

# Statistically and Perceptually Motivated Nonlinear Image Representation

Siwei Lyu and Eero P. Simoncelli

Howard Hughes Medical Institute,  
Center for Neural Science, and  
Courant Institute of Mathematical Sciences  
New York University,  
New York, NY 10003, USA

## ABSTRACT

We describe an invertible nonlinear image transformation that is well-matched to the statistical properties of photographic images, as well as the perceptual sensitivity of the human visual system. Images are first decomposed using a multi-scale oriented linear transformation. In this domain, we develop a Markov random field model based on the dependencies within local clusters of transform coefficients associated with basis functions at nearby positions, orientations and scales. In this model, division of each coefficient by a particular linear combination of the amplitudes of others in the cluster produces a new nonlinear representation with marginally Gaussian statistics. We develop a reliable and efficient iterative procedure for inverting the divisive transformation. Finally, we probe the statistical and perceptual advantages of this image representation, examining robustness to added noise, rate-distortion behavior, and artifact-free local contrast enhancement.

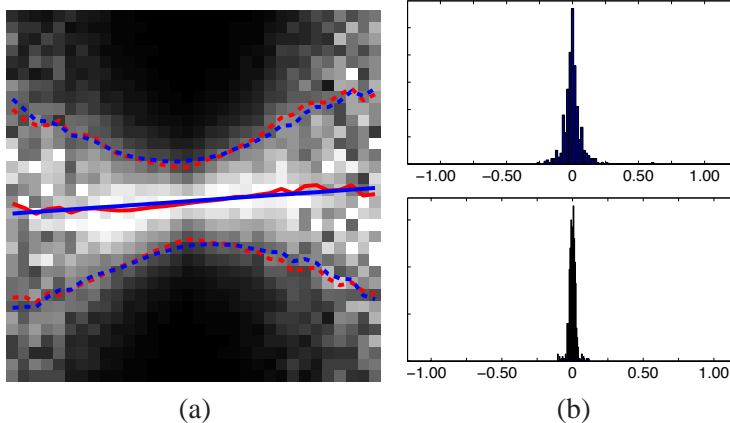
**Keywords:** optimal representation, independent components, Markov random field, divisive normalization, contrast enhancement

## 1. INTRODUCTION

Image analysis often relies on the choice of representation. Most of the literature is based on linear representations, which are well understood, and which can be optimized to minimize statistical dependencies (e.g., using principal components analysis, or more recently, independent components analysis<sup>1,2</sup>). The basis functions obtained using these methods are localized oriented filters at different spatial scales, which we can loosely refer to as “wavelets”. Such representations have been used as an effective substrate for a wide variety of applications in machine vision and image processing. Nevertheless, linear decompositions do not (and probably cannot<sup>3</sup>) completely eliminate higher-order statistical dependencies. It is thus natural to ask whether we might develop better representations by incorporating invertible nonlinear transformations<sup>4</sup>

In recent years, a local gain control model, sometimes known as *divisive normalization* has become popular for modeling biological vision. Each coefficient of a linear transform is divided by a Minkowski combination of a cluster of neighboring coefficient amplitudes. This type of model can explain nonlinearities in the responses of mammalian cortical neurons,<sup>5,6</sup> nonlinear masking phenomenon in human visual perception,<sup>7-9</sup> and has also been empirically shown to produce approximately Gaussian marginal distributions<sup>10</sup> and to reduce statistical dependencies of the original linear representation.<sup>11-15</sup> Recently, several authors have developed invertible image transformations that incorporate divisive normalization.<sup>16-18</sup>

In this paper, we derive a new form of divisive representation from a multi-scale Markov random field model for images. We first develop the model, demonstrating that it can account for basic statistical properties of photographic images, and describe efficient methods for estimating the model parameters. We show that divisive normalization arises naturally as a means of simplifying the statistical properties of the linear coefficients, and



**Fig. 1.** (a) Conditional histogram of two adjacent wavelet coefficients. Grayscale intensities are proportional to probability, with larger values corresponding to brighter pixels. Each column is normalized to fill the full range of intensities. Red solid and dashed lines indicate conditional mean and standard deviation, respectively. Blue solid and dashed lines are best-fitting linear models for the conditional mean and standard deviation. (b) Histograms of relative fitting errors over 4 different images and 20 subbands of the conditional mean (top) and conditional variance (bottom) of the same pair of coefficients.

we develop a reliable and efficient iterative procedure for inverting this divisive normalization transformation. Last, we demonstrate the promise of this representation in terms of perceptual resilience to noise contamination, image compression, and adaptive enhancement of local contrast.

## 2. STATISTICAL IMAGE MODEL

Photographic images exhibit highly non-Gaussian statistical behaviors. When decomposed in a wavelet decomposition, the coefficients tend to have highly kurtotic marginal distributions.<sup>19,20</sup> More importantly, there are higher-order statistical dependencies between coefficients at nearby locations, orientations and scales.<sup>11,21</sup> Such higher-order dependencies can be illustrated by examining the conditional distribution of adjacent pairs of wavelet coefficients.

Consider two coefficients  $(x_1, x_2)$  representing images at adjacent spatial locations, within the same subband of a steerable pyramid,<sup>22</sup> a type of overcomplete multiscale oriented transform. Assuming spatial homogeneity, we examine the conditional histogram of this pair of coefficients, gathered over the whole subband. Fig. 1 (a) shows this conditional histogram, which illustrates several important aspects of the relationship between the two coefficients. First, both the conditional mean *and* variance of  $x_1$  exhibit strong dependency on the value of  $x_2$ . The conditional mean can be approximated with a linear function of  $x_2$  (shown as a solid blue line). The conditional variance can also be fit with a linear function of the mean-corrected square of  $x_2$  (shown as a blue dashed line).<sup>11,12,23</sup> This linear description provides a surprisingly good approximation across different images and subbands. Shown in Fig. 1 (b) are histograms of relative fitting errors, computed as the ratio of the difference between data (i.e., the conditional means and variances) and their fittings and the standard deviation of the data, over 4 different images and 20 subbands of the conditional mean and variance of the same pair of coefficients.

A variety of models have been proposed to capture these local statistical behaviors. Several authors have used a generalized Gaussian to model subband marginal distributions.<sup>24–27</sup> Local joint statistics have been described using spherically invariant random processes,<sup>28</sup> including: binary Gaussian mixtures,<sup>29,30</sup> infinite Gaussian scale mixtures,<sup>23,31,32</sup> elliptical generalized Gaussians,<sup>33</sup> and Bessel K-forms.<sup>34</sup> Local models have been extended to global models by invoking a Markov structure on a tree,<sup>29,30,35</sup> or a spatial Markov random field.<sup>36–40</sup>

## 2.1. Multi-scale Conditionally Gaussian Markov Image Model

Here, we attempt to develop the simplest global model that captures the conditional mean and variance relationships described above. We first assume that the linearity of the pairwise dependency of mean and variance extends to dependency on a generalized neighborhood of coefficients at nearby locations, orientations and scales. Denoting the indices of coefficients within the generalized neighborhood of coefficient  $x_i$  as  $\mathcal{N}(i)$ , we define:

$$\mathcal{E}(x_i|x_j, j \in \mathcal{N}(i)) = \mu_i \equiv \sum_{j \in \mathcal{N}(i)} a_j x_j \quad (1)$$

$$\mathcal{E}((x_i - \mu_i)^2|x_j, j \in \mathcal{N}(i)) = \sigma_i^2 \equiv b + \sum_{j \in \mathcal{N}(i)} c_j (x_j - \mu_j)^2. \quad (2)$$

That is, the conditional mean of a coefficient is a linear function of the neighboring coefficients, and the conditional variance is a linear combination of mean-corrected squares of neighboring coefficients, plus a constant.<sup>11,12</sup> In order to limit the dimensionality of the parameter space, we also assume *homogeneity* in position and scale: the set of parameters  $\{a_j, b, c_j\}$  governing the dependencies are assumed to be the same for each neighborhood, regardless of absolute spatial position or scale.

These constraints do not uniquely specify the image density, either locally or globally. A well-known criterion for selecting a unique density that satisfies such constraints is that of *maximum entropy*.<sup>41</sup> The conditional density  $p(x_i|x_j, j \in \mathcal{N}(i))$  that has the highest entropy possible while still satisfying the constraints of Eqs. (1) and (2) is a Gaussian density, of the form:

$$p(x_i|x_j, j \in \mathcal{N}(i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right). \quad (3)$$

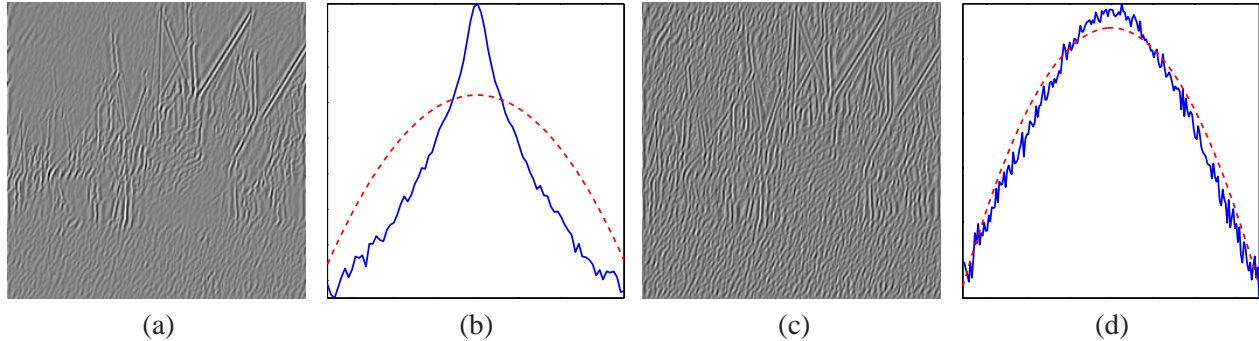
If we further assume that the joint distribution over all coefficients is of maximal entropy, subject to the neighborhood-conditioned marginal specified in Eq. (3), the resulting global model is a Markov Random Field (MRF)<sup>42</sup> over the wavelet coefficients. The Markov property implies that the density for a single coefficient conditioned on the full set of coefficients is equal to the density conditioned only on the coefficients in the neighborhood, and this conditional density is Gaussian as specified in Eq. (3). Despite this, it is important to emphasize that the full density on the coefficients is *not* jointly Gaussian in the conventional sense, since the variance of each coefficient depends on its neighbors.

## 2.2. Parameter Estimation

The parameters in the MRF model  $\{a_j, b, c_j\}$ , need to be estimated from images. For a given statistical model, the most common method for estimating model parameters from data is to maximize the likelihood function. However, the joint density of the MRF-based model is defined implicitly with the ensemble of all 1-D neighborhood conditional distributions, and direct maximization of likelihood is intractable. Instead, we can use *maximum pseudo-likelihood* (MPL)<sup>43</sup> estimation, which finds parameters that maximize the product of all 1-D conditionals,  $\prod_i p(x_i|x_j, \forall j)$ , or equivalently  $\sum_i \log p(x_i|x_j, \forall j)$ . In the case of our MRF model, the MPL parameter estimation reduces to solving

$$\operatorname{argmin}_{a_j, b, c_j} \sum_i \left[ \log \sigma_i^2 + \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right] \quad (4)$$

with the additional constraint that  $\sigma_i^2 > 0$  for all  $i$ . This expression differs substantially from previous least-squares forms for estimating variance conditioned on a neighborhood.<sup>11,12</sup> But it is similar to the method



**Fig. 2.** (a) A subband of the steerable pyramid decomposition of a photographic image. (b) Log histogram of wavelet subband shown in (a) (blue solid line). Also shown is the log of a Gaussian density with the same mean and variance (an inverted parabola, red dashed line). (c) The DNT transformation of the subband in (a). (d) Log histogram of DNT subband shown in (c) (blue solid line). Also shown is a Gaussian with the same mean and variance (red dashed line).

described in [ref. 14], which was developed in the absence of a global statistical model for purposes of eliminating statistical dependencies between coefficients. This nonlinear constrained optimization problem may be solved using numerical conjugate gradient descent methods, which lead to a minimum of the objective function in Eq. (4), but do not guarantee global optimality.

### 3. DIVISIVE NORMALIZATION TRANSFORM

Because the joint density of the coefficients under the MRF model is implicit, it is difficult to perform standard operations (e.g., inference, sampling). But the form of the conditional density leads naturally to an invertible nonlinear image transformation that has significantly simpler statistical structures than the original linear representation. We define the divisive normalization transform (DNT), as

$$r_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2}} = \frac{x_i - \sum_{j \in N(i)} a_j x_j(i)}{\sqrt{b + \sum_{j \in N(i)} c_j (x_j - \sum_{k \in N(j)} a_k x_j)^2}}. \quad (5)$$

Shown in Fig. 2(a) is a wavelet subband of a photographic image, and in (c) is the DNT of this subband. Compared to the wavelet subband, the DNT coefficients are more homogeneous with similar statistics across different locations. Furthermore, as shown in Fig. 2(b) and (d), the DNT representation has a marginal distribution much closer to Gaussian than that of the original subband. A similar observation was first made for local derivative measurements.<sup>10</sup> Under the MRF model, as specified by the 1D conditionals of Eq. (3), each normalized coefficient  $r_i$  has a Gaussian distribution with zero mean and unit variance. Thus, the normalized representation is both homogeneous and marginally Gaussian, unlike the original linear representation. Each DNT coefficient,  $r_i$ , is independent of  $\sigma_i^2$  (see Appendix), and is thus independent of the original coefficients in its neighborhood,  $x_j, j \in N(i)$ . The DNT coefficients are *not* guaranteed to be independent of each other,<sup>15</sup> but empirical observations suggest that they have significantly less dependency than in the original linear representation.<sup>11, 14, 15</sup>

#### 3.1. Inverting the Transformation

Given a DNT representation of an image, we need a method of recovering the original image. There are three basic possibilities. First, we could retain the mean and variance values,  $\mu_i$  and  $\sigma_i^2$ . While this makes inversion straightforward, it requires substantial storage overhead, and is likely to lead to added complexity in applications. Second, we could choose an ordering for the coefficients, and enforce a rule that the conditioning neighborhood for each coefficient contains only coefficients occurring earlier in the list (that is, for each  $i$  and  $j \in N(i)$ , we

require  $j < i$ ). For example, we could only include coefficients at coarser scales in the divisive neighborhood, as was done in a previous implementation of a ‘‘contrast pyramid’’,<sup>44</sup> as well as a previous normalization pyramid.<sup>16</sup> Inversion in this case is again simple, but the restriction to coarse-scale or asymmetric spatial neighborhoods may limit the statistical performance of the model. Finally, we can try to invert the unrestricted DNT, as expressed in Eq. (5). Several iterative procedures to accomplish this have been studied for a normalized version of the blocked discrete cosine transform (DCT),<sup>13,17</sup> however, these procedures cannot be extended beyond small image patches due to the lack of efficient procedure to check convergence condition. Here, we develop a more efficient and reliable procedure that guarantees to converge, and apply it to a wavelet transform.

To describe the inversion procedure, we first rewrite Eq. (5) in matrix/vector form. Denote  $\mathbf{x} = (x_1, \dots, x_D)^T$  as the vectorized wavelet representation of an image (including coefficients in all subbands),  $\mathbf{r} = (r_1, \dots, r_D)^T$  as the corresponding DNT coefficients, and  $\mathbf{b}$  a  $D$ -dimensional vector whose elements are all equal to parameter  $b$ . Furthermore, let  $A$  and  $C$  be the matrices that compute sums over the neighborhoods using weights  $a_j$  and  $c_j$ , respectively. In this notation, for example,  $A\mathbf{x}$  is equivalent to the right side of Eq. (1). These  $D \times D$  matrices implement a generalized form of convolution, and thus are efficient to compute.

Using this notation, the DNT is performed with the following sequence of operations:

$$\mathbf{y} = (\mathcal{I}_D - A)\mathbf{x} \quad (6)$$

$$\mathbf{z} = \mathbf{y} \otimes \mathbf{y} \quad (7)$$

$$\mathbf{u} = \mathbf{z} \oslash (\mathbf{b} + C\mathbf{z}) \quad (8)$$

$$\mathbf{r} = \text{sgn}(\mathbf{y}) \otimes \sqrt{\mathbf{u}} \quad (9)$$

where  $\mathbf{y}$ ,  $\mathbf{z}$  and  $\mathbf{u}$  are intermediate variables,  $\mathcal{I}_D$  is the  $D$ -dimensional identity matrix, the  $\sqrt{\cdot}$  and  $\text{sgn}(\cdot)$  operations are applied to each vector element, and  $\otimes$  and  $\oslash$  denote element-wise multiplication and division, respectively. To invert the transformation, we simply invert each step in Eqs. (6) - (9):

$$\mathbf{u} = \mathbf{r} \otimes \mathbf{r} \quad (10)$$

$$\mathbf{z} = (\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)^\dagger (\mathbf{u} \otimes \mathbf{b}) \quad (11)$$

$$\mathbf{y} = \text{sgn}(\mathbf{r}) \otimes \sqrt{\mathbf{z}} \quad (12)$$

$$\mathbf{x} = (\mathcal{I}_D - A)^\dagger \mathbf{y}. \quad (13)$$

where  $\dagger$  is pseudo-inverse, and  $\mathcal{D}(\cdot)$  is the operator that generates a diagonal matrix, with diagonal elements equal to those of its vector argument. The inversion of Eq. (8) relies on the fact that  $\mathbf{u} \otimes \mathbf{v} = \mathcal{D}(\mathbf{u})\mathbf{v}$  for any vectors  $\{\mathbf{u}, \mathbf{v}\}$ . If the matrices in Eqs. (11) and (13) are full rank, the inversion procedure will converge to the unique correct solution. Otherwise, the system will converge to the least-square optimal solution. In particular, the matrix in Eq. (11) will be stably invertible as long as the eigenvalues of symmetric matrix  $\mathcal{D}(\mathbf{u})C$  are not too close to one.

For small dimensionality  $D$ , the matrix inversion steps in Eqs. (11) and (13) may be computed directly, but for large  $D$  (e.g., a wavelet representation of a typical image) this is computationally prohibitive. Instead we develop an efficient iterative algorithm for these steps. First, we transform Eqs. (11) and (13) into two equivalent linear equations:

$$\begin{aligned} (\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)^T (\mathbf{u} \otimes \mathbf{b}) &= (\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)^T (\mathcal{I}_D - \mathcal{D}(\mathbf{u})C) \mathbf{z} \\ (\mathcal{I}_D - A)^T \mathbf{y} &= (\mathcal{I}_D - A)^T (\mathcal{I}_D - A) \mathbf{x}. \end{aligned} \quad (14)$$

The two equations have the same solution as the original matrix inversion problem. However, the right-hand sides of these equations use symmetric positive definite matrices, and are thus easily solved using conjugate gradient descent,<sup>45</sup> with a guarantee of convergence to the correct solution. The matrix-vector products of Eq. (14) may all be efficiently implemented by linearly combining convolution operations over subbands.

## 4. PERCEPTUAL RELEVANCE

We have developed a DNT representation based on an explicit statistical model. A number of articles in the perception literature have shown that the divisive normalization operation can effectively explain masking phenomena, in which the presence of large coefficients can reduce the visibility of errors in coefficients that are nearby in location, orientation, or scale.<sup>7-9</sup> In this section, we demonstrate the perceptual relevance of the representation by testing its resilience to noise contamination. A perceptually-matched representation should have the property that equal-size distortions in the representation should correspond to equally visible distortions in the image domain. Thus, adding white noise to the representation should lead to equally visible distortions throughout the image. On the other hand, adding white noise in a non-perceptual representation (e.g., the pixel domain) produces distortions that are homogeneously distributed throughout the image, but whose visibility varies greatly depending on the local image context. For example, distortions in smooth regions will be more noticeable than those near discontinuous features.

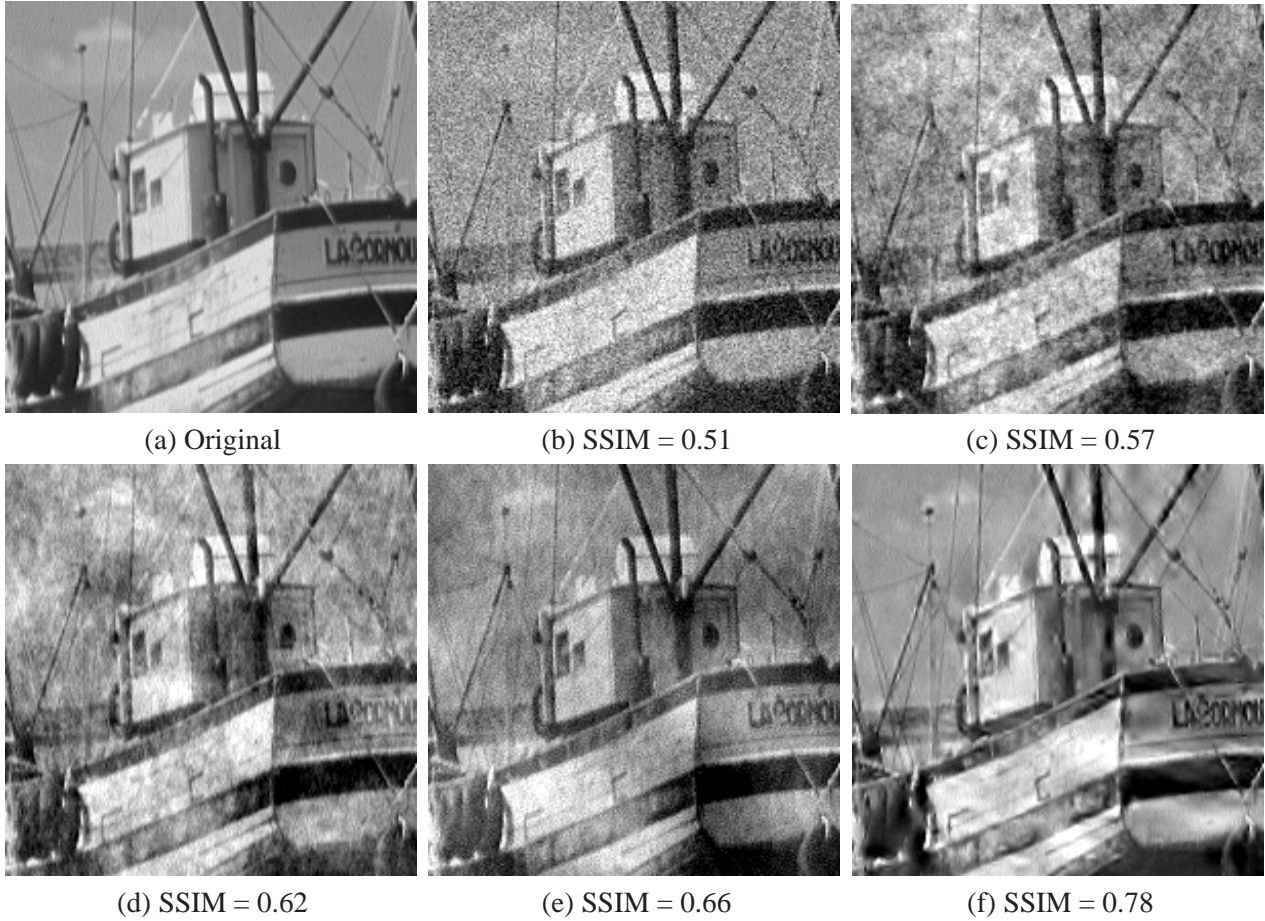
We compare the DNT representation with four other linear representations: raw pixels, Fourier transform, orthogonal wavelet and 8-orientation steerable pyramid. The DNT representation used an 8-orientation steerable pyramid as the front-end linear representation, and a neighborhood containing a  $5 \times 5$  spatial block of coefficients and  $3 \times 3$  blocks in two adjacent orientations and the next coarsest scale. We perturb an image by adding white Gaussian noise to the coefficients of each representation, and adjusting the variance of this noise so that the reconstructed distorted images all have the same mean square error relative to the original image. For wavelet and steerable pyramid representations, the noise amplitude is matched to a given fraction of the standard deviation for each subband. For the Fourier representation, the noise is Gaussian with a power spectrum  $1/(a + bf')$ , where the parameters  $\{a, b, \gamma\}$  are chosen to best fit the Fourier amplitude of the original image.

Shown in Fig. 3 is the result of this procedure for a particular photographic image, with the pixel-domain PSNR set to  $25dB$ . The perceptual quality of the images is dramatically different. For linear representations, the noise is statistically homogeneous and independent of the image content. For the DNT, large perturbations are confined to the vicinity of high contrast image features (e.g., edges), where they are less visible because of masking effects.

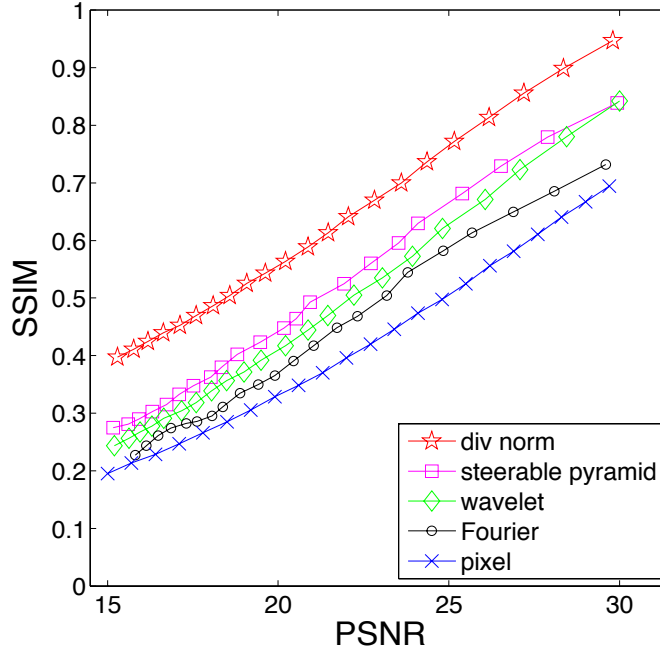
To quantify the perceptual difference between these images as a function of image PSNR, we used a recent perceptual image quality metric, the structural similarity image metric (SSIM).<sup>46</sup> Numerically, the SSIM index lies in the range between 0 and 1, with larger values indicating better image perceptual quality. We performed the analysis on a set of four images, and for PSNR levels ranging from  $15dB$  to  $30dB$ . For each representation and noise level, we generated 10 example images and computed the average SSIM over these. Shown in Fig. 4 are the plots of SSIM for each image representation as a function of the PSNR (in the pixel domain). The plot indicates that across all noise levels, the images for which noise was added in the DNT domain should appear less distorted (according to SSIM) than those for which noise was added to any of the other representations.

## 5. COMPRESSION

Next we consider the DNT as a substrate for image compression. The dual statistical and perceptual properties of the DNT suggest that it should be quite effective at jointly optimizing the tradeoff between rate and distortion.<sup>17</sup> To test this, we estimated rate-distortion curves for both linear and divisive orthogonal wavelet representations, under scalar quantization. For the DNT, we used uniform quantization, with an optimal binsize chosen for each band. For the linear wavelet decomposition, we used a Lloyd-Max quantizer similar to that used in JPEG2000.<sup>47</sup> For each subband, the bin boundaries and values were chosen to optimize the mean squared error of the quantized subband with a given number of bins. For all quantizers, we forced one bin to be centered over the origin.



**Fig. 3.** (a) A photographic image. (b)-(f) Perturbations of the image shown in (a) resulting from adding white Gaussian noise to coefficients of different image representations. For each image, the noise level was adjusted to achieve a PSNR of  $25dB$  in the pixel domain. Also shown are SSIM perceptual quality scores<sup>46</sup> (larger values should correspond to better perceptual appearance). (b) Raw pixel, (c) Fourier domain, (d) wavelet domain, (e) steerable pyramid, (f) DNT domain.



**Fig. 4.** Perceptual distortion of images resulting from addition of white noise in different representations, as a function of PSNR in the pixel domain. See text for details.

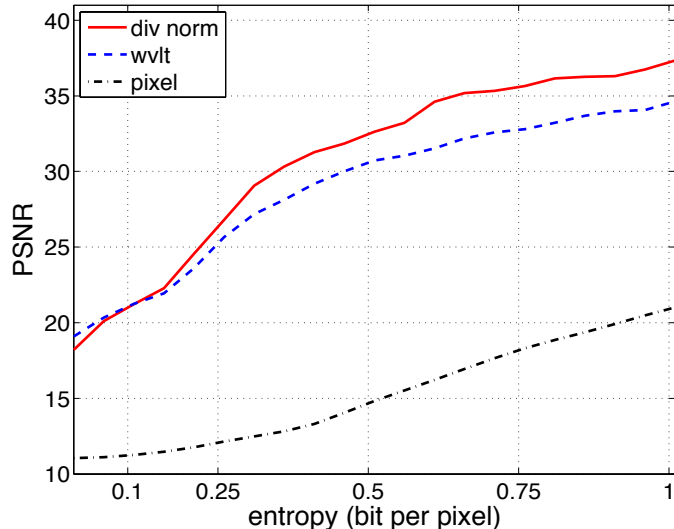
For the rate, we used a direct (plugin) estimate of the Shannon entropy of each subband, and for distortion we used the mean squared error. A rate-distortion curve was computed for each subband, and optimal bit allocation between subbands was accomplished by choosing equal-slope points on these curves. For each such slope, we summed the corresponding entropy and squared error across all bands, producing a single point on the overall rate-distortion curve. Repeating this over many slopes, and interpolating between the resulting points, produces the overall rate-distortion curve for the representation.

For all tests, we used images of size  $512 \times 512$ , subtracted the mean, and then decomposed into subbands of a six-level separable pyramid using a 9-tap quadrature mirror filter set<sup>48</sup>. The DNT is constructed using a  $3 \times 3$  spatial neighborhood with a generic set of parameters learned over a large set of images so as to simplify the encoding procedure.

Shown in Fig. 5 are rate distortion curves for the divisive and linear wavelet representations, as well as for the raw pixel domain (using Lloyd-Max quantization). Each curve is averaged over five standard grayscale images (“boats”, “lena”, “barbara”, “baboon”, and “goldhill”). Note that overall the DNT indicates better coding performance than the linear wavelet, especially for the high bit rate, despite the fact that the linear wavelet curve is based on a more powerful quantization scheme.

To illustrate the perceptual benefits of the DNT, Fig. 6 shows cropped portions (size  $256 \times 256$ ) of two images with the same coding cost (0.06 bits per pixel). The DNT image is seen to contain many detailed structures that are completely absent from the linear wavelet example. At the same time, the DNT image is also smoother, presumably because the quantized coefficients are smoothed during the inverse transform. Below each image is the SSIM perceptual quality measure, which reflects our subjective experience, assigning a higher value to the DNT image.





**Fig. 5.** Rate distortion curves, averaged over five grayscale test images (see text). The x-axis indicates the coding cost, in bits-per-pixel, estimated by marginal Shannon entropy, and the y-axis indicates to the peak-signal-to-noise-ratio (PSNR),  $20 \log_{10}(255/\sigma_e)$ , where  $\sigma_e$  is the standard deviation of the error. The red solid line corresponds to the DNT representation with uniform quantization. The blue dashed line corresponds to a linear wavelet representation with Lloyd-Max quantization. The black dot-dashed line corresponds to Lloyd-Max quantization of the raw pixels.

## 6. CONTRAST ENHANCEMENT

Representing images in the DNT domain mimics important aspects of human visual perception. In this section, we make use of this in enhancing the visual appearance of low contrast regions of photographic images. In general, contrast varies widely across a typical photographic image. The goal of contrast enhancement is to boost contrast in those regions where it is low or moderate, while leaving it unchanged where it is high. This requires that the contrast enhancement be adapted to the local image content.

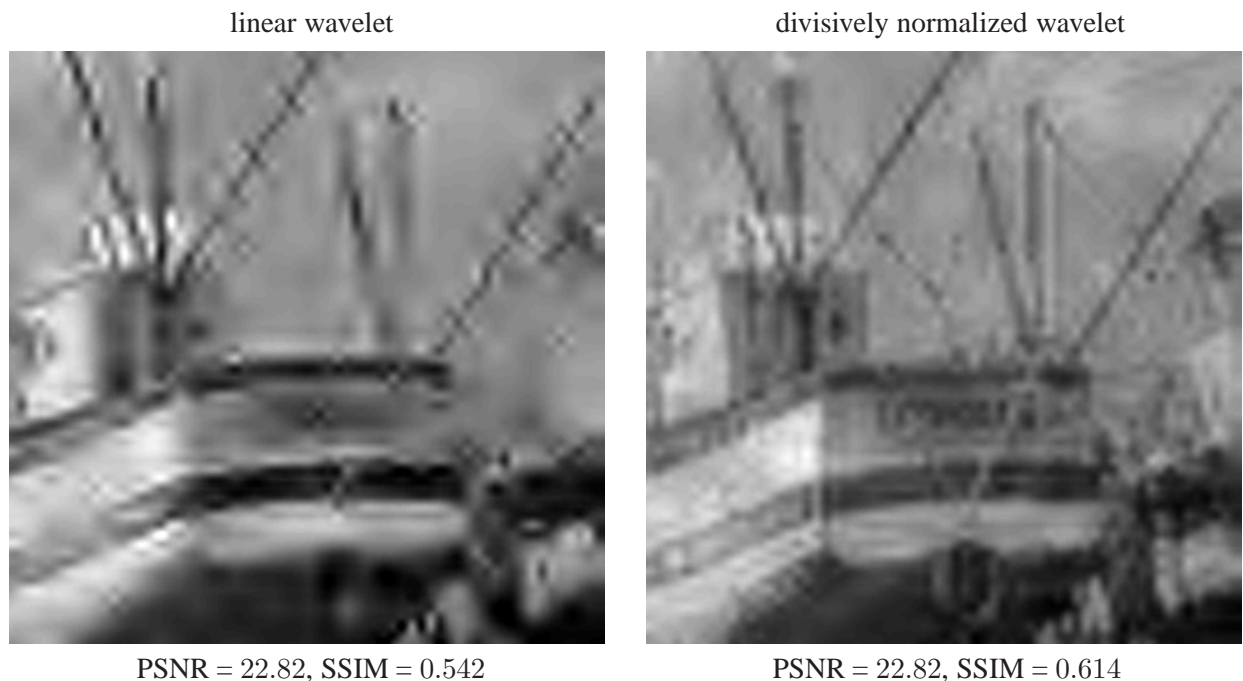
Many recent contrast adjustment methods are based on representing images in the wavelet domain.<sup>7,49–52</sup> A common theme is to use a point-wise nonlinear mapping (typically a power function with exponent less than 1, known as “gamma correction” when applied in the pixel domain) to adjust coefficient values image so that small values are boosted substantially, but large values are boosted only slightly or left alone. Such methods have been shown able to effectively improve the global appearance of the image. However, the independent adjustment of the transform coefficients often introduces substantial artifacts such as ripples and halos around high-contrast features, and these are typically eliminated in a post-processing step.<sup>49,51,52</sup>

### 6.1. Method

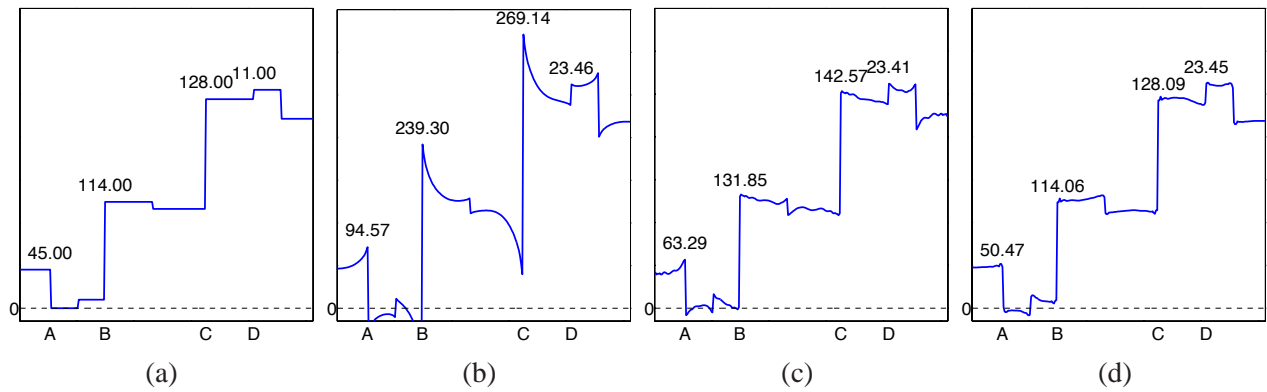
We first transform the steerable pyramid coefficients of the image into the DNT domain. In the DNT domain, each coefficient is individually boosted by a gamma-like nonlinear mapping:

$$g(x) = x \left[ (1 - \epsilon) \frac{|x|}{\tau} + \epsilon \right]^{\gamma-1} \quad (15)$$

where  $\epsilon, \tau$  and  $\gamma$  are adjustable parameters. The parameter  $\gamma \in (0, 1]$  determines the strength of the enhancement effect (small  $\gamma$  produces a large effect, and  $\gamma = 1$  has no effect). Parameter  $\epsilon$  (set to 0.01 in our examples) prevents amplification of noise in low signal areas, and parameter  $\tau$  represents a value that is unmodified by the



**Fig. 6.** Reconstructed image resulting from quantization in the linear wavelet domain (left) vs. quantization in the divisive normalization domain (right). Both representations had an estimated entropy of 0.06 bits/pixel.



**Fig. 7.** (a) A scan line of a test image containing vertical edges. (b)-(d) Contrast enhancement results of the test image with different methods: (b) global high pass filtering (also known as “unsharp masking”), (c) local gamma correction in wavelet domain, (d) global gamma correction in the divisively normalized domain. Numerical values indicate step sizes at four example locations (labeled A, B, C and D). Parameters in each operation were chosen so that the lowest-contrast step in the test image (point D) would be boosted to approximately the same value.

mapping (typically, this is set to the maximum value of the subband). The modified DNT coefficients are then inverted back to the linear pyramid domain, from which the contrast enhanced image is reconstructed. Although there is no guarantee that the nonlinearly modified DNT representation will remain invertible, we have not encountered any failures in practice.

## 6.2. Experiments

To illustrate the efficacy of image contrast enhancement with DNT representation, we first test it on an artificial image containing vertical edges of different strengths. In Fig 7(a), we show a horizontal scan line from this test image. We compared contrast enhancement in the DNT domain with two other methods based on linear image

representations. Fig. 7(b) shows the contrast enhancement result of a standard “unsharp masking” operation, which corresponds to convolution with a filter that boosts high frequency content. The image is first decomposed with a two-band steerable pyramid, then coefficients in each subband are multiplied by a scalar whose value is larger for higher-frequency bands. Note that the overall image contrast is boosted by this simple procedure, but the large and small steps in intensities are increased by the same factor. Fig. 7(c) shows the result of using a more elaborate contrast enhancement method, using a gamma correction in the wavelet domain<sup>49,51,52</sup>. With a two-band steerable pyramid decomposition of the image, the coefficient magnitudes were transformed with a nonlinear function (similar to Eq. (15), but  $|x|$  is replaced by the local standard deviation of coefficients). The contrast enhanced image is then obtained by inverting the pyramid decomposition. As shown in Fig. 7(c), this method alleviates some of the problems of the global high-pass filtering: large steps are now boosted only a small amount. On the other hand, visible ringing artifacts are introduced in the smooth regions around edges. In Fig. 7(d), contrast enhancement resulting from a global gamma correction in the DNT domain is shown. Compared to the two methods based on linear representations, the small steps in the image are boosted substantially while the large steps are only slightly changed, and this is achieved with a minimal introduction of artifacts.

Fig. 8 shows contrast enhancement results of four natural photographic images\*, using the gamma nonlinearity in the DNT domain. We used an 8-orientation, 6-level steerable pyramid, and estimated the DNT parameters for each image individually. The DNT transform is based on a generalized neighborhood including a  $5 \times 5$  spatial cluster of coefficients within the same subband, a  $3 \times 3$  cluster within two adjacent orientation subbands, and a  $3 \times 3$  cluster within the “parent” subband at the next coarsest scale, with parameters learned numerically. In processing the RGB color images, we converted them into the “L\*a\*b” color space, contrast enhanced the luminance channel, and then transformed the resulting image back into RGB space. The  $\gamma$  values are chosen for each individual image to produce visually acceptable results. For all images, the overall contrast is boosted, and details in the low contrast regions (e.g., the upper left corner in the first image showing the windows of a building) are significantly more visible.

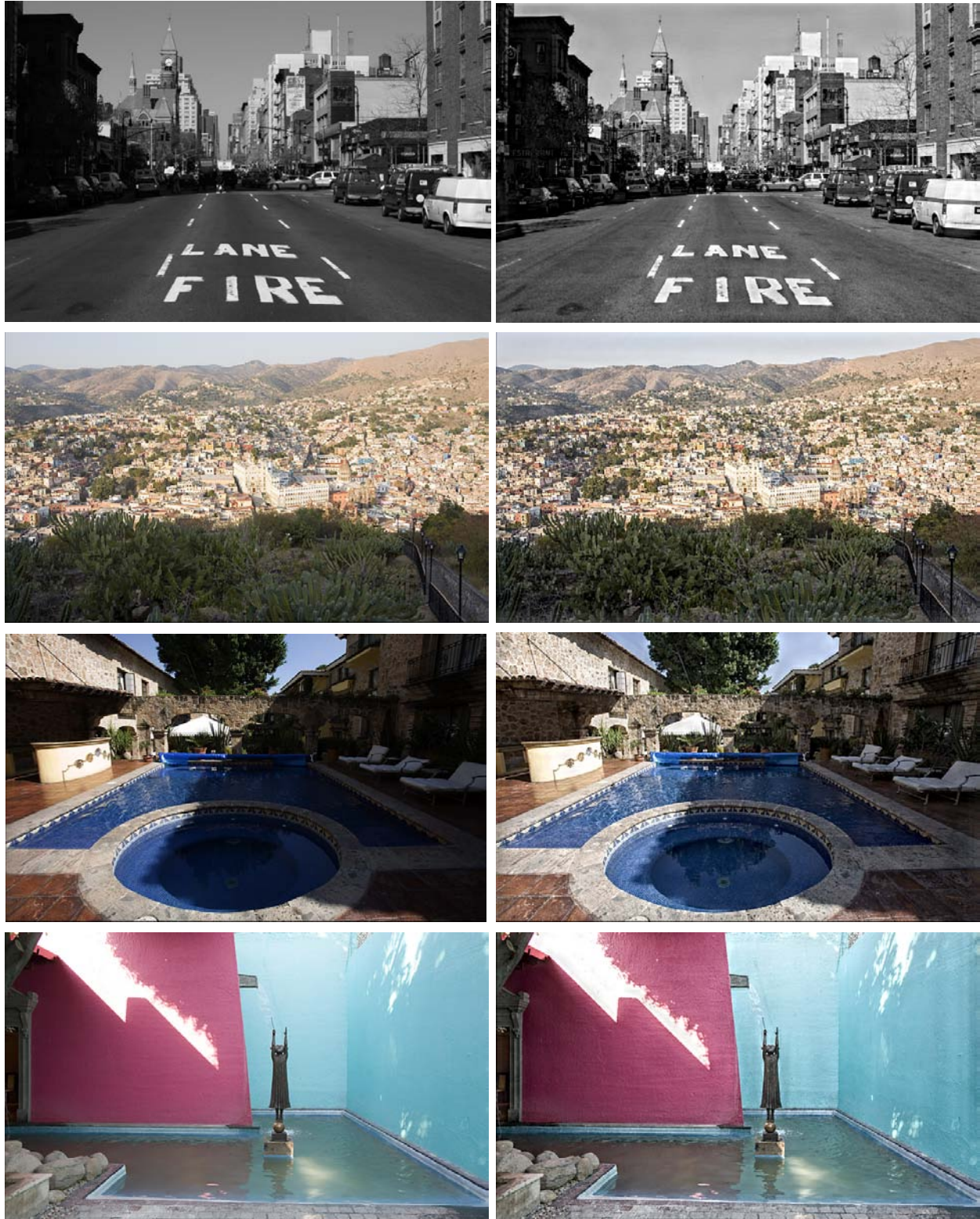
## 7. CONCLUSION

In this paper, we described a Markov random field model for images that captures the conditional mean and variance behavior between coefficients in a multiscale oriented decomposition. We showed that this naturally leads to a nonlinear image representation based on a local divisive normalization transform. We developed methods for estimating the model parameters, and for inverting the divisive normalization transform. We then demonstrated the statistical and perceptual relevance of this image representation, and showed a simple application to contrast enhancement.

We expect this representation to benefit a number of applications in image processing and machine vision. An obvious candidate is compression, in which a previous normalization representation based on a block-DCT decomposition exhibited substantial improvements in perceptual appearance of compressed images!<sup>17</sup> In addition, this representation seems promising as a substrate for applications which require nonlinear image manipulation, as in the contrast enhancement example. For example, blending of images or inpainting to remove artifacts could benefit from this type of representation. We also believe the model may be significantly extended by directly incorporating orientation and phase properties into the local predictions of mean and variance, and we are currently examining these possibilities.

---

\*Images courtesy of N. Bonnier and P. Greenspun.



**Fig. 8.** Contrast enhancement results of natural photographic images. On the left column are the original images, and on the right are the images resulting from gamma-boosting in the DNT domain (see text). Images courtesy of N. Bonnier and P. Greenspun.

## Appendix

We wish to show that if

$$p_x(x) = \int_{-\infty}^{\infty} \mathcal{N}(\mu, \sigma^2) p_{\mu, \sigma^2}(\mu, \sigma^2) d\mu d\sigma^2,$$

where

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

is a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ , then the normalized variable  $r = \frac{x-\mu}{\sqrt{\sigma^2}}$ , is Gaussian-distributed with mean 0 and variance 1, and is independent of  $\sigma^2$ .

**Proof:**

$$\begin{aligned} Pr(r \leq R) &= Pr(x \leq \sqrt{\sigma^2}R + \mu) \\ &= \int_{-\infty}^{\infty} p_{\mu, \sigma^2}(\mu, \sigma^2) d\mu d\sigma^2 \int_{-\infty}^{\sqrt{\sigma^2}R + \mu} \mathcal{N}(\mu, \sigma^2) dx. \end{aligned}$$

Then the PDF of  $r$  is obtained by taking derivative of  $Pr(r \leq R)$  with regards to  $R$ , as:

$$\begin{aligned} p_r(R) &= \frac{d}{dR} Pr(r \leq R) \\ &= \int_{-\infty}^{\infty} p_{\mu, \sigma^2}(\mu, \sigma^2) d\mu d\sigma^2 \frac{d}{dR} \int_{-\infty}^{\sqrt{\sigma^2}R + \mu} \mathcal{N}(\mu, \sigma^2) dx \\ &= \int_{-\infty}^{\infty} p_{\mu, \sigma^2}(\mu, \sigma^2) d\mu d\sigma^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{R^2}{2}\right) = \mathcal{N}(0, 1). \end{aligned}$$

Next, in order to show that  $r$  and  $\sigma^2$  are independent, we need to prove  $p_{r, \sigma^2}(R, Z) = p_r(R) \cdot p_{\sigma^2}(Z)$ . First,

$$\begin{aligned} Pr(r \leq R, \sigma^2 \leq Z) &= Pr(x \leq \sqrt{\sigma^2}R + \mu, \sigma^2 \leq Z) \\ &= \int \int_{-\infty}^Z p_{\mu, \sigma^2}(\mu, \sigma^2) d\mu d\sigma^2 \int_{-\infty}^{\sqrt{\sigma^2}R + \mu} \mathcal{N}(\mu, \sigma^2) dx \end{aligned}$$

Then

$$\begin{aligned} p_{r, \sigma^2}(R, Z) &= \frac{\partial}{\partial R} \frac{\partial}{\partial \sigma^2} Pr(r \leq R, \sigma^2 \leq Z) \\ &= \int_{-\infty}^{\infty} p_{\mu, \sigma^2}(\mu, Z) d\mu \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{R^2}{2}\right) \right] = p_r(R) \cdot p_{\sigma^2}(Z) \end{aligned}$$

## REFERENCES

1. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, pp. 607–609, 1996.
2. A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research* **37**(23), pp. 3327–3338, 1997.
3. M. Bethge, "Factorial coding of natural images: How effective are linear models in removing higher-order dependencies?," *J. Opt. Soc. Am. A* **23**, pp. 1253–1268, June 2006.
4. G. H. Granlund, "In search of a general picture processing operator," *Computer Graphics and Image Processing* **8**(2), pp. 155–173, 1978.

5. D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neural science* **9**, pp. 181–198, 1992.
6. W. S. Geisler and D. G. Albrecht, "Cortical neurons: Isolation of contrast gain control," *Vision Research* **8**, pp. 1409–1410, 1992.
7. J. Foley, "Human luminance pattern mechanisms: Masking experiments require a new model," *J. of Opt. Soc. of Amer. A* **11**(6), pp. 1710–1719, 1994.
8. A. Watson and J. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A* , pp. 2379–2391, 1997.
9. P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *IEEE Int'l. Conf. on Image Proc.*, pp. 982–986, 1994.
10. D. L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems* **5**, pp. 517–548, 1996.
11. E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, **1**, pp. 673–678, IEEE Computer Society, (Pacific Grove, CA), November 2-5 1997.
12. R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," **8**(12), pp. 1688–1701, 1999.
13. J. Malo, R. Navarro, I. Epifanio, F. Ferri, and J. Artigas, "Non-linear invertible representation for joint statistical and perceptual feature representation," *Lect. Not. Comp. Sci.* **1876**, pp. 658–667, 2000.
14. O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience* **4**, pp. 819–825, August 2001.
15. R. Valerio and R. Navarro, "Optimal coding through divisive normalization models of v1 neurons," *Network: Computation with neural systems* , pp. 579–593, 2003.
16. R. Valerio, E. P. Simoncelli, and R. Navarro, "Directly invertible nonlinear divisive normalization pyramid for image representation," in *Visual Content Processing and Representation- Proc 8th Int'l Workshop, VLBV*, N. Garcia, J. Martinez, and L. Salgado, eds., pp. 331–340, Springer, (Madrid, Spain), 18-19 Sep 2003. Lecture Notes in Computer Science, vol. 2849.
17. J. Malo, I. Epifanio, R. Navarro, and E. P. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Trans Image Processing* **15**, pp. 68–80, January 2006.
18. J. M. Gluckman, "Higher order pyramids: an early vision representation," in *European Conference on Computer Vision (ECCV)*, 2006.
19. P. J. Burt, "Fast filter transforms for image processing," *Comp. Graph. Image Proc.* **16**, pp. 20–51, 1981.
20. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**(12), pp. 2379–2394, 1987.
21. B. Wegmann and C. Zetsche, "Statistical dependence between orientation filter outputs used in an human vision based image code," in *Proc Visual Comm. and Image Processing*, **1360**, pp. 909–922, (Lausanne, Switzerland), 1990.
22. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans Information Theory* **38**, pp. 587–607, March 1992. Special Issue on Wavelets.
23. M. K. Mkhçak, I. Kozinstev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Trans. on Sig. Proc.* **6**(12), pp. 300–303, 1999.
24. S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.* **11**, pp. 674–693, July 1989.
25. C. Bouman and K. Sauer, "A generalized Gaussian image model for edge preserving MAP estimation," *IEEE Trans. Image Proc.* **2**, pp. 296–310, July 1993.
26. E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc 3rd IEEE Int'l Conf on Image Proc.*, **1**, pp. 379–382, IEEE Sig Proc Society, (Lausanne), September 16-19 1996.
27. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors," *IEEE Trans. Info. Theory* **45**, pp. 909–919, 1999.
28. L. Parra, C. Spence, and P. Sajda, "Higher-order statistical properties arising from the non-stationarity of natural signals," in *Adv. Neural Information Processing Systems (NIPS\*00)*, T. K. Leen, T. G. Dietterich, and V. Tresp, eds., **13**, pp. 786–792, MIT Press, (Cambridge, MA), May 2001.
29. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.* **46**, pp. 886–902, April 1998.

30. J. Romberg, H. Choi, and R. Baraniuk, "Bayesian wavelet domain image modeling using hidden Markov trees," in *Proc. IEEE Int'l Conf on Image Proc.*, (Kobe, Japan), October 1999.
31. S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Wavelet image coding based on a new generalized Gaussian mixture model," in *Data Compression Conf.*, (Snowbird, Utah), March 1997.
32. M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems (NIPS\*99)*, S. A. Solla, T. K. Leen, and K.-R. Müller, eds., **12**, pp. 855–861, MIT Press, (Cambridge, MA), May 2000.
33. J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, **1**, pp. 541–547, 1999.
34. A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probability," *IEEE Pat. Anal. Mach. Intell.* **28**(9), 2002.
35. M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," *Applied and Computational Harmonic Analysis* **11**, pp. 89–123, July 2001.
36. S. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Pat. Anal. Mach. Intell.* **19**(11), 1997.
37. S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *IEEE Conf Computer Vision and Patt Rec.*, **II**, pp. 860–867, June 2005.
38. P. Gehler and M. Welling, "Product of edge-perts," in *Adv. Neural Information Processing Systems (NIPS\*05)*, Y. Weiss, B. Schölkopf, and J. Platt, eds., **18**, MIT Press, (Cambridge, MA), 2006.
39. S. Osindero, M. Welling, and G. Hinton, "Topographic product models applied to natural scene statistics," *Neural Computation* **18**(2), 2005.
40. S. Lyu and E. P. Simoncelli, "Statistical modeling of images with fields of Gaussian scale mixtures," in *Adv. Neural Information Processing Systems (NIPS\*06)*, B. Schölkopf, J. Platt, and T. Hofmann, eds., **19**, MIT Press, (Cambridge, MA), May 2007.
41. E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.* **106**, pp. 620–630, 1957.
42. P. Winkler, *Image Analysis, Random Fields And Markov Chain Monte Carlo Methods*, Springer, 2nd ed., 2003.
43. J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B* **36**(2), pp. 192–225, 1974.
44. E. Peli, "Contrast in complex images," *J Opt Soc Am A* **7**(10), pp. 2032–2040, 1990.
45. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, Cambridge, 2nd ed., 2002.
46. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Perceptual image quality assessment: From error visibility to structural similarity," *tip* **13**(4), pp. 600–612, 2004.
47. D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic, 2002.
48. E. P. Simoncelli and E. H. Adelson, "Subband transforms," in *Subband Image Coding*, J. W. Woods, ed., ch. 4, pp. 143–192, Kluwer Academic Publishers, Norwell, MA, 1990.
49. R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *ACM SIGGRAPH*, 2002.
50. J. L. Starck, F. Murtagh, E. Candes, and D. L. Donoho, "Gray and color image contrast enhancement by the curvelet transform," pp. 706–717, 2003.
51. Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," in *ACM Transactions on Graphics (Siggraph Proceedings)*, **24**(3), pp. 836–844, 2005.
52. N. Bonnier and E. P. Simoncelli, "Locally adaptive multiscale contrast optimization," in *Proc 12th IEEE Int'l Conf on Image Proc.*, **I**, pp. 949–952, IEEE Computer Society, (Genoa, Italy), September 11-14 2005.