

# Image Denoising using Gaussian Scale Mixtures in the Wavelet Domain

Javier Portilla\*  
Universidad de Granada

Vasily Strela  
Drexel University

Martin J. Wainwright  
University of California, Berkeley

Eero P. Simoncelli  
New York University

September 29, 2002

## Abstract

We describe a method for removing noise from digital images, based on a statistical model of the coefficients of an overcomplete multi-scale oriented basis. Neighborhoods of coefficients at adjacent positions and scales are modeled as the product of two independent random variables: a Gaussian vector and a hidden positive scalar multiplier. The latter modulates the local variance of the coefficients in the neighborhood, and is thus able to account for the empirically observed correlation between the amplitudes of pyramid coefficients. Under this model, the Bayesian least squares estimate of each coefficient reduces to a weighted average of the local linear (Wiener) estimate over all possible values of the hidden multiplier variable. We demonstrate through simulations with images contaminated by additive Gaussian noise of known covariance that the performance of this method substantially surpasses that of previously published methods, both visually and in terms of mean squared error. In addition, we demonstrate the performance of the algorithm in removing sensor noise from high-ISO digital camera images.

The artifacts arising from many imaging devices are quite different from the images that they contaminate, and this difference allows humans to “see past” the artifacts to the underlying image. The goal of image restoration is to relieve human observers from this task (and perhaps even to improve upon their abilities) by reconstructing a plausible estimate of the original image from the distorted or noisy observation. A prior probability model for both the noise and for uncorrupted images is of central importance for this application.

---

\*During the development of this work, VS was on leave from Drexel University, and was supported by an AMS Centennial Fellowship. MW was supported by a NSERC-1967 Fellowship. JP and EPS were supported by an NSF CAREER grant and an Alfred P. Sloan Fellowship to EPS, and by the Howard Hughes Medical Institute. JP was also supported by an FPI fellowship from the Spanish Government.

Modeling the statistics of natural images is a challenging task, partly because of the high dimensionality of the signal. Two basic assumptions are commonly made in order to reduce dimensionality. The first is that the probability structure may be defined *locally*. Typically, one makes a Markov assumption, that the probability density of a pixel, when conditioned on a set of neighbors, is independent of the pixels beyond the neighborhood. The second is an assumption of spatial *homogeneity*: the distribution of values in a neighborhood is the same for all such neighborhoods, regardless of absolute spatial position. Although the most common model arising from these two assumptions is a Gaussian Markov random field, the restriction to second-order processes is not required, and is problematic for image modeling, where the complexity of local structures are not well described by Gaussian densities. A useful framework for capturing higher order statistics comes from augmenting a simple parametric model for local dependencies (e.g., Gaussian) with a set of “hidden” random variables that govern the parameters (e.g., variance). Such hidden Markov models have become widely used, for example, in speech processing.

The power of statistical image models can be substantially improved by transforming the signal from the pixel domain to a new representation, in which higher-order statistical properties are simplified. Over the past decade, it has become standard to begin computer-vision and image processing tasks by decomposing the image with a set of multi-scale band-pass oriented filters. This kind of representation, loosely referred to as a wavelet decomposition, is effective at decoupling the high-order statistical features of natural images. In addition, it shares some basic properties of neural responses in the primary visual cortex of mammals (e.g., [1]), and thus is adapted to efficiently represent the visually relevant features of images. A number of researchers have developed homogeneous local probability models for images in the wavelet domain. Specifically, the marginal distributions of wavelet coefficients are highly kurtotic, and can be described using suitable long-tailed distributions. Recent work has investigated the dependencies between wavelet coefficients, and found that the amplitudes of coefficients of similar position, orientation and scale are highly correlated. These dependencies may be captured using a model in which local variances are governed by a hidden random field (see [2] for a review).

In this paper, we use a model in which each neighborhood of coefficients is described as a *Gaussian scale mixture* [3]: a product of a Gaussian random vector, and an independent hidden random scalar multiplier. We have previously demonstrated that this model can account for both marginal and joint distributions amongst wavelet pyramid coefficients [4, 2], and we and others have used such models for denoising [5, 6, 7]. Here, we develop the local noise removal solution into a full Bayesian least squares estimator. We apply our solution to images corrupted by simulated additive Gaussian noise of known variance. We also show promising results in denoising digital camera images corrupted by high-ISO sensor noise.

## 1 Background: Statistical image models and denoising

Contemporary models of image statistics are rooted in the television engineering of the 1950s (see [8] for review), which relied on a characterization of the autocovariance function for

purposes of optimal signal representation and transmission. This work, and nearly all work since, assumes that image statistics are spatially homogeneous (i.e., strict-sense stationary). Another common assumption in image modeling is that the statistics are invariant, when suitably normalized, to changes in spatial scale. This assumption has been justified on the intuitive grounds that scenes are equally likely to be viewed from different distances, which corresponds roughly to a change in scale in the image domain. In addition, a large number of authors have provided empirical evidence for this assumption (see reviews mentioned above). The translation- and scale-invariance assumptions, coupled with an assumption of Gaussianity, provides the baseline model for images found throughout the engineering literature: images are described as samples of a Gaussian random field, with variance falling as  $f^{-\gamma}$  in the frequency domain.

## 1.1 Non-Gaussian image properties

In recent years, models have been developed to account for non-Gaussian behaviors of image statistics. One can see from casual observation that individual images are highly inhomogeneous: they typically contain many regions that are smooth, interspersed with “features” such as contours, or surface markings. This is reflected in the observed marginal distributions of bandpass filter responses, which show a large peak at zero, and tails that fall significantly slower than a Gaussian of the same variance [9, 10, 11] (see Fig. 1). When one seeks a linear transformation that maximizes the non-Gaussianity<sup>1</sup> of the marginal responses, the result is a basis set of oriented filters of different sizes with roughly an octave in bandwidth [e.g., 12, 13]. Due to the combination of these qualitative properties, as well as an elegant mathematical framework, multi-scale wavelet decompositions have emerged as the representation of choice for many image processing applications. The originally developed versions of these representations were orthonormal and biorthogonal. In either case, the representation is exactly complete in a linear algebraic sense: the number of coefficients is equal to the number of image pixels. But the constraint of completeness leads to clear visual artifacts in many applications (i.e., “aliasing” or “ringing”), and in the context of denoising, many authors have suggested the use of overcomplete decompositions [e.g., 14, 15, 16].

In addition to the non-Gaussian marginal behavior, the joint responses of bandpass filters exhibit important non-Gaussian statistical behavior. In particular, even when they are second-order decorrelated, wavelet coefficients corresponding to pairs of basis functions of similar position, orientation and scale exhibit striking dependencies [17, 18]. Casual observation indicates that large-amplitude coefficients are sparsely distributed throughout the image, and tend to occur in clusters. In particular, it can be seen empirically that (a) edges and lines produce significant simultaneous responses over a range of several octaves in scale; (b) corners, crosses, points and other spatially localized but non-oriented features produce simultaneous responses in more than one orientation; and (c) all of these types of feature produce significant simultaneous responses at neighboring spatial locations within their responding subbands. The conditional histograms of such pairs of coefficients indicates

---

<sup>1</sup>Different authors have used different measures of non-Gaussianity, but have obtained similar results.

that the standard deviation of a coefficient scales roughly linearly with the amplitude of nearby coefficients [18, 19, 4] (see Fig. 1).

## 1.2 Variance-adaptive models

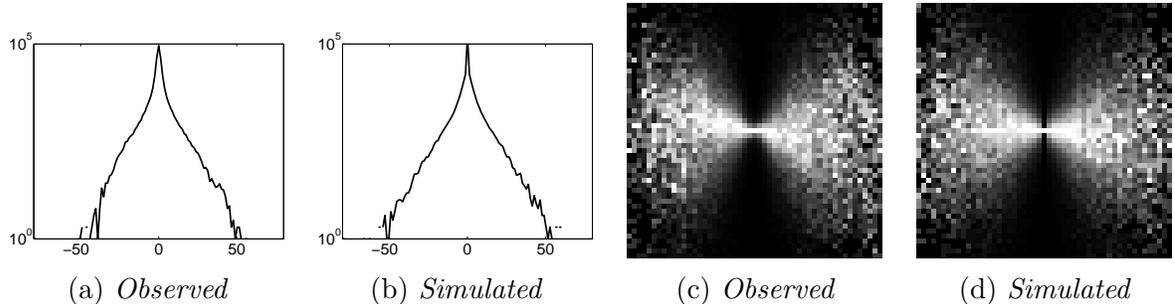
The dependency between local coefficient amplitudes, as well as the associated marginal behaviors, can be modeled using a random field with a spatially fluctuating variance. In the autoregressive conditional heteroskedastic (ARCH) models [e.g., 20], the local variance of the residual is expressed as a linear combination of the past squared innovation values and local variances. These have proven useful for many real signals that suffer from abrupt fluctuations, followed by relative “calm” periods (stock market prices, for example). Similar models have been developed for visual images, in which measure of local signal strength is estimated from a spatial neighborhood in the pixel domain [21], or from a collection of wavelet coefficients at nearby positions, scales, and orientations [18, 22, 23].

An elegant model that formalizes these characteristics is a homogeneous random field whose nodes are modulated by hidden scaling variables. A particularly useful example arises from the product of a Gaussian vector and a hidden scalar multiplier, known as a *Gaussian scale mixture* [3]. GSM distributions represent an important subset of the *elliptically symmetric distributions*, which are those that can be defined as functions of a quadratic norm of the random vector. Embedded in a random field, these kinds of models have been found useful in the speech-processing community [24]. GSM models have recently been found effective in describing visual images. For example, Baraniuk and colleagues have used a discrete hidden multiplier variable to characterize the two modes of behavior corresponding to smooth or low-contrast textured regions and features [25, 26]. A related discrete-multiplier model was developed by Spence and Parra [27]. In other work, authors have assumed that the local variance is governed by a continuous multiplier [4, 5, 6, 2]. Continuous Gaussian mixtures are able to capture the strongly leptokurtotic behavior of the marginal densities of natural image wavelet coefficients, as well as the correlation in their local amplitudes. Figure 1 shows a comparison of the empirical joint statistics of an example pair of wavelet coefficients to their simulated statistics under a local GSM model.

## 1.3 Empirical Bayes denoising using variance-adaptive models

More than 20 years ago, Lee [21] suggested a two-step procedure for image denoising, in which one first estimates the local signal variance from a neighborhood of observed pixels, and then (proceeding as if this were the true variance) applies the standard linear least squares (LLS) solution. This method is a type of *empirical Bayes* estimator [28], in that a parameter of the local model is first estimated from the data, and this estimate is subsequently used to estimate the signal.

This two-step denoising solution can be applied to any of the variance-adaptive models described in the previous section, and is substantially more powerful when applied in the wavelet domain [18, 29]. Chang et. al. used a local estimate of variance in order to set



**Fig. 1.** Comparison of coefficient statistics from an example image subband (a vertical subband of the *Boats* image) with those arising from simulation of a local GSM model. Model parameters (covariance matrix and the multiplier prior density) are estimated by maximizing the likelihood of the observed set of wavelet coefficients. (a,b) log marginal histograms. (c,d): conditional histograms of two adjacent coefficients. Brightness corresponds to probability, except that each column has been independently rescaled to fill the range of display intensities.

the level of a threshold estimator [23]. In the context of a GSM model, we have previously developed a maximum likelihood (ML) estimator [6]. Mihcak et. al. used a maximum a posteriori (MAP) estimator based on an exponential marginal prior [5], as did Li and Orchard [30], whereas Portilla et. al. used a lognormal prior [7]. Wainwright et. al. developed a tree-structured Markov model to provide a global description for the set of multiplier variables [2]. The combination of multi-scale representation and variance-adaptive modeling has provided the most effective methods for removing homogeneous additive noise from natural images to date.

Despite its success, the two-step empirical Bayes approach is suboptimal, even when the local variance estimator is optimal, because the second step does not take into account the uncertainty associated with the estimate of variance. In this paper we derive a least squares optimal single-step Bayesian estimator, which takes into account the uncertainty in the variance. We use a neighborhood of coefficients drawn from two subbands at adjacent scales, thus taking advantage of the strong statistical coupling observed through scale in multi-scale representations, and we explicitly model the covariance between neighbor coefficients, using the vectorial form of the LLS solution.

## 2 Image probability model

As a preprocessing stage for image modeling, we use a fixed multi-scale linear decomposition whose basis functions are spatially localized, oriented, and roughly one octave in bandwidth. As described in the previous section, it is now generally agreed that the use of overcomplete representations is advantageous, in order to avoid aliasing artifacts that plague critically-sampled representations such as orthogonal wavelets. A widely followed approach is to use orthogonal or or biorthogonal basis functions, but without decimating the sub-bands [e.g., 15]. Even when the the critical sampling constraint has been removed, significant improvement comes from re-designing smoother filters that achieve rotation-

invariance [14, 29, 6, 16]. For the current paper, we have used a particular variant of an overcomplete tight frame representation known as a “steerable pyramid” [14]. The primary basis functions of these transforms are polar-separable in the Fourier domain, and are related by translation, dilation, and rotation. They are spatially localized, smooth and symmetric, orientation-tuned, and one octave in bandwidth. Sampling of all subbands is above the Nyquist limit. Details of the representation are provided in Appendix A. Other authors have developed representations with similar properties [31, 32, 16].

## 2.1 Gaussian scale mixtures

Consider an image decomposed into oriented subbands at multiple scales. We denote as  $x_c^{s,o}(n,m)$  the coefficient corresponding to a linear basis function at scale  $s$ , orientation  $o$ , and centered at spatial location  $(2^s n, 2^s m)$ . We denote as  $\mathbf{x}^{s,o}(n,m)$  a *neighborhood* of coefficients clustered around this *reference coefficient*. The neighborhood may include coefficients from other subbands (i.e., corresponding to basis functions at nearby scales and orientations), as well as from the same subband. The structure for the neighborhood used to obtain our results is described in section 4. For notational simplicity, we drop the superscripts  $s, o$  and indices  $(n, m)$  in the following development.

We assume the coefficients within each local neighborhood are characterized by a Gaussian scale mixture (GSM) model. Formally, a random vector  $\mathbf{x}$  is a Gaussian scale mixture [3] if and only if it can be expressed as the product of a zero-mean Gaussian vector  $\mathbf{u}$  and an independent positive scalar random variable  $\sqrt{z}$ :

$$\mathbf{x} \stackrel{d}{=} \sqrt{z}\mathbf{u}, \quad (1)$$

where  $\stackrel{d}{=}$  indicates equality in distribution. The variable  $z$  is known as the *multiplier*. The vector  $\mathbf{x}$  is thus an infinite mixture of Gaussian vectors, whose density is determined by the covariance matrix  $\mathbf{C}_{\mathbf{u}}$  of vector  $\mathbf{u}$  and the mixing density,  $p_z(z)$ :

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}) &= \int dz p(\mathbf{x}|z) p_z(z) \\ &= \int dz \frac{\exp(-\mathbf{x}^T(z\mathbf{C}_{\mathbf{u}})^{-1}\mathbf{x}/2)}{(2\pi)^{N/2}|z\mathbf{C}_{\mathbf{u}}|^{1/2}} p_z(z), \end{aligned} \quad (2)$$

where  $N$  is the dimensionality of  $\mathbf{x}$  and  $\mathbf{u}$  (in our case, the size of the neighborhood). Without loss of generality, one can assume  $\mathbb{E}\{z\} = 1$ , which implies  $\mathbf{C}_{\mathbf{x}} = \mathbf{C}_{\mathbf{u}}$ .

The conditions under which a random vector may be represented using a GSM have been studied [3]. The GSM family includes a variety of well-known families of random variables such as the  $\alpha$ -stable family (including the Cauchy distribution), the generalized Gaussian (or stretched exponential) family and the symmetrized Gamma family [2]. GSM densities are symmetric and zero-mean, and they have leptokurtotic marginal densities (i.e., heavier tails than a Gaussian). A key property of the GSM model is that the density of  $\mathbf{x}$  is Gaussian when conditioned on  $z$ . Also, the normalized vector  $\mathbf{x}/\sqrt{z}$  is Gaussian. Some authors have suggested division by a local estimate of standard deviation as a means of “Gaussianizing” a highly kurtotic signal [8, 18].

## 2.2 GSM model for wavelet coefficients

As explained in Section 1 and illustrated in Fig. 1, a GSM model can account for both the shape of wavelet coefficient marginals and the strong correlation between the amplitudes of neighbor coefficients [4, 2]. In order to construct a global model for images from this local description, one must specify both the neighborhood structure of the coefficients, and the distribution of the multipliers. The definition of (and calculations using) the global model is considerably simplified by partitioning the coefficients into non-overlapping neighborhoods. One can then specify either a marginal model for the multipliers (treating them as independent variables) [33], or specify a joint density over the full set of multipliers [2]. Unfortunately, the use of disjoint neighborhoods leads to noticeable denoising artifacts at the discontinuities introduced by the neighborhood boundaries.

An alternative approach is to use a GSM as a local description of the behavior of the cluster of coefficients centered at each coefficient in the pyramid. Since the neighborhoods overlap, each coefficient will be a member of many neighborhoods. The local model implicitly defines a global (Markov) model, described by the conditional density of a coefficient in the cluster given its surrounding neighborhood, assuming conditional independence on the rest of the coefficients. But the structure of the resulting model is such that performing statistical inference (i.e., computing Bayes estimates) in an exact way is intractable. In this paper, we simply solve the estimation problem for the reference coefficient at the center of each neighborhood independently. The choice of neighborhood is described in section 4.

## 2.3 Prior density for multiplier

To complete the model, we need to specify the probability density,  $p_z(z)$ , of the multiplier. Several authors have suggested the generalized Gaussian (stretched exponential) family of densities as an appropriate description of wavelet coefficient marginal densities [11, 34, 35]:  $p_x(x) \propto \exp\left(-\left|\frac{x}{s}\right|^p\right)$ , where the scaling variable  $s$  controls the width of the distribution, and the exponent  $p$  controls the shape (in particular, the heaviness of the tails), and is typically estimated to lie in the range [0.5, 0.8] for image subbands. Although they can be expressed as GSMs, the density of the associated multiplier has no closed form expression, and thus this solution is difficult to implement.

In previous work [7], we noted that the density of the log coefficient magnitude,  $\log|x|$ , may be written as a convolution of the densities of  $\log|u|$  and  $\log\sqrt{z}$ . Since the density of  $u$  is known, this means that estimation of the density of  $\log\sqrt{z}$  may be framed as a deconvolution problem. The resulting estimated density may be approximated by a Gaussian, and thus we proposed a lognormal prior for the  $z$ . This solution has two important drawbacks. First, it assumes all coefficients in a neighborhood have the same marginal statistics, and thus they must typically all belong to the same subband. Second, it is derived in terms of the noise-free coefficients, and it is difficult to extend it for use in the noisy case.

We have also investigated a more direct maximum likelihood approach for estimating a

nonparametric  $p_z(z)$  from an observed set of neighborhood vectors:

$$\hat{p}_z(z) = \arg \max_{p_z(z)} \sum_{m=1}^M \log \left( \int_0^\infty dz p(\mathbf{x}_m|z)p_z(z) \right), \quad (3)$$

where the sum is over the neighborhoods. Note that the estimate,  $\hat{p}_z(z)$ , must be constrained to positive values, and must have unit area. We have developed an efficient algorithm for computing this solution numerically. One advantage of the ML solution is that it is easily extended for use with the noisy observations, by replacing  $\mathbf{x}_m$  the noisy observation.

A fourth choice is a so-called *noninformative prior* [36], which has the advantage that it does not require the fitting of any parameters to the noisy observation. We have examined the most widely used solution, known as *Jeffrey's prior* (see Ref. [36]). In the context of estimating the multiplier  $z$  from coefficients  $\mathbf{x}$ , this takes the form:

$$p_z(z) \propto \sqrt{I(z)}, \quad I(z) = \mathbb{E} \left\{ -\frac{\partial^2 \log p(\mathbf{x}|z)}{\partial z^2} \right\}$$

where  $I(z)$  is the Fisher information matrix. Computing this for the GSM model is straightforward:

$$\begin{aligned} -\frac{\partial^2 \log p(\mathbf{x}|z)}{\partial z^2} &= \frac{\partial^2}{\partial z^2} \left[ N \log(z) + \log |\mathbf{C}_\mathbf{u}| + \frac{\mathbf{x}^T \mathbf{C}_\mathbf{u}^{-1} \mathbf{x}}{z} \right] \\ &= \frac{N}{z^2} + \frac{\mathbf{x}^T \mathbf{C}_\mathbf{u}^{-1} \mathbf{x}}{z^3}. \end{aligned}$$

Taking the square root of the expectation, and using the fact that  $\mathbb{E}\{\mathbf{x}^T \mathbf{C}_\mathbf{u}^{-1} \mathbf{x}\} = z$  we obtain Jeffrey's prior:

$$p_z(z) \propto \frac{1}{z}, \quad (4)$$

which corresponds to a constant prior on  $\log(z)$ . Note that this prior does not constitute a legitimate probability density, as its integral does not converge. Nevertheless, it is common to ignore this fact and use such a solution as a pseudo-prior, whenever the integrals involved in the estimation converge. In our case, to ensure the existence of the posterior  $p(z|\mathbf{y})$  of (13), we have set the pseudo-prior to zero in the interval  $[0, z_{\min})$ , where  $z_{\min}$  is a very small positive constant (see section 4 for details).

Of the alternatives described above, we have found (as expected) that the ML-estimated nonparametric prior produces the best results for denoising the pyramid coefficients. Note, however, that a least squares optimal estimate for the pyramid coefficients does *not* necessarily lead to a least-squares optimal estimate for the image pixels, since the pyramid representation is overcomplete. We were surprised to find that the noninformative prior typically leads to better denoising performance in the image domain (roughly +0.15 dB, on average). Given that it is also simpler and more efficient to implement, we have used it for all of the results shown in the following sections.

### 3 Image denoising

Our procedure for image denoising uses the same top-level structure as most previously published approaches: (1) decompose the image into pyramid subbands at different scales and orientations; (2) denoise each subband; and (3) invert the pyramid transform, obtaining the denoised image. We assume the image is corrupted by independent additive Gaussian noise of known covariance. A vector  $\mathbf{y}$  corresponding to a neighborhood of  $N$  observed coefficients of the pyramid representation can be expressed as:

$$\mathbf{y} = \mathbf{x} + \mathbf{w} = \sqrt{z}\mathbf{u} + \mathbf{w}. \quad (5)$$

Note that the assumed GSM structure of the coefficients, coupled with the assumption of independent additive Gaussian noise, means that the three random variables on the right side of (5) are independent.

Both  $\mathbf{u}$  and  $\mathbf{w}$  are zero-mean Gaussian vectors, with associated covariance matrices  $\mathbf{C}_{\mathbf{u}}$  and  $\mathbf{C}_{\mathbf{w}}$ . The density of the observed neighborhood vector conditioned on  $z$  is a zero-mean Gaussian, with covariance  $\mathbf{C}_{\mathbf{y}|z} = z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}$ :

$$p(\mathbf{y}|z) = \frac{\exp(-\mathbf{y}^T(z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}})^{-1}\mathbf{y}/2)}{\sqrt{(2\pi)^N |z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}|}}. \quad (6)$$

We assume the noise covariance is known apriori in the image domain. Since  $\mathbf{w}$  is derived from the image through the (linear) pyramid transformation, it is straightforward to compute the noise covariance matrix  $\mathbf{C}_{\mathbf{w}}$ . Given  $\mathbf{C}_{\mathbf{w}}$ , the signal covariance  $\mathbf{C}_{\mathbf{u}}$  can be computed from the observation covariance matrix  $\mathbf{C}_{\mathbf{y}}$ . Taking the expectation of  $\mathbf{C}_{\mathbf{y}|z}$  over  $z$  yields:

$$\mathbf{C}_{\mathbf{y}} = \mathbb{E}\{z\}\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}},$$

We are free to choose  $\mathbb{E}\{z\}$ , so we set it to one, resulting in:

$$\mathbf{C}_{\mathbf{u}} = \mathbf{C}_{\mathbf{y}} - \mathbf{C}_{\mathbf{w}}. \quad (7)$$

We ensure that  $\mathbf{C}_{\mathbf{u}}$  is positive semidefinite by performing an eigenvector decomposition and setting any negative eigenvalues to zero.

#### 3.1 Bayes least squares estimator

For each neighborhood, we wish to estimate  $x_c$ , the reference coefficient at the center of the neighborhood, from  $\mathbf{y}$ , the set of observed (noisy) coefficients. The Bayes least squares (BLS) estimate is just the conditional mean:

$$\begin{aligned} \mathbb{E}\{x_c|\mathbf{y}\} &= \int dx_c x_c p(x_c|\mathbf{y}) \\ &= \int dx_c \int_0^\infty dz x_c p(x_c, z|\mathbf{y}) \\ &= \int dx_c \int_0^\infty dz x_c p(x_c|\mathbf{y}, z) p(z|\mathbf{y}) \\ &= \int_0^\infty dz p(z|\mathbf{y}) \mathbb{E}\{x_c|\mathbf{y}, z\}, \end{aligned} \quad (8)$$

where we have assumed uniform convergence in order to exchange the order of integration. Thus, the solution is written as the average of the Bayes least squares estimate of  $x$  when conditioned on  $z$ , weighted by the posterior density,  $p(z|\mathbf{y})$ . We now describe each of these individual components.

### 3.2 Local Wiener estimate

The advantage of the GSM model is that the coefficient neighborhood vector  $\mathbf{x}$  is Gaussian when conditioned on  $z$ . This fact, coupled with the assumption of additive Gaussian noise means that the expected value inside the integral of (8) is simply a local linear (Wiener) estimate. Writing this for the full neighborhood vector:

$$\mathbb{E}\{\mathbf{x}|\mathbf{y}, z\} = z\mathbf{C}_{\mathbf{u}}(z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}})^{-1}\mathbf{y}, \quad (9)$$

We can simplify the dependence of this expression on  $z$  by diagonalizing the matrix  $z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}$ . Specifically, let  $\mathbf{S}$  be the symmetric square root of the positive definite matrix  $\mathbf{C}_{\mathbf{w}}$  (i.e.,  $\mathbf{C}_{\mathbf{w}} = \mathbf{S}\mathbf{S}^T$ ), and let  $\{\mathbf{Q}, \mathbf{\Lambda}\}$  be the eigenvector/eigenvalue expansion of the matrix  $\mathbf{S}^{-1}\mathbf{C}_{\mathbf{u}}\mathbf{S}^{-T}$ . Then:

$$\begin{aligned} z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}} &= z\mathbf{C}_{\mathbf{u}} + \mathbf{S}\mathbf{S}^T \\ &= \mathbf{S}(z\mathbf{S}^{-1}\mathbf{C}_{\mathbf{u}}\mathbf{S}^{-T} + \mathbf{I})\mathbf{S}^T \\ &= \mathbf{S}\mathbf{Q}(z\mathbf{\Lambda} + \mathbf{I})\mathbf{Q}^T\mathbf{S}^T. \end{aligned} \quad (10)$$

Note this this diagonalization does not depend on  $z$ , and thus only needs to be computed once for each subband. We can now simplify (9):

$$\begin{aligned} \mathbb{E}\{\mathbf{x}|\mathbf{y}, z\} &= z\mathbf{C}_{\mathbf{u}}\mathbf{S}^{-T}\mathbf{Q}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{Q}^T\mathbf{S}^{-1}\mathbf{y} \\ &= z\mathbf{S}\mathbf{S}^{-1}\mathbf{C}_{\mathbf{u}}\mathbf{S}^{-T}\mathbf{Q}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{Q}^T\mathbf{S}^{-1}\mathbf{y} \\ &= z\mathbf{S}\mathbf{Q}\mathbf{\Lambda}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{Q}^T\mathbf{S}^{-1}\mathbf{y} \\ &= z\mathbf{M}\mathbf{\Lambda}(z\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{v}, \end{aligned} \quad (11)$$

where  $\mathbf{M} = \mathbf{S}\mathbf{Q}$ , and  $\mathbf{v} = \mathbf{M}^{-1}\mathbf{y}$ . Finally, we restrict the estimate to the reference coefficient, as needed for the solution of (8):

$$\mathbb{E}\{x_c|\mathbf{y}, z\} = \sum_{n=1}^N \frac{zm_{cn}\lambda_n v_n}{z\lambda_n + 1}, \quad (12)$$

where  $m_{ij}$  represents an element ( $i$ -th row,  $j$ -th column) of the matrix  $\mathbf{M}$ ,  $\lambda_n$  are the diagonal elements of  $\mathbf{\Lambda}$ ,  $v_n$  the elements of  $\mathbf{v}$ , and  $c$  is the index of the reference coefficient within the neighborhood vector.

### 3.3 Posterior distribution of the multiplier

The other component of the solution given in (8) is the distribution of the multiplier, conditioned on the observed neighborhood values. We use Bayes rule to compute this:

$$p(z|\mathbf{y}) = \frac{p(\mathbf{y}|z)p_z(z)}{\int_0^\infty d\alpha p(\mathbf{y}|\alpha)p_z(\alpha)}. \quad (13)$$

As discussed in section 2.3, we choose a noninformative Jeffrey’s prior, corrected at the origin, for the function  $p_z(z)$ . The conditional density  $p(\mathbf{y}|z)$  is given in (6), and its computation may be simplified using the relationship in (10) and the definition of  $\mathbf{v}$ :

$$p(\mathbf{y}|z) = \frac{\exp(-\frac{1}{2} \sum_{n=1}^N \frac{v_n^2}{z\lambda_n+1})}{\sqrt{(2\pi)^N |\mathbf{C}_w| \prod_{n=1}^N (z\lambda_n + 1)}}. \quad (14)$$

Summarizing our denoising algorithm:

1. Decompose the image into subbands
2. For each subband (except the lowpass residual):
  - (a) Compute neighborhood noise covariance,  $\mathbf{C}_w$
  - (b) Estimate noisy neighborhood covariance,  $\mathbf{C}_y$
  - (c) Estimate  $\mathbf{C}_u$  from  $\mathbf{C}_w$  and  $\mathbf{C}_y$  using (7)
  - (d) For each neighborhood:
    - i. For each value  $z$  in the integration range:
      - A. Compute  $\mathbb{E}\{x_c|\mathbf{y}, z\}$  using (12)
      - B. Compute  $p(\mathbf{y}|z)$  using (14)
      - C. Compute  $p(z|\mathbf{y})$  using (13) and (4)
    - ii. Compute  $\mathbb{E}\{x_c|\mathbf{y}\}$  numerically using (8)
3. Reconstruct the denoised image from the processed subbands and the lowpass residual

## 4 Implementation

We decompose the image into subbands using a specialized variant of the steerable pyramid [14]. The representation consists of oriented bandpass bands at 8 orientations and 5 scales, 8 oriented highpass residual subbands, and one lowpass (non-oriented) residual band, for a total of 49 subbands. A detailed description of the decomposition is given in Appendix A.

We have hand-optimized the neighborhood structure (i.e., choice of spatial positions, scales and orientations). A  $3 \times 3$  region surrounding the reference coefficient, together with the coefficient at the same location and orientation at the next coarser scale (the *parent*), maximizes the denoising performance, on average. We denote this generalized neighborhood as  $3 \times 3 + p$ . In previous work on compression, inclusion of the parent coefficient was also found to provide a significant improvement in performance [19]. Other authors have also incorporated scale-to-scale interactions [e.g., 37, 25, 26]. When denoising the subbands at the coarsest scale, which have no parent subband, we simply use the  $3 \times 3$  spatial neighborhood. Note that the actual spatial extent of the neighborhood depends on the scale of the subband – the basis functions grow in size as  $2^s$  – as is appropriate under the

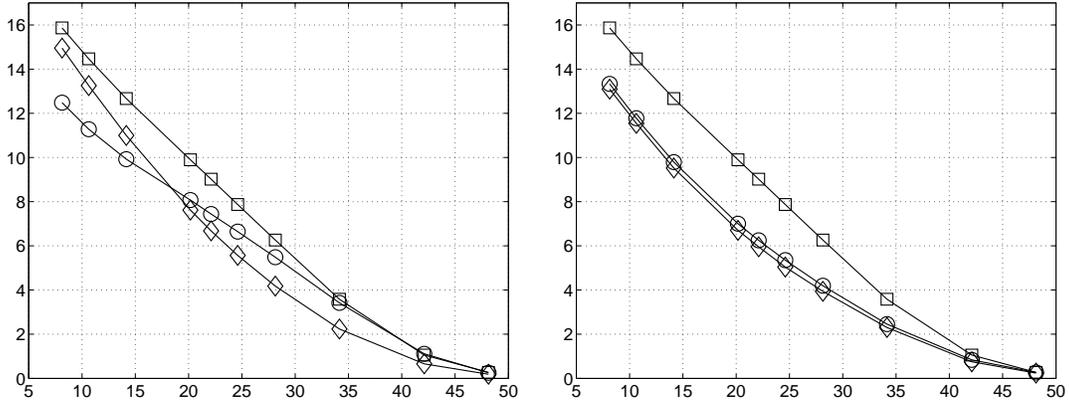
assumption that image statistics are scale-invariant [38, 39]. Note also that since the *parent* subband is sampled at half the density of the reference subband, it must be upsampled and interpolated in order to obtain values for neighborhoods at every choice of reference coefficient.

In our implementation, the integral of (8) is computed numerically. This requires us to choose the (finite) range and sample spacing over which the calculations are performed, which must be a compromise between accuracy and computational cost. The typical shape of  $p(\mathbf{y}|z)$  (considered as a function of  $z$  for each observed  $\mathbf{y}$ ) extends very far and smoothly as  $z$  grows towards infinity, but can exhibit rapid variations as it approaches zero. As such, we have sampled  $z$  with logarithmically uniform spacing. This adequately captures the shape of the function, in terms of the performance of the algorithm, with only  $S_z = 11$  samples. These samples are spread uniformly over a range extending from  $\mu_l - 13\sigma_l$  to  $\mu_l + 9\sigma_l$ , where  $\{\mu_l, \sigma_l\}$  are estimates of the mean and standard deviation of  $\log(z)$ . These are obtained from the variance and kurtosis of each subband using the method of moments described in [7].

The computational cost of the pyramid transform (both forward and inverse) scales as  $I_x I_y \log_2(I_x I_y)$ , where  $(I_y, I_x)$  are the dimensions of the image. The computational cost of the estimation procedure scales as  $\left(I_x + \frac{(N_x + B_x)}{2}\right) \left(I_y + \frac{(N_y + B_y)}{2}\right) N K S_z$ , where  $N_{x,y}$  are the dimensions of the spatial subband neighborhood (3 in our case),  $B_{x,y}$  the dimensions of the bandpass convolution kernels (roughly 9 in our implementation),  $N$  the full size of the neighborhood (10 in our case),  $K$  the number of orientations, and  $S_z$  the number of samples used for the distributions over  $z$ . The terms added to the image dimensions correspond to the padded boundary region that must be estimated in order to properly reconstruct the image. As a guide, running times in our current unoptimized Matlab implementation, on a Linux workstation with 1.7 GHz Intel Pentium-III CPU, are roughly 40 seconds for  $256 \times 256$  images. Finally, the primary memory cost is due to storage of the pyramid coefficients (roughly  $7KN_xN_y/3$  floating point numbers).

## 5 Results

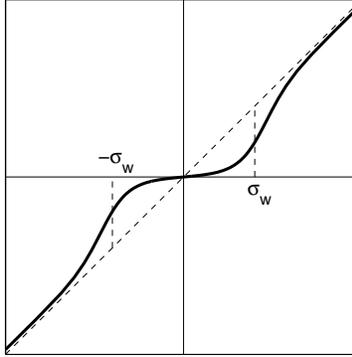
We have tested our method on a set of 8-bit grayscale test images, each contaminated with computer-generated additive Gaussian white noise at 10 different variances. Further information about the images is provided in Appendix B. Table 1 shows error variances of the denoised images, expressed as peak signal-to-noise ratios (PSNR) in decibels, for the full range of input noise levels. Note that for all images, there is very little improvement at the lowest noise level. This makes sense, since the “clean” images in fact include quantization errors and thus have an implicit PSNR of 58.9dB. At the other extreme, improvement is substantial (roughly 17dB in the best cases). The reduction in noise (in dB), averaged over the first three images, is plotted in Fig. 2, as a function of the noise level.



**Fig. 2.** Comparison of our denoising method (squares) to other methods. Curves depict the increase in PSNR as a function of input PSNR, averaged over three representative images (*Lena*, *Barbara* and *Boats*). **Left:** Comparison to two restricted models: A Gaussian model in which the local covariance is estimated for each subband in the pyramid domain (diamonds), and a GSM model with a neighborhood of size one (circles). **Right:** Comparison to a soft-thresholding method in an undecimated wavelet decomposition (*circles*) [40] and a local variance-adaptive linear method in the image domain (*diamonds*) [21]. Parameters for both methods are hand-optimized for each noisy image: an optimal threshold level for the first method, and an optimal neighborhood size (ranging from 3 for low noise levels to 11 for high noise levels) for the second.

$\sigma$ / PSNR	<i>Lena</i>	<i>Barb</i>	<i>Boats</i>	<i>Fgrpt</i>	<i>House</i>	<i>Peprs</i>
1 / 48.13	48.45	48.33	48.43	48.47	48.83	48.37
2 / 42.11	43.24	43.26	43.00	43.06	44.06	43.00
5 / 34.15	38.46	37.75	36.99	36.70	38.67	37.30
10 / 28.13	35.60	33.99	33.59	32.48	35.36	33.75
15 / 24.61	33.92	31.83	31.71	30.19	33.63	31.75
20 / 22.11	32.69	30.31	30.39	28.63	32.39	30.34
25 / 20.17	31.71	29.13	29.37	27.48	31.41	29.26
50 / 14.15	28.61	25.50	26.36	24.16	28.26	25.96
75 / 10.63	26.85	23.66	24.79	22.41	26.41	24.02
100 / 8.13	25.64	22.63	23.74	21.25	25.08	22.65

**Table 1.** Denoising performance, expressed as peak Signal-to-Noise Ratio,  $20 \log_{10}(255/\sigma_e)$  in dB, where  $\sigma_e$  is the standard deviation of the error.



**Fig. 3.** Nonlinear pointwise estimation function resulting from restriction of our method to a neighborhood of size one.

### 5.1 Comparison to restricted versions

In order to understand the relative contribution of various aspects of our method, we considered two restricted versions of our model that are representative of the two primary denoising concepts found in the literature. The first is a Gaussian model, arising from the restriction of our model to a prior density  $p_z(z)$  which is a delta function concentrated at the signal variance. This model is globally Gaussian, but the signal covariance is modeled only locally (over the extent of the neighborhood) for each pyramid subband. As such, this denoising solution may be viewed as a regularized version of the classical linear (Wiener filter) solution. In order to implement this, we simply estimate each coefficient using (12), with  $z$  set to one.

The second restricted form of our model uses a neighborhood containing only the reference coefficient (i.e.,  $1 \times 1$ ). Under these conditions, the model describes only the marginal density of the coefficients, and the estimator reduces to application of a scalar function to the observed noisy coefficients. Pointwise estimators have become quite popular in the image denoising literature, and typically are chosen to perform a type of thresholding operation, suppressing low-amplitude values while retaining high-amplitude values. The concept was developed originally in the television engineering literature (where it is known as “cor-ing” [e.g., 41]), and specific functions have been derived under a variety of formulations, including minimax optimality under a smoothness condition [40, 42], and Bayesian estimation under non-Gaussian priors [34, 43, 29, 35]. The function resulting from the reduction of our model to a single-element neighborhood is shown in Fig. 3, and is a smooth threshold operation, similar to those derived in [29] for a generalized Gaussian prior.

Figure 2 shows a comparison of our full model and these two reduced forms. Note that the pointwise (soft-thresholding) solution outperforms the jointly Gaussian solution at high PSNR (low noise) levels, but the opposite is true for high noise levels. The full model incorporates the advantages of the two subcases, and thus outperforms both of them.

We have also examined the relative importance of other features of our model. Table 2 shows the decrease in PSNR that results when each feature is removed. The first column

$\sigma$ / PSNR	<i>Prnt</i>	<i>Ori8</i>	<i>BLS</i>	<i>Cov</i>	<i>Bdr</i>
10 / 28.13	-0.02	0.18	0.12	0.41	0.12
25 / 20.17	0.05	0.28	0.23	1.04	0.15
50 / 14.15	0.10	0.28	0.32	0.97	0.15

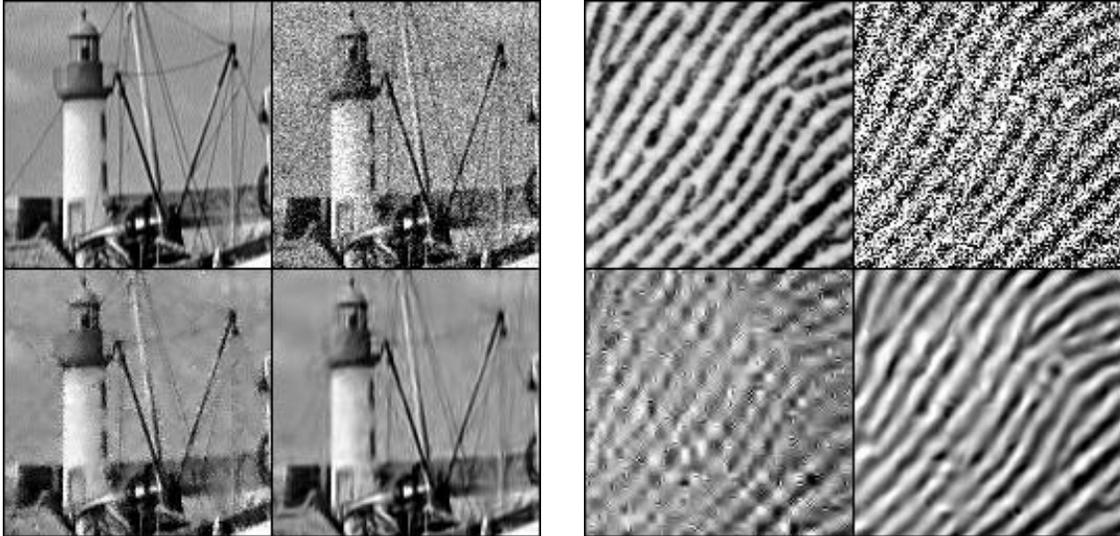
**Table 2.** Reduction in denoising performance (dB) resulting from removal of model components, shown at 3 different noise contamination rates. Results are averaged over *Lena*, *Barbara* and *Boats*. See text for further information.

shows that removal of the coarse-scale parent from the neighborhood reduces performance significantly only at high noise levels. The second column shows that reducing the number of orientation bands from  $K = 8$  to  $K = 4$  leads to a significant reduction in performance. The next column (*BLS*) demonstrates that the full BLS estimator gives significantly better performance than a two-step estimator (a MAP estimate of the local multiplier, followed by a linear estimate of the coefficient, as in [7]). The next column (*Cov*) shows importance of modeling the full covariance of both noise and signal, as opposed to just the variance. This is due partly to the high correlation amongst coefficients in the overcomplete representation, and also to the adaptation to specific spectral features of each image. Finally, the use of mirror extension (as opposed to periodic boundary handling) leads to a surprisingly large improvement.

## 5.2 Comparison to other approaches

We have compared our method to two well-known and widely available denoising algorithms: a local variance-adaptive linear method in the pixel domain [21], as implemented by the MATLAB function *wiener2*, and a soft thresholding method [40], using an undecimated pyramid based on the Daubechies 4-tap wavelet filter. In both cases, a single parameter (the neighborhood size or the threshold) was hand-optimized to maximize performance. Results are shown in Fig. 2. Our method is seen to outperform the other two over the entire range of noise levels, and by roughly 3dB over the range from 10-25dB that is commonly studied. Figure 4 provides a visual comparison of example images denoised using two of these algorithms. Our method produces artifacts that are significantly less visible, and at the same time is able to better preserve the features of the original image.

Finally, we have compared our method to some of the best available published results [22, 23, 5, 30, 44, 45], and these are shown in Fig. 5. Since there are many slightly different versions of these images available on the internet, we have either verified with the authors that we are using the same images (Refs. [30, 44, 45]), or have simply replotted comparisons made by the referred authors (Refs. [22, 23, 5]). The improvement of our method with respect to others is apparent. Figure 6 provides a visual comparison of an example image (*Barbara*) denoised using the algorithm of Li *et al.* [30], which is based on a variance-adaptive model in an overcomplete separable wavelet representation. Our method is seen to provide less artifacts as well as better preservation of edges and other details. Note, however, that the noisy images were created using different samples of noise, and thus the artifacts in the two images appear at different locations. The separation of diagonal orientations in the



**Fig. 4.** Comparison of denoising results on two images (cropped to  $128 \times 128$  for visibility of the artifacts). **Left:** *Boats* image. Top-left: original image. Top-right: noisy image, PSNR = 22.11dB ( $\sigma = 20$ ). Bottom-left: denoising result using adaptive local Wiener in the image domain [21], PSNR = 28.0dB. Bottom-right: our method, PSNR = 30.4dB. **Right:** *Fingerprint* image. Top-left: original image. Top-right: noisy image, PSNR = 8.1dB ( $\sigma = 100$ ). Bottom-left: denoising result using soft thresholding [40] with a single hand-optimized threshold, PSNR = 18.0dB. Bottom-right: our method, PSNR = 21.2dB.

steerable pyramid allows more selective removal of the noise in diagonally oriented image regions (see parallel diagonal lines on the left side of the face).

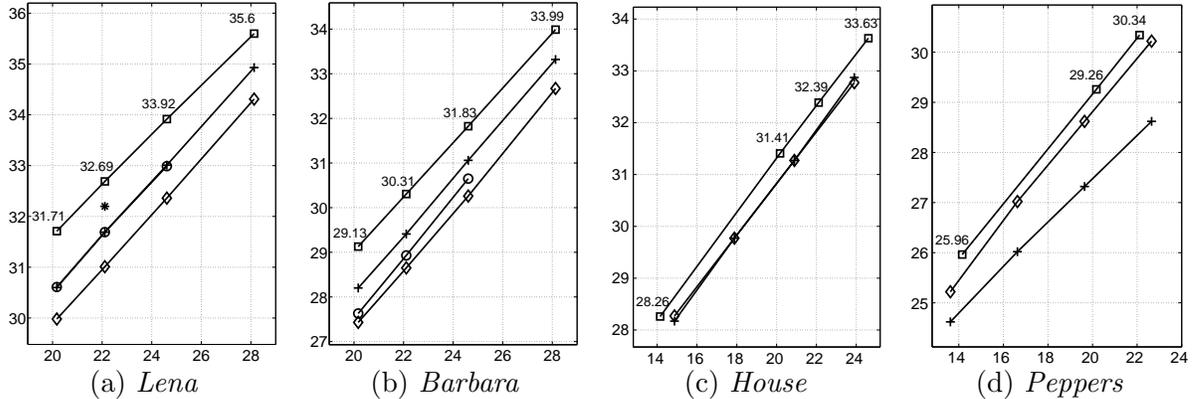
### 5.3 Denoising digital camera images

As a more realistic test of our approach, we have obtained images from a Canon G1 consumer digital camera, which uses a  $2160 \times 1440$  CCD sensor quantized to 10 bits. We studied and calibrated the properties of the sensor noise, as described in Appendix C. We found two important deviations from the white noise model assumed in section 5. First, the noise is highly correlated, both spatially and across color channels. We found that attempting to incorporate the correlations across color channels did not significantly improve denoising performance, and thus the results shown in this section are obtained by considering each channel in isolation. The second deviation from the white model is that the noise is strongly dependent on the signal (see Fig. 7(c)).

In order to capture the signal dependence, we assumed a model of Gaussian noise, in which the variance in each pixel is a deterministic function of the true (noise-free) image intensity at that pixel. We estimated this function through calibration measurements (see Appendix C for details). In the subband domain, we assumed the following noise model:

$$\mathbf{y} = \sqrt{z}\mathbf{u} + \sqrt{\alpha_x}\mathbf{w}.$$

where  $\alpha_x$  is a secondary multiplier that controls the local noise variance, with  $\mathbb{E}\{\alpha_x\} = 1$ . We estimate this noise multiplier as an average of the noise variance at each pixel



**Fig. 5.** Comparison of denoising performance of several recently published methods. Curves depict output PSNR as a function of input PSNR. Square symbols indicate our results, taken from Table 1. (a,b): diamonds [5]; circles [23]; crosses [30]; asterisk [45]. (c,d): crosses [22]; diamonds [44].

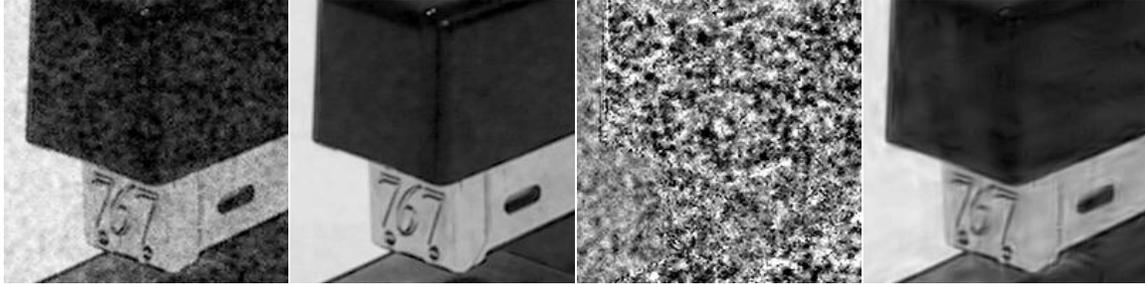


**Fig. 6.** Comparison of denoising results on *Barbara* image (cropped to  $150 \times 150$  for visibility of the artifacts). Left to right: Original image; Noisy image ( $\sigma = 25$ , PSNR = 20.2dB); Results of Li *et al.* [30] (PSNR = 28.2dB); Our method (PSNR = 29.1dB).

over a spatial neighborhood roughly equal in size to the associated basis function for each coefficient. The pixel variances are estimated by applying the calibration curve described in Appendix C, to a rough estimate of the clean pixels (obtained by simple lowpass filtering). Once the values  $\alpha_x$  have been computed, the estimation of coefficients proceeds as before, except that the terms  $(z\lambda_n + 1)$  in (14) and (12) must be replaced with  $(z\lambda_n + \alpha_x)$ .

Under laboratory conditions, we can obtain a good approximation of a noiseless reference image corresponding to a given noisy one, thus allowing us to quantify the denoising performance. We photograph a fixed rigid scene at two different ISO settings (specifically, 50 and 400), compensating for this change in sensitivity by altering the exposure time. The low-ISO image has significantly less noise, but slightly different contrast, and is spatially shifted by a small fraction of a pixel (a small movement of the camera between exposures was unavoidable). To compensate for these changes, we use a standard least squares method to fit the pixel values of the noisy image with a linear combination of the three color channels and a  $3 \times 3$  spatial neighborhood from the clean picture. We use the result as a reference image for the noisy observation.

Figure 7 shows the blue channels of an example of the noisy (ISO 400) picture, the reference (ISO 50) image, the noise (difference, amplified for visibility), and the result of applying



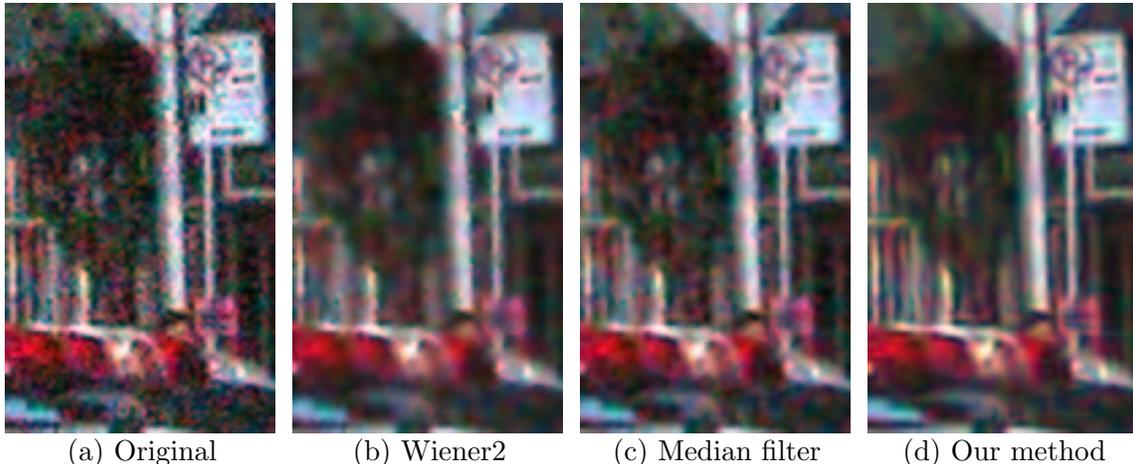
**Fig. 7.** Denoising on cropped blue channel. Left to right: Noisy (400 ISO) digital photograph; Reference (50 ISO) image; Noise (difference between the two); Denoising result. See text for details.



**Fig. 8.** Denoising a digital camera photograph. **Left:** original image, shot at ISO=400 on a Canon PowerShot G1, cropped to  $512 \times 512$  pixels. Estimated noise level is approximately 27.9dB. **Right:** denoised image, using the spatially variant noise model. Noise reduction is approximately 6dB.

our modified spatially variant algorithm. Note the strong noise spatial covariance and its dependency on the signal intensity. Averaging the squared error in the 3 color channels, we have obtained an input PSNR of 28.10 dB, and an output PSNR of 34.06 dB. Using a spatially adaptive method in the space domain [21] (Matlab function *wiener2*) leads to a loss of 2.20 dB in performance. This noise reduction is consistent with the results shown in the graph of Fig. 2.

Figure 8 shows denoising results for a 400 ISO digital photograph taken on the street with daylight illumination. In Fig. 9, we show cropped details of this image, and compare our algorithm to two widely available algorithms: Matlab's *wiener2* function, and Adobe Photoshop's median filter. Parameters for each of these were hand-optimized (see figure caption). Visually, our algorithm achieves a much better compromise between preservation of details and removal of noise. Even so, a noticeable amount of noise remains, especially in the low frequencies, producing local oscillations in the chrominance of the depicted objects.



**Fig. 9.** Denoising a digital camera photograph. (a) Original, cropped from Fig. 8; (b) Denoised using Matlab’s *wiener2* function, with a  $5 \times 5$  neighborhood [21]; (c) Denoised using median filter in Adobe Photoshop, with radius=1; (d) Our method.

## 6 Conclusions

We have presented a denoising method based on a local Gaussian Scale Mixture model in an overcomplete wavelet representation. Our statistical model differs from previous approaches in a number of important ways. First, many previous models have used either separable orthogonal wavelets, or redundant versions of such wavelets. In contrast, our model is based on an overcomplete tight frame that is free from aliasing, with basis functions that are selective for orientation. The increased redundancy of the representation results in improved performance, and we believe that the smoothness and symmetry of the kernels also contribute. Second, we have explicitly modeled the covariance between neighboring coefficients (for both signal and noise), as opposed to considering only marginal responses or local variance. Thus, the model captures correlations induced by the overcomplete representation as well as correlations inherent in the underlying image, and it can handle Gaussian noise of arbitrary power spectral density. Third, we have included a neighbor from the same orientation and spatial location at a coarser scale (a *parent*), as opposed to considering only spatial neighbors within each subband. This modeling choice is consistent with the empirical findings of strong statistical dependence across scale in natural images [e.g., 8, 37]. Note, however, that the inclusion of the parent results in only a modest increase in performance compared to the other elements shown in table 2. We believe this is because the relationship between responses at different scales is more complicated than what we have captured with local covariance, and ultimately requires a more sophisticated model (see below).

In addition to these modeling differences, there are also differences between our denoising method and previous methods based on hidden-variable models [23, 6, 5, 7, 2]. First, we compute the full optimal Bayesian least squares solution, as opposed to first estimating the local variance, and then using this to estimate the coefficient. We have shown empirically that this approach yields a very important improvement in the results. Also, we use the vectorial form of the LLS solution. These enhancements, together with a reasonable choice

for the prior of the hidden multiplier, result in a substantial improvement in the quality of the denoised images, unprecedented both in visual and numerical (PSNR) terms.

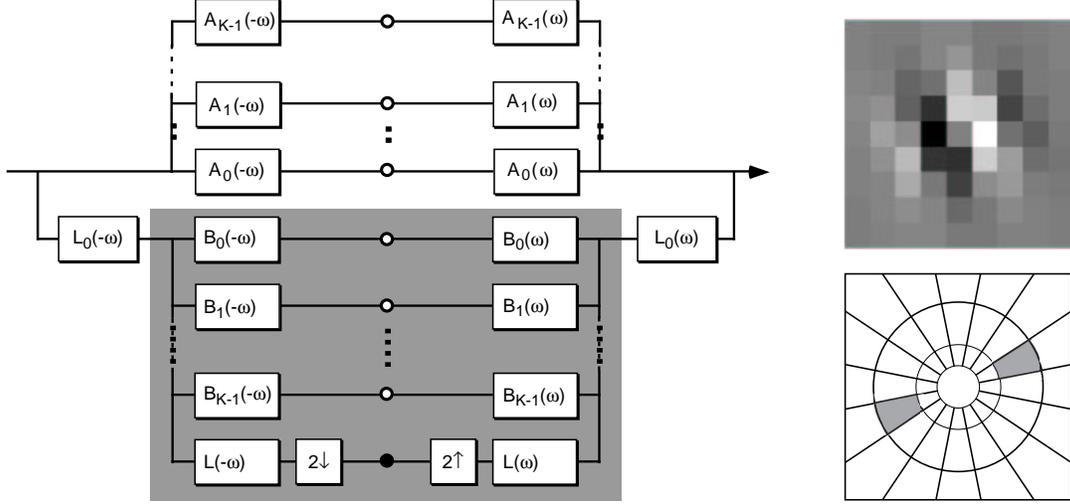
We have also applied our method to a digital camera image, after calibration of the sensor noise properties. In particular, we have calibrated and incorporated two important features of the sensor noise: spatial correlation, and signal-dependence. The results presented here, although preliminary in nature, are quite promising. We are also extending the denoising solution to address the complete image restoration problem, by incorporating a model of image blur. Preliminary results for both extensions appear quite promising.

We believe that the current image model can be improved in a number of ways. It would be desirable to develop a method for efficiently estimating a prior for the multiplier in each subband by maximizing the joint likelihood of the observed subbands, as opposed to the somewhat heuristic choice of noninformative prior we have used in our implementation. It would also be useful to develop a fully blind solution, in which the noise parameters (covariance) are estimated from the observed (noisy) image. In addition, it is worth exploring the transformation of the local GSM model into an explicit Markov model on a graph with cycles, as opposed to the tree-structured models previously developed [25, 2]. This conceptual simplification would facilitate other applications requiring a conditional local density model (e.g., synthesis or coding). Finally, major improvements are likely to come from statistical models that include dependencies beyond those associated with covariance and magnitude, such as phase congruency relationships between complex wavelet coefficients [e.g., 46], that can capture the local structural properties of typical image features.

## A The steerable pyramid

We use a transform known as a *steerable pyramid* [14, 47] to decompose images into frequency subbands. The transform is implemented in the Fourier domain, allowing exact reconstruction of the image from the subbands, as well as a flexible choice of the number of orientations ( $K$ ) and scales ( $J$ ). A software implementation (in MatLab) is available at <http://www.cns.nyu.edu/~lcv/software.html>. As with conventional orthogonal wavelet decompositions, the pyramid is implemented by recursively splitting an image into a set of oriented subbands, and a lowpass residual band which is subsampled by a factor of two along both axes. Unlike conventional orthogonal wavelet decompositions, the oriented bands are not subsampled, and the subsampling of the lowpass band does not produce aliasing artifacts, as the lowpass filter is designed to obey the Nyquist sampling criterion. When performing convolutions, the boundaries are handled by mirror extension (reflection) of the image, thereby maintaining continuity. Since it is a tight frame, the transformation may be inverted by convolving each subband with its associated complex-conjugated filter and adding the results. The redundancy factor of this overcomplete representation is (for  $J \rightarrow \infty$ )  $\frac{7}{3}K + 1$ .

The system diagram for the transform is shown in Fig. 10. The filters are polar-separable



**Fig. 10. Left:** System diagram for the extended version of the steerable pyramid used in this paper [14]. The input image is first split into a lowpass band and a set of high-pass oriented bands. The lowpass band is then split into a lower-frequency band and a set of oriented subbands. The pyramid recursion consists of inserting the diagram contents of the shaded region at the the lowpass branch (solid circle). **Right:** Basis function corresponding to an example oriented subband, and idealized depiction of the frequency domain partition ( $K = 8$ ,  $J = 2$ ), with gray region corresponding to this basis function.

in the Fourier domain, where they may be written as:

$$L(r, \theta) = \begin{cases} \cos\left(\frac{\pi}{2} \log_2\left(\frac{4r}{\pi}\right)\right), & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 1, & r \leq \frac{\pi}{4} \\ 0, & r \geq \frac{\pi}{2} \end{cases}$$

$$B_k(r, \theta) = H(r)G_k(\theta), \quad k \in [0, K-1],$$

where  $r, \theta$  are polar frequency coordinates, and

$$H(r) = \begin{cases} \cos\left(\frac{\pi}{2} \log_2\left(\frac{2r}{\pi}\right)\right), & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 1, & r \geq \frac{\pi}{2} \\ 0, & r \leq \frac{\pi}{4} \end{cases}$$

$$G_k(\theta) = \frac{(K-1)!}{\sqrt{K}[2(K-1)]!} \left[ 2 \cos\left(\theta - \frac{\pi k}{K}\right) \right]^{K-1}.$$

The recursive procedure is initialized by splitting the input image into lowpass and oriented highpass portions, using the following filters:

$$L_0(r, \theta) = L\left(\frac{r}{2}, \theta\right), \quad A_k(r, \theta) = H\left(\frac{r}{2}\right)G_k(\theta).$$

Figure 10 also shows the impulse response of an example oriented filter (for  $K = 8$ ), at the highest resolution level, together with its (purely imaginary) Fourier transform.

## B Origin of the test images

All test images used in obtaining our results are available on the internet from <http://www.cns.nyu.edu/lcv/denoise>. Five of the images, commonly known as *Lena*, *Bar-*

*bara*, *Boats*, *House* and *Peppers*, are available in various image database collections and widely used in the image processing literature. The first three are  $512 \times 512$  and the last two are  $256 \times 256$ . We also included a  $512 \times 512$  image of a fingerprint, which unlike the other images, is a homogeneous texture. Unfortunately, most test images are available in more than one version, with differences between them due to cropping, scanning, or conversion from color to gray-level.

Among the three or four popular versions of *Lena*, we chose the one that is perhaps the most standard, from the image data base of the Waterloo fractal compression project at <http://links.uwaterloo.ca/greyset2.base.html>. We have confirmed that this is the same version used in [30] and [45], as shown in the comparison of Fig. 5(a), but have not verified that it is identical to that used in [5] or [23]. The *Barbara* image was obtained from Marco Schmidt’s standard test images database at <http://jiu.sourceforge.net/testimages/index.html>. This version has been previously used in [30], who, in turn, have made a comparison to [5, 23]. The *Boats* image was taken from University of Southern California SIPI image database at <http://sipi.usc.edu/services/database/database.cgi>. The *House* and *Peppers* images were generously provided by Dr. Pižurica, for proper comparison to the results reported in [44].

## C Digital camera sensor: model and calibration

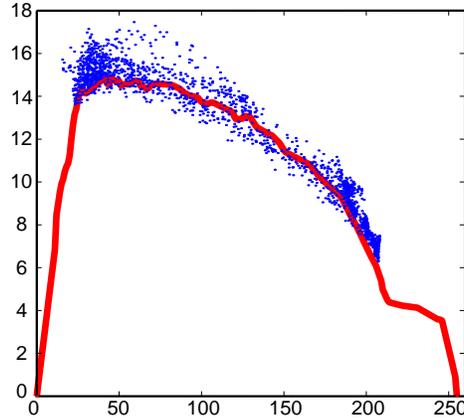
We performed a calibration procedure in order to model and quantify the noise properties of digital images gathered with a Canon PowerShot G1 consumer digital camera.<sup>2</sup> We took photographs of uniform gray-level patches (using a Kodak Q60 test target) under reasonably uniform lighting conditions. Exposures were made with the sensor gain set to the equivalent of ISO 400 (the highest, and noisiest, setting on this camera). For each color channel and within each patch, we extracted a central rectangular portion of pixels, measured the mean intensity of this patch, subtracted the best fitting plane (to allow for linear illumination variation), and then measured the variance of the residual.

The relationship between mean and standard deviation for the blue channel (the noisiest of the three) is plotted as a solid line in the graph shown in Fig. 11. We have added the two endpoints, at the minimal and maximal intensity levels (0 and 255), since an image patch with either of these mean values must necessarily have a variance of zero.

As a confirmation of the applicability of this calibration curve to images taken under more general conditions, this graph also contains points corresponding to the estimated mean and standard deviation of the laboratory image described in section 5.3. Specifically, we estimate the local mean intensity of the 400-ISO laboratory image by convolving with a Gaussian kernel and the local standard deviation of the noise by subtracting the corresponding reference image. We see a good agreement between the calibration curve and the

---

<sup>2</sup>We thank Nicolas Bonnier for assistance in digital camera characterization and calibration.



**Fig. 11.** Agreement between the calibration curve (continuous line) and the empirically obtained values of local noise standard deviation vs. local intensity mean, for the blue channel. See text for details.

laboratory image.

## References

- [1] N Graham, *Visual pattern analyzers*, Oxford University Press, New York, 1989.
- [2] 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*, vol. 11, no. 1, pp. 89–123, July 2001.
- [3] D Andrews and C Mallows, “Scale mixtures of normal distributions,” *J. Royal Stat. Soc.*, vol. 36, pp. 99–, 1974.
- [4] M J Wainwright and E P Simoncelli, “Scale mixtures of Gaussians and the statistics of natural images,” in *Adv. Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press.
- [5] M K Mihçak, I Kozintsev, K Ramchandran, and P Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Trans. Sig. Proc.*, vol. 6, no. 12, pp. 300–303, December 1999.
- [6] V Strela, J Portilla, and E Simoncelli, “Image denoising using a local Gaussian scale mixture model in the wavelet domain,” in *Proc SPIE, 45th Annual Meeting*, San Diego, July 2000.
- [7] J Portilla, V Strela, M Wainwright, and E Simoncelli, “Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain,” in *Proc 8th IEEE Int’l Conf on Image Proc*, Thessaloniki, Greece, Oct 7-10 2001, pp. 37–40.
- [8] D L Ruderman, “The statistics of natural images,” *Network: Computation in Neural Systems*, vol. 5, pp. 517–548, 1996.
- [9] D J Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [10] John G. Daugman, “Entropy reduction and decorrelation in visual coding by oriented neural receptive fields,” *IEEE Trans. Biomedical Engineering*, vol. 36, no. 1, pp. 107–114, 1989.
- [11] S G Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Pat. Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

- [12] B A Olshausen and D J Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, pp. 607–609, 1996.
- [13] A J Bell and T J Sejnowski, “The ‘independent components’ of natural scenes are edge filters,” *Vision Research*, vol. 37, no. 23, pp. 3327–3338, 1997.
- [14] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger, “Shiftable multi-scale transforms,” *IEEE Trans Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992, Special Issue on Wavelets.
- [15] R R Coifman and D L Donoho, “Translation-invariant de-noising,” in *Wavelets and statistics*, A Antoniadis and G Oppenheim, Eds. Springer-Verlag lecture notes, San Diego, 1995.
- [16] J Starck, E J Candes, and D L Donoho, “The curvelet transform for image denoising,” *IEEE Trans. Image Proc.*, vol. 11, no. 6, pp. 670–684, June 2002.
- [17] 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*, Lausanne, Switzerland, 1990, vol. 1360, pp. 909–922.
- [18] E P Simoncelli, “Statistical models for images: Compression, restoration and synthesis,” in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1997, pp. 673–678.
- [19] R W Buccigrossi and E P Simoncelli, “Image compression via joint statistical characterization in the wavelet domain,” *IEEE Trans Image Proc*, vol. 8, no. 12, pp. 1688–1701, December 1999.
- [20] T Bollersley, K Engle, and D Nelson, “ARCH models,” in *Handbook of Econometrics V*, B Engle and D McFadden, Eds. 1994.
- [21] J S Lee, “Digital image enhancement and noise filtering by use of local statistics,” *IEEE Pat. Anal. Mach. Intell.*, vol. PAMI-2, pp. 165–168, March 1980.
- [22] M Malfait and D Roose, “Wavelet-based image denoising using a Markov random field a priori model,” *IEEE Trans. Image Proc.*, vol. 6, pp. 549–565, April 1997.
- [23] S G Chang, Bin Yu, and Martin Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” in *Fifth IEEE Int’l Conf on Image Proc*, Chicago, October 1998.
- [24] Helmut Brehm and Walter Stammers, “Description and generation of spherically invariant speech-model signals,” *Signal Processing*, vol. 12, pp. 119–141, 1987.
- [25] M S Crouse, R D Nowak, and R G Baraniuk, “Wavelet-based statistical signal processing using hidden Markov models,” *IEEE Trans. Signal Proc.*, vol. 46, pp. 886–902, April 1998.
- [26] 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.
- [27] C Spence and L Parra, “Hierarchical image probability (HIP) model,” in *Adv. Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, MIT Press.
- [28] H. Robbins, “The empirical Bayes approach to statistical decision problems,” *Ann. Math. Statistics*, vol. 35, pp. 1–20, 1964.
- [29] E P Simoncelli, “Bayesian denoising of visual images in the wavelet domain,” in *Bayesian Inference in Wavelet Based Models*, P Müller and B Vidakovic, Eds., chapter 18, pp. 291–308. Springer-Verlag, New York, 1999, Lecture Notes in Statistics, vol. 141.
- [30] Xin Li and Michael T. Orchard, “Spatially adaptive image denoising under overcomplete expansion,” in *IEEE Int’l Conf on Image Proc*, Vancouver, September 2000.

- [31] A B Watson, “The cortex transform: rapid computation of simulated neural images,” *Comp. Vis. Graphics Image Proc.*, vol. 39, pp. 311–327, 1987a.
- [32] N Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.
- [33] V Strela, “Denoising via block Wiener filtering in wavelet domain,” in *3rd European Congress of Mathematics*, Barcelona, July 2000, Birkhäuser Verlag.
- [34] E P Simoncelli and E H Adelson, “Noise removal via Bayesian wavelet coring,” in *Third Int’l Conf on Image Proc*, Lausanne, September 1996, vol. I, pp. 379–382.
- [35] P Moulin and J Liu, “Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors,” *IEEE Trans. Info. Theory*, vol. 45, pp. 909–919, 1999.
- [36] G E P Box and C Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1992.
- [37] Jerome Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans Sig Proc*, vol. 41, no. 12, pp. 3445–3462, December 1993.
- [38] D L Ruderman, “Origins of scaling in natural images,” *Vision Research*, vol. 37, pp. 3385–3398, 1997.
- [39] A Taberner, J Portilla, and R Navarro, “Duality of log-polar image representations in the space and the spatial-frequency domains,” *IEEE Trans. Signal Proc.*, vol. 47, no. 9, pp. 2469–2479, September 1999.
- [40] D Donoho, “Denoising by soft-thresholding,” *IEEE Trans. Info. Theory*, vol. 43, pp. 613–627, 1995.
- [41] J P Rossi, ,” *JSMPTTE*, vol. 87, pp. 134–140, 1978.
- [42] D Leporini and J C Pesquet, “Multiscale regularization in Besov spaces,” in *31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1998.
- [43] F Abramovich, T Sapatinas, and B W Silverman, “Wavelet thresholding via a Bayesian approach,” *J R Stat Soc B*, vol. 60, pp. 725–749, 1998.
- [44] A Pižurica, W Philips, I Lemahieu, and M Acheroy, “A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising,” *IEEE Trans. Image Proc.*, vol. 11, no. 5, pp. 545–557, May 2002.
- [45] J L Starck, D L Donoho, and E Candes, “Very high quality image restoration,” in *Proc. SPIE conf. Signal and Image Processing*, A Laine, M Unser, and A Aldroubi, Eds., San Diego, August 2001, vol. 4478, pp. 9–19.
- [46] J Portilla and E P Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int’l Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.
- [47] W T Freeman and E H Adelson, “The design and use of steerable filters,” *IEEE Pat. Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, 1991.