Nonlinear Image Representation Using Divisive Normalization

Siwei Lyu and **Eero P. Simoncelli**
Howard Hughes Medical Institute, and Center for Neuroscience, New York University

Siwei Lyu: lsw@cns.nyu.edu; Eero P. Simoncelli: eero@cns.nyu.edu

Abstract

In this paper, we describe a nonlinear image representation based on divisive normalization that is designed to match the statistical properties of photographic images, as well as the perceptual sensitivity of biological visual systems. We decompose an image using a multi-scale oriented representation, and use Student's t as a model of the dependencies within local clusters of coefficients. We then show that normalization of each coefficient by the square root of a linear combination of the amplitudes of the coefficients in the cluster reduces statistical dependencies. We further show that the resulting divisive normalization transform is invertible and provide an efficient iterative inversion algorithm. Finally, we probe the statistical and perceptual advantages of this image representation by examining its robustness to added noise, and using it to enhance image contrast.

1. Introduction

Choosing a good representation is of fundamental importance to image processing and computer vision applications. Most recent efforts in finding image representations focus on linear transforms optimized to minimize statistical dependencies. The basis functions obtained using these methods are localized oriented filters at different spatial scales, which are loosely referred as "wavelets". Such transforms have been used as an effective substrate for a wide variety of applications in computer vision and image processing. Nevertheless, linear transforms do not completely eliminate higher-order statistical dependencies in photographic images [20, 2]. It is thus natural to develop invertible nonlinear transforms that can reduce such higher-order statistical dependencies.

Several recent image representations [25, 16, 15, 10] include spatially varying *divisive normalization* (DN) as simple nonlinear map, where each component in a cluster of coefficients is divided by the square root of a linear combination of the squared amplitudes of its neighbors. Divisive Normalization was originally motivated by observed properties of biological vision, where it was used to explain nonlinearities in the responses of cortical neurons [e.g., 12, 9], nonlinear masking phenomenon in visual perception [e.g., 8, 27, 24], and has also been empirically shown to reduce statistical dependencies of the original linear representation [e.g., 20, 4, 19].

In this paper, we provide a more flexible form of the DN representation. We first justify the Student's t model as a good description of some observed statistical properties of photographic images in the wavelet domain. We then show that DN transform is an approximation to remove higherorder dependencies from the Student's t model. Extending the divisive normalization transform to images of arbitrary sizes, we show an efficient iterative inversion algorithm that is guaranteed to converge. We probe the statistical and perceptual advantages of this image representation by demonstrating that it is more robust to additive noise than conventional linear transforms. We also apply a simple method of adaptive contrast enhancement to the DN representation, demonstrating that the resulting images are nearly artifact-free compared to the results of applying such enhancement in a linear representation.

2. Photographic Image Wavelet Statistics

Photographic images exhibit highly non-Gaussian statistical behaviors. When decomposed in a multi-scale oriented decomposition, the coefficients tend to have highly kurtotic marginal distributions [5, 7]. More importantly, there are higher-order statical dependencies between coefficients at nearby locations, orientations and scales [28, 4] that cannot be modeled by Gaussian models. Such non-Gaussian statistical properties of photographic image wavelets can be more clearly illustrated by examining the conditional distribution of adjacent pairs of wavelet coefficients. Consider (x_1, x_2) representing two coefficients at adjacent spatial locations, within the same subband of a wavelet decomposition[1]. We examine the conditional histogram of this pair of coefficients, gathered over the whole subband as in [20]. Fig.1 shows this conditional histogram, which illustrates several important aspects of the relationship between the two coefficients. First, the conditional mean and variance of x_1 exhibit strong dependency on the value of x_2. In this case, 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).

3. Student's t Model

The behavior of multi-scale decomposition coefficients as described in previous section can be captured with a Student's t model of a cluster of coefficients. Formally, a *D*-dimensional Student's t variable **x** with zero mean has density:

$$p(\mathbf{x}|\textstyle\sum, \alpha, \beta) \propto \left(1+\frac{\alpha}{2}\mathbf{x}^T\textstyle\sum^{-1}\mathbf{x}\right)^{-\beta-\frac{D}{2}} \quad (1)$$

where $\Gamma(\cdot)$ is the Gamma function, Σ is a positive definite matrix, and α, β are two model parameters. Previously, Student's t has been proposed as an image model mainly for its convenience in computation [23, 18, 29]. We show that for a Student's t variable **x**, the conditional mean and variance of its ith component given the rest,

---

[1]In this work, we use steerable pyramid [21], which a particular type of overcomplete wavelet transform. However, the observed regularities are fairly robust across different multi-scale decompositions.

$$\mathcal{E}(x_i|x_j,\ j \neq i)=\mu_i \ \text{and} \ \mathrm{var}(x_i|x_j,\ j \neq i)=\sigma_i^2, \quad (2)$$

can be expressed as:

$$\mu_i=\sum_{j \neq i} a_j\, x_j \ \text{and} \ \sigma_i^2=b+\sum_{j \neq i} c_j(x_j-\mu_j)^2. \quad (3)$$

That is, the conditional mean of one wavelet 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. When $D = 2$, this gives the exact dependency as shown in Fig.1.

4. Divisive Normalization

Student's t is a special case of the more general density family of Gaussian scale mixtures [1]. As such, it has been shown that a Student's t variable **x** can be equivalently expressed as the product of two independent random variables as

$$\mathbf{x}=\mathbf{u}\,\sqrt{z}, \quad (4)$$

where **u** is a D-dimensional zero-mean Gaussian variable with covariance matrix Σ, and z is a positive-valued variable with an inverse Gamma density of parameters (a, β), as:

$$p_z(z|\alpha,\ \beta)=\frac{1}{\alpha^\beta \Gamma(\beta)} z^{-\beta-1}\exp\left(-\frac{1}{\alpha z}\right). \quad (5)$$

Note that the Gaussian scale mixtures formulation of a Student's t variable provides a separation of the statistical dependencies in **x**: the Gaussian component **u** carries only second-order dependencies, while the latent variable z creates higher-order dependencies by correlating all the Gaussian components through a common scaling factor. If we use a decorrelation transform (e.g., PCA) to diagonalize the covariance matrix Σ first, the Gaussian component **u** then has zero dependency, i.e., all components in **u** are mutually independent. We can then remove any additional higherorder dependencies by dividing **x** by z. According to Eq.(4), such an operation results in the independent Gaussian variable **u**.

However, as z is a latent variable, its value is not observable. Since we only have access to variable **x**, one natural approximation to the division of z is to first obtain an estimate of z from **x**, $\hat{z}$, and then use $\mathbf{x}/\sqrt{\hat{z}}$ to approximate $\mathbf{x}/\sqrt{z}$. In particular, we can consider the maximum a posterior (MAP) estimate $\hat{z}_{MAP} = \arg\max_z p(z|\mathbf{x})$, or the the minimum mean squared error (MMSE) estimate, computed as $\hat{z}_{MAP} = E(z|\mathbf{x})$. In the case of Student's t model, however, both estimators take the following form

$$\propto \sqrt{b+\sum_j c_j x_j^2}, \quad (6)$$

where $b=\dfrac{1}{\alpha}$ and c_j s correspond to the diagonals of the diagonalized covariance matrix Σ. Due to space limitations, we cannot provide the proof here. Combining with the front-end decorrelation transform, we can approximately remove statistical dependencies in the Student's t variable by the following transform:

$$r_i = \frac{x_i - \sum_j a_j\, x_j(i)}{\sqrt{b + \sum_j c_j \left(x_j - \sum_k a_k\, x_j\right)^2}}, \quad (7)$$

where parameters a_k are computed from the eigen-vectors of matrix Σ.

The transform described in Eq.(7) coincides with the divisive normalization transform as described in [19], which has been shown to empirically reduce statistical dependencies of multi-scale decomposition coefficients of photographic images. The parameters in the divisive normalization transform, $\{b, c_k, a_k\}$ can be obtained by maximizing the likelihood of blocks of multi-scale decomposition coefficients under the model [23]. Shown in Fig.2(a) is a subband of a photographic image, and in (c) is the DNT of this subband. Compared to the original 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 a Gaussian than the original subband.

5. Iterative Inversion of DNT

The divisive normalization transform, Eq.(7), is nonlinear. Nevertheless it can be inverted [15], and here we develop a more efficient iterative procedure for doing so. To describe the inversion procedure, we first rewrite the forward DN transform in matrix and vector form. First denote $\mathbf{x} = (x_1, \cdots, x_D)^T$ as the vectorized multi-scale oriented representation of an image, including coefficients in all subbands, $\mathbf{r} = (r_1, \cdots, r_D)^T$ as its corresponding DN transform coefficients, and $\mathbf{1}$ is a D dimensional vector with all component being one. Furthermore, let A and C be the matrices formed by weights a_is and c_is, respectively. Note that these matrices, though may be huge in size due to their dimensionality of $D \times D$, have sparse block-Toeplitz (convolutional) [11] structures determined by the neighborhood configuration and the spatial homogeneity assumptions. With these notations and introducing intermediate variables, the DN transform is equivalent to the following sequence of operations:

$$\mathbf{y} = (\mathscr{I}_D - A)\mathbf{x} \quad (8)$$

$$\mathbf{z} = (y_1^2, \cdots, y_D^2)^T \quad (9)$$

$$\mathbf{u} = \mathbf{z} \oslash (b\mathbf{1} + C\mathbf{z}) \quad (10)$$

$$\mathbf{r} = \left(\text{sign}(y_1)\sqrt{u_1}, \cdots, \text{sign}(y_D)\sqrt{u_D} \right)^T \quad (11)$$

where $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{u}$ are all D dimensional vectors, $\mathcal{I}_D$ is the $D \times D$ identity matrix, and $\oslash$ denotes the element-wise division operator.

The inversion of the DN transformation is thus composed of reversing the steps in Eq.(8) – (11), as

$$\mathbf{u} = \left(r_1^2, \cdots, r_D^2 \right)^T \quad (12)$$

$$\mathbf{z} = b(\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)^{-1}\mathbf{u} \quad (13)$$

$$\mathbf{y} = \left(\text{sign}(r_1)\sqrt{z_1}, \cdots, \text{sign}(r_D)\sqrt{z_D} \right)^T \quad (14)$$

$$\mathbf{x} = (\mathcal{I}_D - A)^{-1}\mathbf{y}. \quad (15)$$

where $\otimes$ denotes element-wise multiplication, and $\mathcal{D}(\cdot)$ is the operator that takes a vector and turns it into the diagonal of a diagonal matrix. All steps are straightforward, except (13), which requires some explanation. First we make use of the fact that for two vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{u} \otimes \mathbf{v} = \mathcal{D}(\mathbf{u})\mathbf{v}$. So from Eq.(10) we have,

$$\mathbf{z} = \mathbf{u} \otimes (b\mathbf{1} + C\mathbf{z})$$
$$\Rightarrow \mathbf{z} = b\mathbf{u} + \mathcal{D}(\mathbf{u})C\mathbf{z}$$
$$\Rightarrow (\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)\mathbf{z} = b\mathbf{u}.$$

Inverting the matrix then gives Eq.(13).

Most of the computational cost in inverting the DN transform is spent on the two matrix inversions, of Eq.(13) and (15). For small D such as image patches, these steps may be computed directly. For large D (typically the case for multi-scale oriented representation of a whole image) it is computationally inhibitive to perform matrix inversion directly. On the other hand, note that inversion of the two big matrices are not necessary as we really need just to solve the two corresponding linear equations

$$(\mathcal{I}_D - \mathcal{D}(\mathbf{u})C)\mathbf{z} = b\mathbf{u}$$
$$(\mathcal{I}_D - A)\mathbf{x} = \mathbf{y}.$$

The matrices on the left hand of these two linear equations, though potentially very large, are highly structured. Analogous to the Landweber algorithm for solving the Fredholm integral equation of the first type [13], our problem is equivalent to solving these two linear equations:

$$(\mathscr{I}_D - \mathscr{D}(\mathbf{u})C)^T (\mathbf{u} \otimes \mathbf{b}) = (\mathscr{I}_D - \mathscr{D}(\mathbf{u})C)^T (\mathscr{I}_D - \mathscr{D}(\mathbf{u})C)\mathbf{z}$$
$$(\mathscr{I}_D - A)^T \mathbf{y} = (\mathscr{I}_D - A)^T (\mathscr{I}_D - A)\mathbf{x}. \tag{16}$$

The advantage of re-expressing the problem this way is that the right-hand sides of these equations are symmetric positive definite matrices, and thus they can be solved using conjugate gradient descent [17], with a guarantee of convergence to the correct solution. In computing the gradients, we take advantage of the fact that the matrix-vector product between a block-Toeplitz matrix and a vector is equivalent to a convolution [11], which may be efficiently implemented with fast Fourier transform.

6. Perceptual Relevance

We have described a DNT representation from the statistical viewpoint. A substantial body of work in the perception literature has shown that the divisive normalization operation can effectively explain masking phenomena, in which the presence of large coefficients reduces the visibility of errors in coefficients that are nearby in location, orientation, or scale [e.g., 8, 27, 24]. 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 generate equally visible distortions at all locations within the image. On the other hand, adding white noise in a non-perceptual representation (e.g., the pixel domain) will produce distortions of unequal visibility throughout the image, depending on the local image context.

We compare the DNT representation with four other linear representations: raw pixels, Fourier transform, orthogonal wavelet and steerable pyramid. The DNT representation used a steerable pyramid as the front-end linear representation, and a neighborhood including a 5 × 5 spatial neighbor block and 3 × 3 neighbor blocks in two adjacent orientations and upper scales. We perturb an image by adding white Gaussian noise to the coefficients of each representation, and adjusting the strength of this noise so that distorted images have the same mean square error relative to the original image. For the Fourier, wavelet, and steerable pyramid representations, the noise amplitude is matched to a given fraction of the standard deviation for each frequency band or subband to reduce visibility.

Shown in Fig 3 is the result of this procedure on a particular photographic image, with the fixed output PSNR set to 25dB. The perceptual quality of the images is drastically different. For linear representations, the noise affects the entire image uniformly, while in the DNT domain, 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) [26]. Numerically, SSIM index ranges between 0 and 1, with larger values indicating better image perceptual quality. We repeated this experiment on a set of four images, and over a wide range of image PSNRs (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 output PSNR. The plot indicates that the images for which noise was added in DNT domain are predicted by SSIM to appear less distorted than all other images, and across all noise levels.

7. Contrast Enhancement

Representing images in the divisive normalized domain reflects certain 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. All of this must be done without generating artifacts.

Many recent contrast adjustment methods are based on representing images in the wavelet domain [e.g., 6, 22, 8, 14, 3]. A common theme is to use a point-wise non-linear mapping (typically a power function with exponent less than 1, also called "gamma correction") 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 artifacts such as ripples and halos around high-contrast features, and these need to be eliminated in a post-processing step [6, 14, 3].

7.1. Method

We first transform the steerable pyramid coefficients of the image into the DNT domain, using a DNT transform using a generalized neighborhood system a 5×5 spatial neighborhood, a 3×3 neighborhood for two adjacent orientations subbands and a 3×3 neighborhood for upper scale subband, and optimally estimated parameters. In the DNT domain, each coefficient is individually boosted by a gamma correction type non-linear map:

$$g(x) = x \left[(1-\varepsilon)\frac{|x|}{\tau} + \varepsilon \right]^{\gamma-1} \quad (17)$$

where ε, τ and γ are adjustable parameters. The parameter $\gamma \in [0, 1)$ determines the strength of the enhancement effect (small γ produces a large effect, and $\gamma = 1$ has no effect). Parameter ε (set to 0.01) prevents amplification of noise in low signal areas, and parameter τ

represents a value that is unmodified by the mapping (typically, this is set to the maximum value of a subband). The modified DNT coefficients are then transformed back to the wavelet domain, from which the contrast enhanced image is reconstructed. Although there is no guarantee that the nonlinearly modified DNT representation will be invertible (see [15]), we have not encountered any failures in practice.

7.2. Experiments

To illustrate the efficacy of image contrast enhancement with DNT representation, we first test it on an artificial image containing vertical edges. In Fig.5(a), we show a horizontal scan line from the test image, and mark the amplitudes of the edges at four different locations (point A, B, C and D), covering a range of different strengths.

We compared contrast enhancement with DNT representation with two other methods using linear image representations. Fig.5(b) shows the contrast enhancement result of a widely used global "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 band. Note that the overall image contrast is boosted by this simple procedure, but the large jumps and low jumps in intensities are increased by an equal factor. Fig.5(c) shows the result of using a more elaborate contrast enhancement method, using a gamma correction in the wavelet domain [6, 14, 3]. With a two-band steerable pyramid decomposition of the image, the coefficient magnitudes were transformed with a gamma correction-like non-linear function (similar to Eq.(17), but the $|x|$ 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.5(c), this method alleviates some of the problems of the global high-pass filtering, as high intensity jumps are boosted only a small amount. On the other hand, artifacts are introduced in the smooth regions around edges. In Fig.5 (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 low contrasts in the images are greatly boosted while the large jumps are only slightly changed. This is achieved with only a minimal introduction of artifacts.

Fig.6 shows contrast enhancement results of four natural photographic images[2] with divisive normalized representation. In processing the RGB color images, we first convert them into the L*a*b* color space, contrast enhancing the L* channel, and transformed the resulting image back to the RGB space. The γ values are chosen for each individual image to produce visually acceptable results. Note that across 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) have become significantly more visible.

---

[2]Images courtesy of N. Bonnier and P. Greenspun.

8. Conclusion

In this paper, we start with a justification of the Student's t model for images in the wavelet domain. We then showed that this naturally leads to a non-linear 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: a previous normalization representation based on a block-DCT decomposition exhibited substantial improvements in perceptual appearance of compressed images [15]. 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 believe the model may be significantly extended by directly incorporating local orientation and phase properties into the local predictions of mean and variance, and we are currently working in this direction.

References

1. Andrews DF, Mallows CL. Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B. 1974; 36(1):99–102.

2. Bethge M. Factorial coding of natural images: How effective are linear models in removing higher-order dependencies? J Opt Soc Am A. Jun; 2006 23(6):1253–1268.

3. Bonnier, N.; Simoncelli, EP. Locally adaptive multiscale contrast optimization. Proc 12th IEEE Int'l Conf on Image Proc; Genoa, Italy. September 11–14 2005; p. 949-952.

4. Buccigrossi RW, Simoncelli EP. Image compression via joint statistical characterization in the wavelet domain. 1999; 8(12):1688–1701.

5. Burt PJ. Fast filter transforms for image processing. Comp Graph Image Proc. 1981; 16:20–51.

6. Fattal, R.; Lischinski, D.; Werman, M. ACMSIGGRAPH. 2002. Gradient domain high dynamic range compression.

7. Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Amer. 1987; 4(12):2379–2394.

8. Foley J. Human luminence pattern mechanisims: Masking experimants require a new model. J of Opt Soc of Amer A. 1994; 11(6):1710–1719.

9. Geisler WS, Albrecht DG. Cortical neurons: Isolation of contrast gain control. Vision Research. 1992; 8:1409–1410. [PubMed: 1455713]

10. Gluckman, JM. Higher order pyramids: an early vision representation. European Conference on Computer Vision (ECCV); 2006.

11. Gray RM. Toeplitz and circulant matrices: A review. Foundations and Trends in Communications and Information Theory. 2006; 2(3):155–239.

12. Heeger DJ. Normalization of cell responses in cat striate cortex. Visual neural science. 1992; 9:181–198.

13. Landweber L. An iterative formula for Freholm integral equations of the first kind. Amer J Math. 1951; 73:615–624.

14. Li Y, Sharan L, Adelson EH. Compressing and companding high dynamic range images with subband architectures. ACM Transactions on Graphics (Siggraph Proceedings). 2005; 24:836–844.

15. Malo J, Epifanio I, Navarro R, Simoncelli EP. Nonlinear image representation for efficient perceptual coding. IEEE Trans on Image Proc. Jan; 2006 15(1):68–80.

16. Malo J, Navarro R, Epifanio I, Ferri F, Artigas J. Non-linear invertible representation for joint statistical and perceptual feature representation. Lect Not Comp Sci. 2000; 1876:658–667.

17. Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes. 2. Cambridge: 2002.

18. Roth S, Black M. Fields of experts: A framework for learning image priors. 2005; 2:860–867.

19. Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. Nature Neuroscience. Aug; 2001 4(8):819–825.

20. Simoncelli, EP. Statistical models for images: Compression, restoration and synthesis. Proc 31st Asilomar Conf on Signals, Systems and Computers; Pacific Grove, CA. November 2–5 1997; p. 673-678.

21. Simoncelli EP, Freeman WT. The steerable pyramid: A flexible architecture for multi-scale derivative computation. 1995; III:444–447.

22. Starck, JL.; Murtagh, F.; Candes, E.; Donoho, DL. Gray and color image contrast enhancement by the curvelet transform. 2003. p. 706-717.

23. Teh Y, Welling M, Osinderoi S, Hinton G. Energy-based models for sparse overcomplete representations. 2003; 4:1235–1260.

24. Teo, PC.; Heeger, DJ. Perceptual image distortion. IEEE Int'l. Conf. on Image Proc; 1994; p. 982-986.

25. Valerio, R.; Simoncelli, EP.; Navarro, R. Lecture Notes in Computer Science. Vol. 2849. Springer; 2003. Directly invertible nonlinear divisive normalization pyramid for image representation; p. 331-340.

26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Perceptual image quality assessment: From error visibility to structural similarity. tip. 2004; 13(4):600–612.

27. Watson A, Solomon J. A model of visual contrast gain control and patern masking. J Opt Soc Amer A. 1997:2379–2391.

28. Wegmann B, Zetzsche C. Statistical dependencies between orientation filter outputs used in human vision based image code. Proc Visual Comm And Image Proc. 1990; 1360:909–922.

29. Weiss, Y.; Freeman, WT. What makes a good model of natural images?. Computer Vision and Pattern Recognition; 2007; CVPR '07. IEEE Conference; 2007.
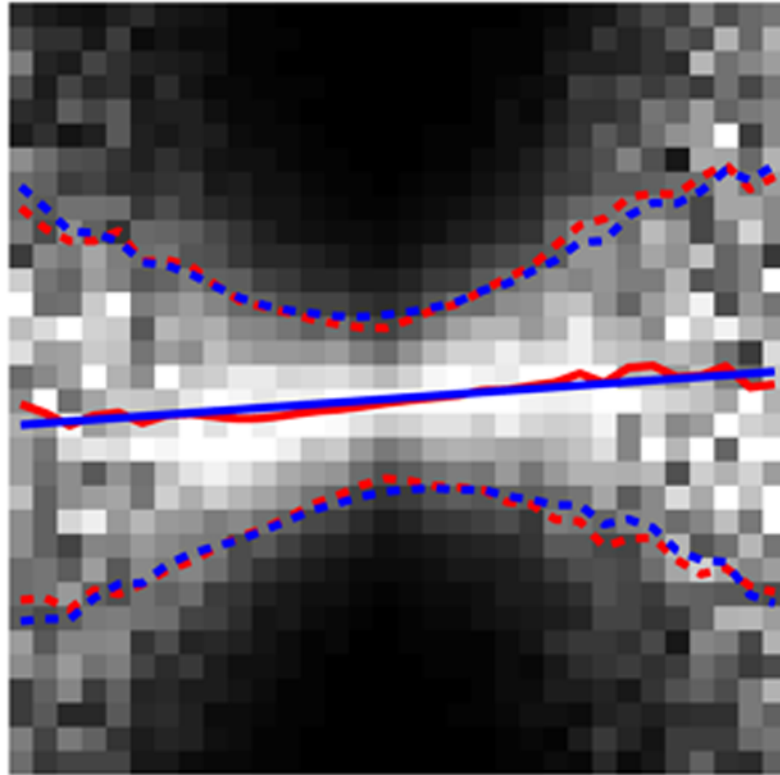
Fig. 1.
Conditional histogram of two adjacent wavelet coefficients. Grayscale intensities are proportional to probability, with larger values corresponding to brighter pixels. Each column is normalized to fill the full range of intensities. Red solid and dashed lines indicate conditional mean and standard deviation, respectively. Blue solid and dashed lines are best-fitting linear models for the conditional mean and standard deviation.
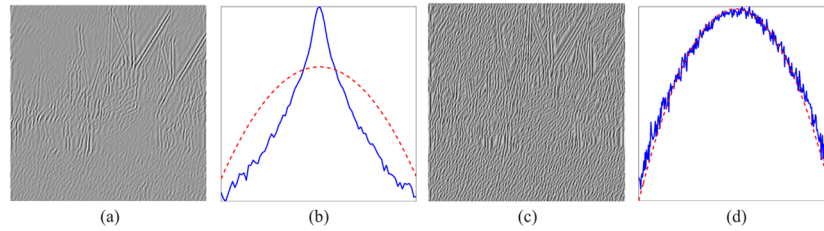
Fig. 2.
(a) A subband of the steerable pyramid decomposition of a photographic image. (b) Log histogram of subband coefficients shown in (a) (blue solid line). (c) The DNT transformation of the subband in (a). (d) Log histogram of DNT subband coefficients shown in (c) (blue solid line). In both (b) and (d) a Gaussian with the same mean and variance (red dashed line) is shown for comparison.

(a) Original   (b) SSIM = 0.51   (c) SSIM = 0.71

(d) SSIM = 0.62   (e) SSIM = 0.66   (f) SSIM = 0.75
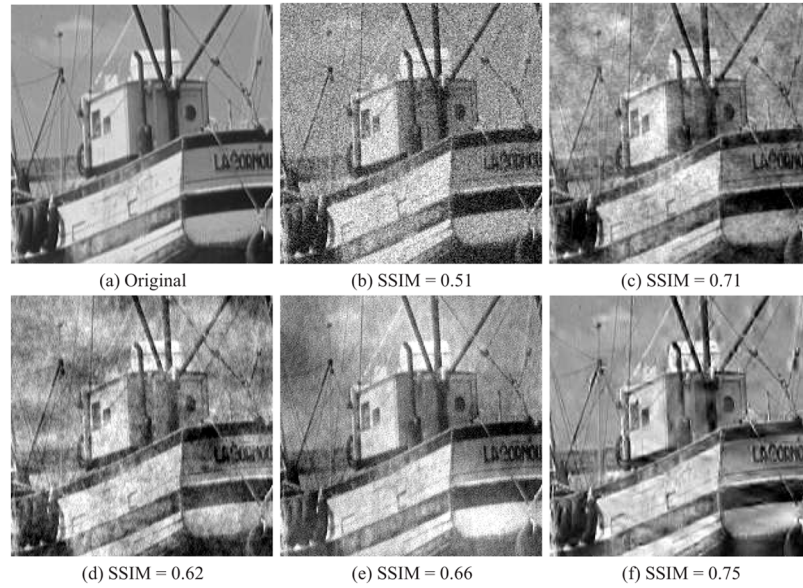
Fig. 3.
(a) A photographic image. (b)–(f) Perturbations of the image shown in (a) resulting from adding white Gaussian noise to coefficients of different image representations. All perturbed images have PSNR of 25*dB* in the pixel domain, and are shown along with SSIM scores. (b) Raw pixel, (c) Fourier domain, (d) wavelet domain, (e) steerable pyramid, (f) DNT domain.
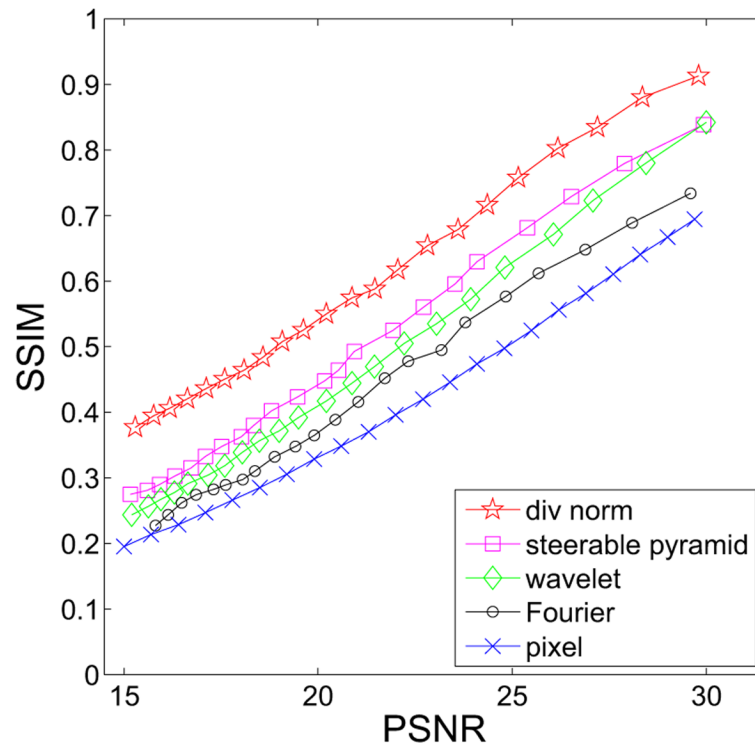
Fig. 4.
Effects of noise perturbation for different image representations. See text for details.
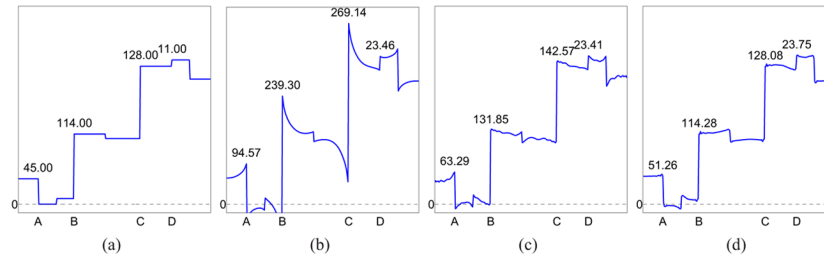
Fig. 5.
(a) A scan line of a test image of vertical edges. (b)–(d) Contrast enhancement results of the test image with different methods: (b) global high pass filtering ("unsharp masking"), (c) local gamma correction in wavelet domain, (d) global gamma correction in DNT domain. Parameters in each operation were chosen so that the smallest intensity jump in the test image (point D) would be boosted to the same value.
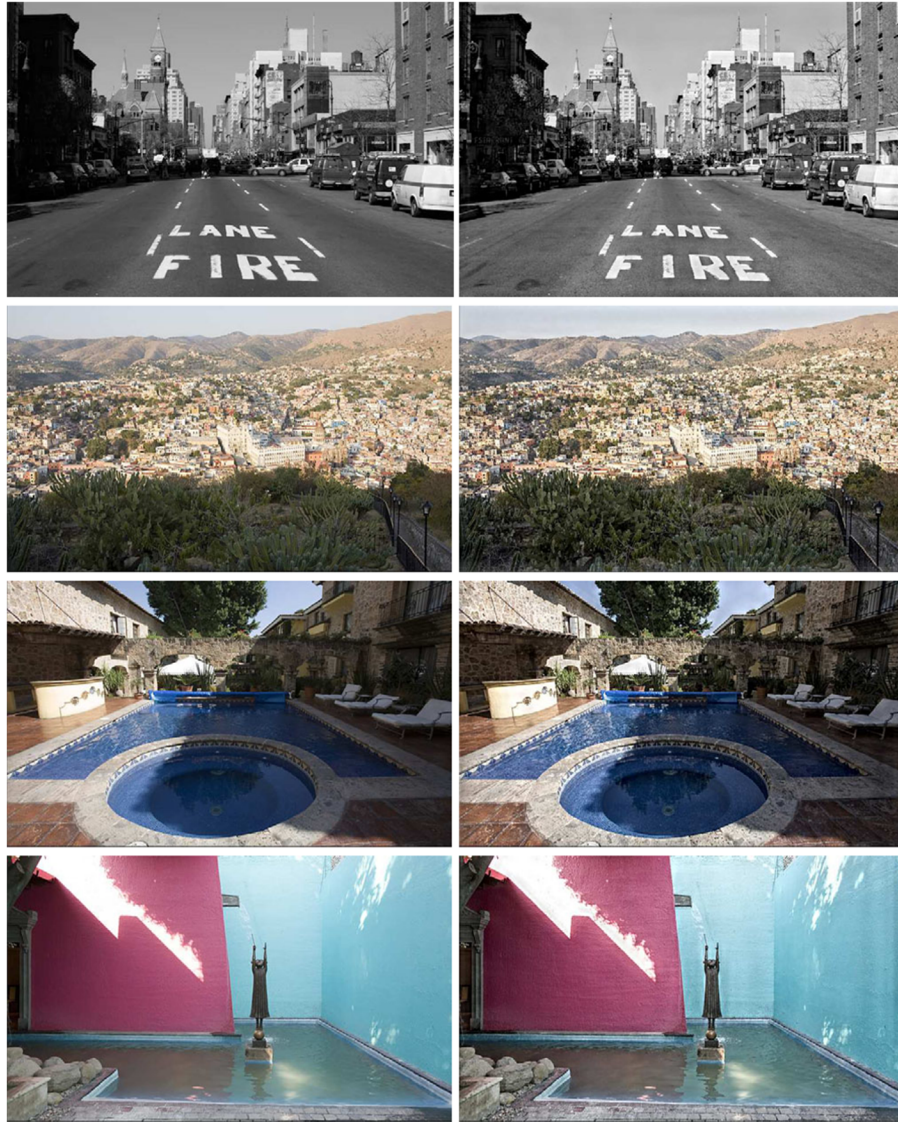
Fig. 6.
Contrast enhancement results of natural photographic images. On the left column are the original images, and on the right are the corresponding contrast enhanced images with divisive normalization representations. Images courtesy of N. Bonnier and P. Greenspun.