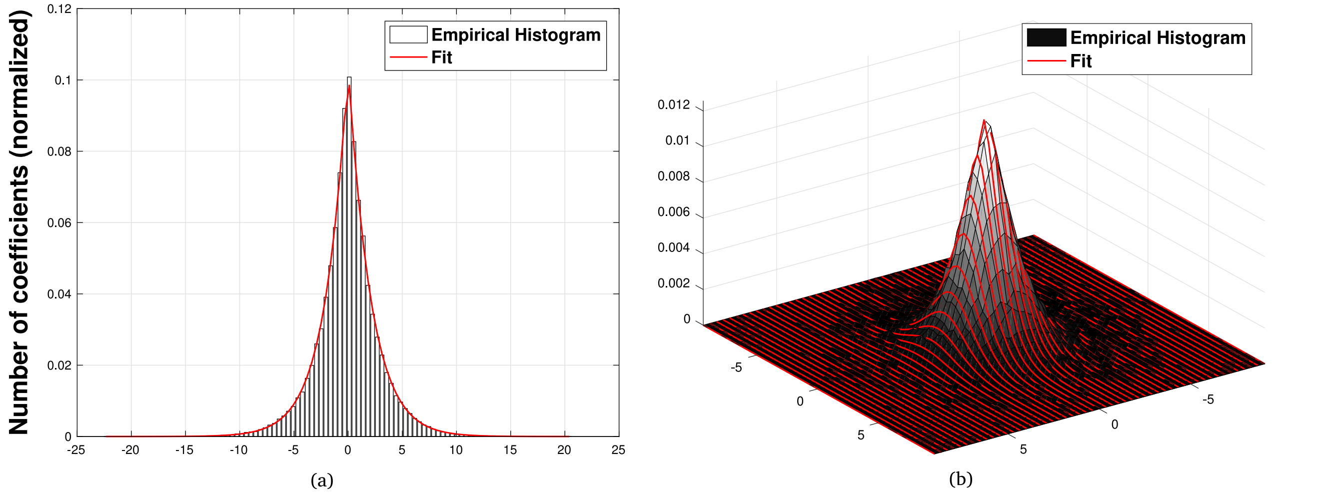


Fig. 2. Visual illustration of the parameter estimation procedure. est = estimated parameters, $\mathcal{I}$ = identity matrix. (a) $d = 1$, $s = 0.5$, $\Sigma = 1$; $s_{est} = 0.52$, $\Sigma_{est} = 1.1744$; KLD = 0.0015 and (b) $d = 2$, $s = 0.7$, $\Sigma = \mathcal{I}$; $s_{est} = 0.7$, $\Sigma_{est} = [1.0044\ 0.0005;\ 0.0005\ 0.9905]$; KLD = 0.0422.

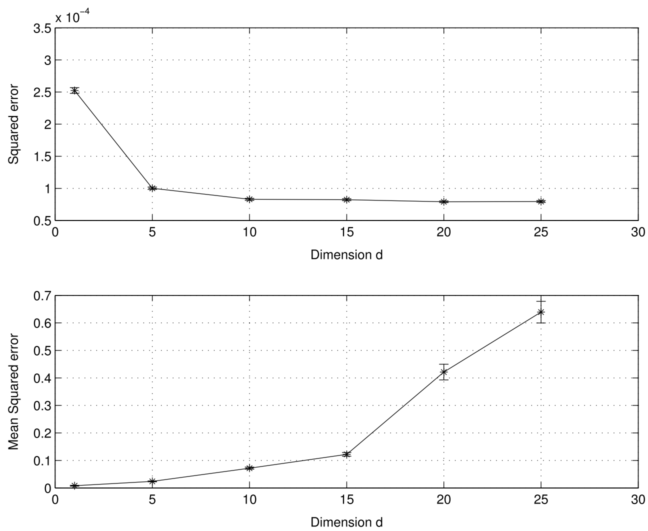


Fig. 3. The mean and standard error bars of squared errors between actual and estimated shape parameter (top) and the covariance matrix (bottom) across 1000 sample draws of a MVGG, as a function of the dimension d.

where d is the dimension, s is a shape parameter (scalar), $\boldsymbol{\Sigma}$ is the scale parameter (matrix) and $\Gamma(\cdot)$ 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, $\mathbf{x}_1 \dots \mathbf{x}_N$, compute the sample version of Mardia's multivariate kurtosis coefficient – $\gamma_2(\mathbf{x}) = E[(\mathbf{x}^T \boldsymbol{\Sigma}^{-1}\mathbf{x})^2] - d(d+2)$ –as [49]

$$\hat{\gamma}_2(\mathbf{x}_1 \dots \mathbf{x}_N) = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i^T \mathbf{S}^{-1}\mathbf{x}_i)^2 - d(d+2), \tag{2}$$

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

$$\gamma_2(\mathbf{x}) = \frac{d^2\Gamma(\frac{d}{2s})\Gamma(\frac{d+4}{2s})}{\Gamma^2(\frac{d+2}{2s})} - d(d+2). \tag{3}$$

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 $\mathbf{V}(\mathbf{x})$ [45] as

$$\mathbf{V}(\mathbf{x}) = \frac{2^{1/s}\Gamma(\frac{d+2}{2s})}{d\,\Gamma(\frac{d}{2s})}\boldsymbol{\Sigma}, \tag{4}$$

where $\mathbf{V}(\mathbf{x})$ can be replaced by the sample covariance $\mathbf{S}$ [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.

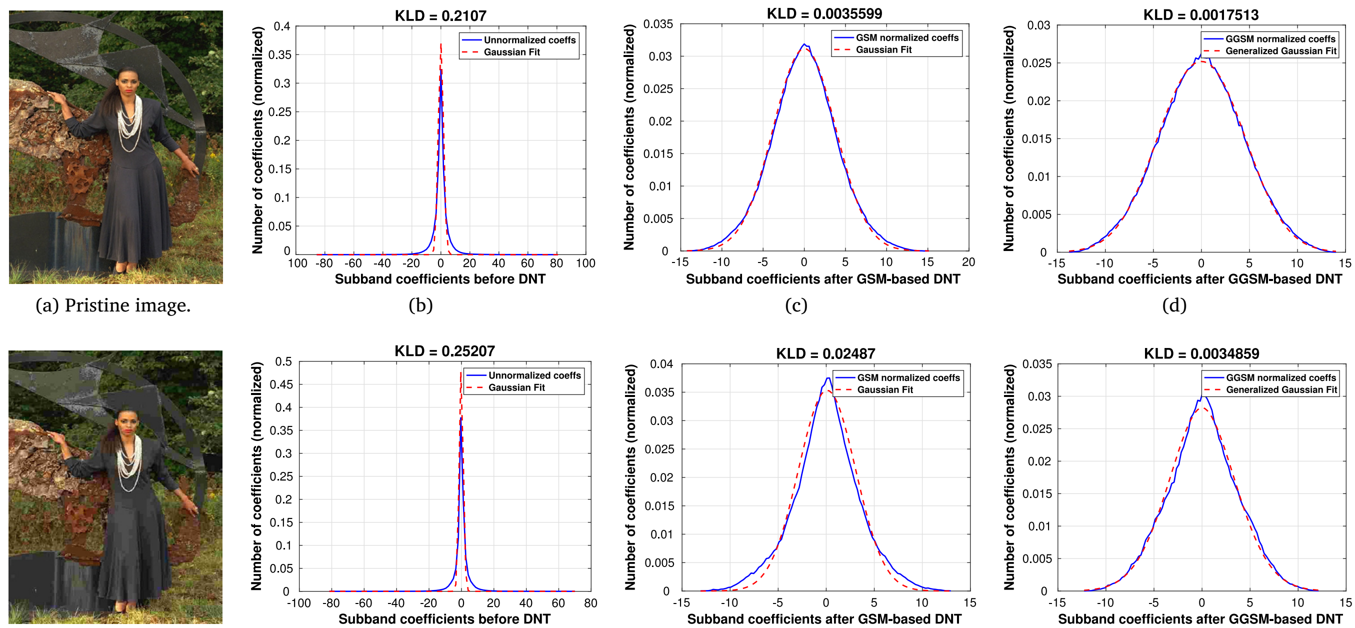
In order to demonstrate that the above technique estimates the parameters reliably, we drew multiple samples from MVGG distributions with random values for the covariance matrix and shape parameter using the technique described in [45]. Multiple such draws (1000) were performed for each $d = 1, 5, 10, 15, 20, 25$ 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 $\mathbf{x}$ is a Generalized Gaussian scale mixture (GGSM) if $\mathbf{x} \equiv \sqrt{z}\cdot\mathbf{u}$, where $\equiv$ denotes equality in probability distribution, $\mathbf{u}$ is a zero-mean Multivariate Generalized Gaussian (MVGG) random vector with scale parameter $\boldsymbol{\Sigma}_\mathbf{u}$ 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 $\mathbf{x}$ given the variance field z, is given by

$$p(\mathbf{x}|z) = \frac{\Gamma\left(\frac{d}{2}\right)\cdot s\cdot z^{-d/2}}{\pi^{d/2}\Gamma\left(\frac{d}{2s}\right)2^{d/2s}|\boldsymbol{\Sigma}_\mathbf{u}|^{1/2}}\cdot\exp\left\{-\frac{z^{-s}}{2}(\mathbf{x}^T\boldsymbol{\Sigma}_\mathbf{u}^{-1}\mathbf{x})^s\right\}. \tag{5}$$



(a) Pristine image. (b) (c) (d)

(e) JPEG compressed image. (f) (g) (h)

Fig. 4. (a), (e) A pristine and JPEG compressed image from the LIVE database; (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.

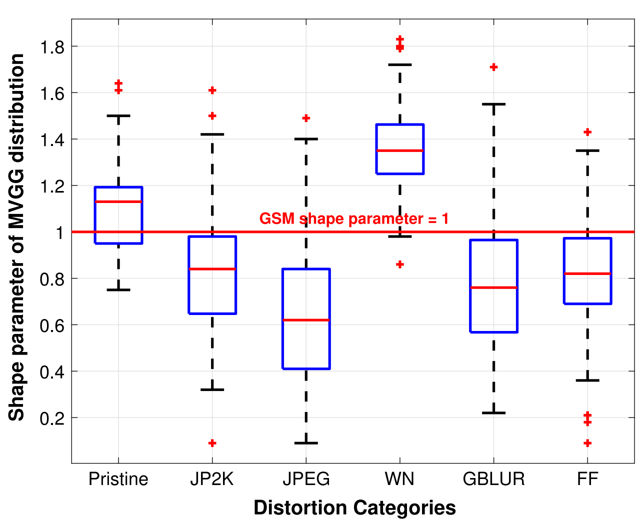


Fig. 5. Box plot of the estimated shape parameter values of an MVGG distribution model of the first subband of the pristine and distorted images of the LIVE database. The implicit assumption of the shape parameter value ($s = 1$) under the GSM model is highlighted in red.

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

$$\begin{aligned}\hat{z} &= \underset{z}{\operatorname{argmax}}(\log p(\mathbf{x}|z)) \\ &= \underset{z}{\operatorname{argmax}}\left(-\frac{d}{2}\log(z) - \frac{1}{2}z^{-s}(\mathbf{x}^T\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{x})^s\right) \\ &= \underset{z}{\operatorname{argmin}}\left(\frac{d}{2}\log(z) + \frac{1}{2}z^{-s}(\mathbf{x}^T\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{x})^s\right).\end{aligned} \tag{6}$$

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

$$\hat{z} = \left(\frac{s}{d}\right)^{1/s}(\mathbf{x}^T\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{x}). \tag{7}$$

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

Initialize:
  $\mathbf{x}_0$: original subband coefficients
  $s_0, \boldsymbol{\Sigma}_0 \leftarrow f(\mathbf{x}_0)$ ▷ Detailed in Section 2.2.1
while $|s_{n+1} - s_n| \geq \epsilon$ **do**
  $\hat{z}_n \leftarrow \left(\frac{s_n}{d}\right)^{1/s_n}(\mathbf{x}_0^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_0)$ ▷ From Eq. 7
  $\mathbf{x}_n \leftarrow \mathbf{x}_0/\sqrt{\hat{z}_n}$ ▷ Divisive Normalization
  $s_{n+1}, \boldsymbol{\Sigma}_{n+1} \leftarrow f(\mathbf{x}_n)$ ▷ Detailed in Section 2.2.1
end while
return $\hat{z}_n$

In the above-mentioned variance field estimation procedure, $s, \boldsymbol{\Sigma} \leftarrow f(\cdot)$ 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

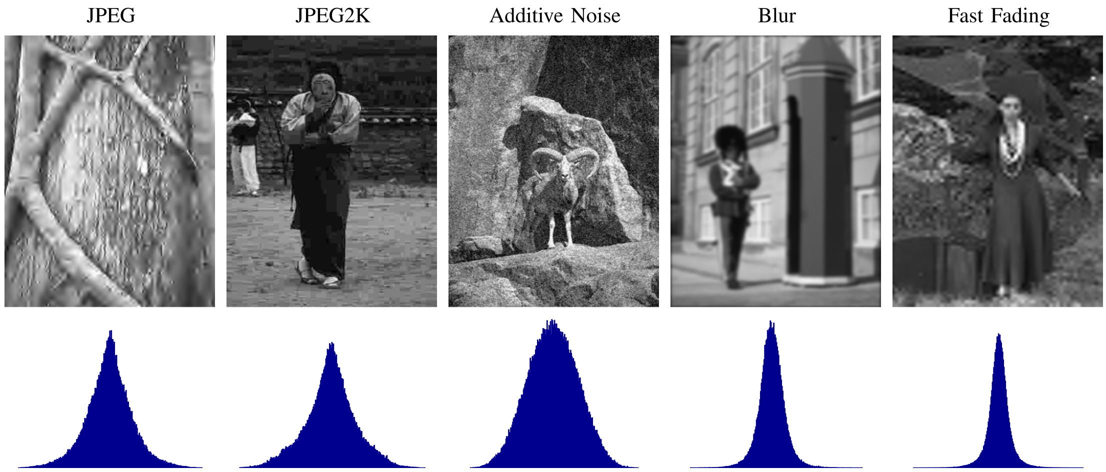


Fig. 6. Distorted images (first row) and histograms of normalized coefficients (second row) from one subband. Note that the shape of the distribution is a function of the distortion.

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_c at each subband, $\mathbf{x}$ contains 15 coefficients, including 9 from the same subband (3×3 neighborhood around x_c), 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 GSM 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 GSM 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. Fig. 6 shows five images corrupted by different distortions and their coefficient histograms from one subband of the steerable pyramid decomposition after GGSM-based divisive normalization. One may observe that the divisively-normalized coefficients of the distorted images follow a generalized Gaussian distribution.

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,

Table 2
Brief summary of the DIIVINE features and their computation procedure. For detailed description refer [23].

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

Table 3
Median classification accuracy across 100 train–test trials on LIVE, CSIQ, and TID13 image databases. 'Avg' column includes the average accuracies weighted by the number of images from each database.

	LIVE	CSIQ	TID13	Avg
GSM-Classifier (%)	*83.38*	*64.94*	*43.83*	*54.68*
GGSM-Classifier (%)	***86.96***	***67.54***	***44.42***	***56.19***

Table 5
Median Spearman's rank ordered correlation coefficient (SROCC) across 100 train–test trials on all distortion categories of LIVE IQA database. *Italicized* algorithms are NR IQA algorithms, others are FR IQA algorithms.

	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.920	0.915	0.984	0.835	0.905	0.897
SSIM (SS)	0.942	0.954	0.965	0.929	0.949	0.919
DIIVINE-GSM	*0.903*	*0.905*	*0.982*	*0.928*	*0.867*	*0.916*
DIIVINE-GGSM	*0.901*	*0.927*	*0.982*	*0.927*	*0.900*	***0.931***

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 of the classifier is tested on the test set. To ensure that performance 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 INtegrity Evaluation (DIIVINE) 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 DIIVINE-GGSM and the previous one DIIVINE-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 DIIVINE-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 DIIVINE-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

Table 4
Median Spearman's Rank Ordered Correlation Coefficient (SROCC) and Pearson's Linear Correlation Coefficient (PLCC) across 100 train–test trials on LIVE, CSIQ, TID13, LIVE Challenge, and LIVE Multiply Distorted IQA databases. *Italicized* algorithms are NR IQA algorithms, others are FR IQA algorithms. Best two NR-IQA algorithms for each database are boldfaced.

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

Table 6
Median SROCC when trained and tested on different databases.

Training DB	LIVE		CSIQ		TID13	
Testing DB	CSIQ	TID13	LIVE	TID13	CSIQ	LIVE
DIIVINE-GSM	*0.860*	*0.864*	*0.641*	*0.694*	***0.741***	*0.759*
DIIVINE-GGSM	***0.871***	***0.872***	***0.859***	***0.791***	*0.723*	***0.821***

database [55] and the LIVE Multiply Distorted database [64]. Table 4 compares the performance in terms of Spearman's rank order correlation coefficient (SROCC), which measures the monotonicity between the predicted objective scores and the subjective opinion scores, and Pearson's Linear Correlation Coefficient (PLCC), which captures the degree of linear relationship between the two. Similar training strategy was used as in distortion classification to report the SROCC and PLCC results — median of 80% train–20% test splits with no content overlap over 100 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]:,

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

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 DIIVINE-GGSM performed better than its predecessor DIIVINE-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 NSS-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 DIIVINE-GGSM against DIIVINE-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, DIIVINE-GGSM showed an overall improved performance over DIIVINE-GSM, as indicated in Table 6. The philosophy behind the DIIVINE-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

[1] H.B. Barlow, Possible principles underlying the transformation of sensory messages, Sensory Commun. (1961) 217–234.
[2] E.P. Simoncelli, B.A. Olshausen, Natural image statistics and neural representation, Ann. Rev. Neurosci. 24 (1) (2001) 1193–1216.
[3] A. Srivastava, A.B. Lee, E.P. Simoncelli, S.C. Zhu, On advances in statistical modeling of natural images, J. Math. Imaging Vis. 18 (1) (2003) 17–33.
[4] E.P. Simoncelli, B.A. Olshausen, Natural image statistics and neural representation, Annu. Rev. Neurosci. 24 (1) (2001) 1193–1216.
[5] B.A. Olshausen, D.J. Field, Natural image statistics and efficient coding, Network 7 (1996) 333–339.
[6] H.-w. Chang, M.-h. Wang, Sparse correlation coefficient for objective image quality assessment, Signal Process., Image Commun. 26 (10) (2011) 577–588.
[7] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Comp. Vision and Pattern Reco., 2012, IEEE, 2012, pp. 1098–1105.
[8] D. Temel, M. Prabhushankar, G. AlRegib, UNIQUE: Unsupervised image quality estimation, IEEE Signal Process. Lett. 23 (10) (2016) 1414–1418.
[9] Z. Gu, L. Zhang, X. Liu, H. Li, J. Lu, Learning quality-aware filters for no-reference image quality assessment, in: Intl. Conf. on Multimedia and Expo, ICME, 2014, pp. 1–6.
[10] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on hvs, in: Proc. of the Second Intl. Workshop on Video Proc. and Quality Metrics, vol. 4, 2006.
[11] M. Wainwright, E. Simoncelli, Scale mixtures of Gaussians and the statistics of natural images, Adv. Neural Inf. Process. Syst. 12 (1) (2000) 855–861.
[12] M.J. Wainwright, O. Schwartz, E.P. Simoncelli, Natural image statistics and divisive normalization: Modeling nonlinearities and adaptation in cortical neurons, Stat. Theor. Brain (2002) 203–222.
[13] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Tran. Image Process. 12 (11) (2003) 1338–1351.
[14] J. Portilla, E.P. Simoncelli, Image restoration using gaussian scale mixtures in the wavelet domain, in: Int'L Conf. Image Process., 2003, vol. 2, IEEE, 2003, pp. II–965.
[15] W. Dong, G. Shi, Y. Ma, X. Li, Image restoration via simultaneous sparse coding: Where structured sparsity meets gaussian scale mixture, Int. J. Comput. Vis. 114 (2–3) (2015) 217–232.
[16] H.R. Sheikh, A.C. Bovik, G. De Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Trans. Image Process. 14 (12) (2005) 2117–2128.
[17] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Tran. Image Process. 20 (8) (2011) 2378–2386.
[18] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Top. Signal Process. 3 (2) (2009) 202–211.
[19] R. Soundararajan, A.C. Bovik, RRED indices: Reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2) (2012) 517–526.
[20] C.G. Bampis, P. Gupta, R. Soundararajan, A.C. Bovik, SpEED-QA: Spatial efficient entropic differencing for image and video quality, IEEE Signal Process. Lett. 24 (9) (2017) 1333–1337.
[21] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
[22] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Tran. Image Process. 21 (12) (2012) 4695–4708.
[23] 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.
[24] Y. Zhang, A.K. Moorthy, D.M. Chandler, A.C. Bovik, C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes, Signal Process., Image Commun. 29 (7) (2014) 725–747.
[25] D. Ghadiyaram, A.C. Bovik, Feature maps driven no-reference image quality prediction of authentically distorted images, in: SPIE Human Vision and Elect. Imag., 2015 pp. 93940J–93940J.
[26] M.A. Saad, A.C. Bovik, C. Charrier, Blind prediction of natural video quality, IEEE Tran. Image Process. 23 (3) (2014) 1352–1365.
[27] A.K. Moorthy, A.C. Bovik, Visual quality assessment algorithms: what does the future hold?, Multimedia Tools Appl. 51 (2) (2011) 675–696.

[28] R. Soundararajan, A.C. Bovik, Video quality assessment by reduced reference spatio-temporal entropic differencing, IEEE Trans. Circuits Syst. Video Technol. 23 (4) (2013) 684–694.
[29] A. Mittal, M.A. Saad, A.C. Bovik, A completely blind video integrity oracle, IEEE Trans. Image Process. 25 (1) (2016) 289–300.
[30] A.C. Bovik, Automatic prediction of perceptual image and video quality, Proc. IEEE 101 (9) (2013) 2008–2024.
[31] M.-J. Chen, L.K. Cormack, A.C. Bovik, No-reference quality assessment of natural stereopairs, IEEE Tran. Image Process. 22 (9) (2013) 3379–3391.
[32] M.-J. Chen, C.-C. Su, D.-K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process., Image Commun. 28 (9) (2013) 1143–1155.
[33] C.-C. Su, L.K. Cormack, A.C. Bovik, Depth estimation from monocular color images using natural scene statistics models, in: IVMSP Workshop, 2013 IEEE 11th, IEEE, 2013, pp. 1–4.
[34] C.-C. Su, L.K. Cormack, A.C. Bovik, Bayesian depth estimation from monocular natural images, J. Vis. 17 (5) (2017) 22–22.
[35] P. Gupta, C.G. Bampis, A.C. Bovik, Natural scene statistics for noise estimation, in: IEEE Southwest Symp. on Image Anal. and Interp., SSIAI, Las Vegas, NV, 2018.
[36] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[37] H.R. Sheikh, A.C. Bovik, G. De Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Tran. Image Process. 14 (12) (2005) 2117–2128.
[38] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E. Yang, A.C. Bovik, Quality-aware images, IEEE Trans. Image Process. 15 (6) (2006) 1680–1689.
[39] M. Saad, A. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[40] D. Temel, G. AlRegib, PerSIM: Multi-resolution image quality assessment in the perceptually uniform color domain, in: IEEE Int'l Conf. Image Process, IEEE, 2015, pp. 1682–1686.
[41] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, On between-coefficient contrast masking of DCT basis functions, in: Proc. of the Third Int'L. Workshop on Video Proc. and Quality Metrics, vol. 4, 2007.
[42] 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.
[43] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multiscale transforms, IEEE Trans. Inform. Theory 38 (2) (1992) 587–607.
[44] M.J. Wainwright, E.P. Simoncelli, Scale mixtures of gaussians and the statistics of natural images, in: Nips, 1999, pp. 855–861.
[45] E. Gómez, M. Gomez-Viilegas, J. Marin, A multivariate generalization of the power exponential family of distributions, Comm. Statist. Theory Methods 27 (3) (1998) 589–600.
[46] R. Kwitt, P. Meerwald, A. Uhl, Color-image watermarking using multivariate power-exponential distribution, in: IEEE Int'l Conf. Image Process., 2009, pp. 4245–4248.
[47] M. Coban, R. Mersereau, Adaptive subband video coding using bivariate generalized Gaussian distribution model, in: IEEE Intl. Conf. Img. Proc., 2002, vol. 4, IEEE, 2002, pp. 1990–1993.
[48] G. Verdoolaege, P. Scheunders, Geodesics on the manifold of multivariate generalized gaussian distributions with an application to multicomponent texture discrimination, Int. J. Comput. Vis. 95 (3) (2011) 265.
[49] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhyā 36 (2) (1974) 115–128.
[50] T. Cover, J. Thomas, Elements of Information Theory, Wiley-Interscience, 2006.
[51] E. Gómez-Sánchez-Manzano, M. Gómez-Villegas, J. Marín, Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications, Comm. Statist. Theory Methods 37 (6) (2008) 972–985.
[52] A.K. Moorthy, A.C. Bovik, Statistics of natural image distortions, in: IEEE Int'l Conf. on Acoustics Speech and Signal Process., 2010, pp. 962–965, 2010.
[53] 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) 011006–011006.
[54] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Image database TID2013: Peculiarities, results and perspectives, Signal Process., Image Commun. 30 (2015) 57–77.
[55] D. Ghadiyaram, A.C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (1) (2016) 372–387.
[56] R. Sekuler, R. Blake, Perception, McGraw Hill, 2002.
[57] M. Jenadeleh, M.E. Moghaddam, BIQWS: Efficient wakeby modeling of natural scene statistics for blind image quality assessment, Multimedia Tools Appl. 76 (12) (2017) 13859–13880.
[58] W. Hachicha, M. Kaaniche, A. Beghdadi, F.A. Cheikh, No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models, Signal Process., Image Commun. 54 (2017) 107–117.
[59] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, A.C. Bovik, Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment, IEEE Signal Process. Mag. 34 (6) (2017) 130–141.
[60] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2018) 206–219.
[61] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Signal Process., Image Commun. 29 (8) (2014) 856–863.
[62] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimedia 17 (1) (2015) 50–63.
[63] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error measurement to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[64] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: Signals, Systems and Computers (ASILOMAR), 2012, pp. 1693–1697.
[65] B. Schölkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (5) (2000) 1207–1245.
[66] D.M. Corey, W.P. Dunlap, M.J. Burke, Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations, J. Gen. Psychol. 125 (3) (1998) 245–261.
[67] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, A.C. Bovik, No-reference quality assessment of screen content pictures, IEEE Tran. Image Process. 26 (8) (2017) 4005–4018.
[68] K. Gu, D. Tao, J.-F. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, IEEE Trans. Neural Netw. Learn. Syst. (2017).