

# Image Denoising Using Derotated Complex Wavelet Coefficients

Mark Miller, *Member, IEEE*, and Nick Kingsbury, *Member, IEEE*

**Abstract**—A method for removing additive Gaussian noise from digital images is described. It is based on statistical modeling of the coefficients of a redundant, oriented, complex multiscale transform. Two types of modeling are used to model the wavelet coefficients. Both are based on Gaussian scale mixture (GSM) modeling of neighborhoods of coefficients at adjacent locations and scales. Modeling of edge and ridge discontinuities is performed using wavelet coefficients derotated by twice the phase of the coefficient at the same location and the next coarser scale. Other areas are modeled using standard wavelet coefficients. An adaptive Bayesian model selection framework is used to determine the modeling applied to each neighborhood. The proposed algorithm succeeds in providing improved denoising performance at structural image features, reducing ringing artifacts and enhancing sharpness, while avoiding degradation in other areas. The method outperforms previously published methods visually and in standard tests.

**Index Terms**—Complex, denoising, image, interscale phase, restoration, wavelet.

## I. INTRODUCTION

WAVELET transforms have emerged as the premier tool for image denoising, due to the statistically useful properties of wavelet coefficients of natural images. The sparseness property of wavelet coefficients and tendency of wavelets bases to diagonalize images allows us to break the problem into modeling a small number of “neighboring” coefficients (in space and scale) to reduce the dimensionality and improve the tractability of the problem.

State-of-the-art Gaussian scale mixture (GSM) denoising algorithms employing over-complete multiscale transforms achieve impressive results by modeling images according to the activity within neighborhoods of wavelet coefficients and attenuating coefficients heavily in “inactive” image regions to remove noise [1]. However, under the basic GSM model, there is no distinction between the different basic components of images that cause a neighborhood to be active. This commonly

Manuscript received October 8, 2006; revised July 19, 2007. Published August 13, 2008 (projected). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pier Luigi Dragotti.

M. Miller is with the Signal Processing and Communications Group, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ U.K. (e-mail: m.a.miller.02@cantab.net).

N. Kingsbury is with the Signal Processing and Communications Group, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ U.K., and also with Trinity College, Cambridge CB2 1TQ U.K. (e-mail: ngk@eng.cam.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.926146



Fig. 1. Popular test images. (a) Lena. (b) Barbara.

leads to ringing artifacts in the vicinity of edge and ridge discontinuities when removing medium and high levels of noise.

Consider the popular *Lena* and *Barbara* images shown in Fig. 1. The active regions of these images could be broadly decomposed into two categories: areas of texture such as *Lena*'s boa, and multiscale edge and ridge features such as the outline of her figure and the edges of background objects. We call the latter *structural features*. Texture can also be broadly decomposed into two types (or perhaps more accurately a spectrum between two extremes), periodic texture such as *Barbara*'s clothing and tablecloth and more random texture such as *Lena*'s boa.

For areas of significant image activity, we propose a dual model framework which divides the active areas into two components: 1) structural features and 2) other features including texture.

Structural features will be modeled using interscale phase relationships of complex wavelet coefficients. The method is based on the novel “derotated” coefficients introduced in [2]. The transform used is the dual tree complex wavelet transform (DT-CWT) [3]. This specific modeling of structural features is combined with standard modeling of other features using complex wavelet coefficients in an adaptive Bayesian model selection framework. Both models employ the GSM approach to deal with the spatially varying energy statistics of typical images within each wavelet subband.

The content of this paper is organized as follows. Section II provides relevant background information. Section III details the proposed denoising algorithm. Section IV discusses the constitution of the neighborhoods to which the modeling is applied. A summary of the algorithm is provided in Section V, and results are given in Section VI. Sections VII and VIII contain conclusions and suggestions for future work.

## II. BACKGROUND

### A. Dual Tree Complex Wavelet Transform

The dual tree complex wavelet transform [3] uses a dual tree of real wavelet filters to generate the real and imaginary parts of complex wavelet coefficients. This introduces a limited amount of redundancy and allows the transform to provide approximate shift invariance and directionally selectivity filters, while preserving the usual properties of perfect reconstruction and computational efficiency.

For a  $d$ -dimensional input, an  $L$  scale DT-CWT outputs an array of real scaling coefficients corresponding to the lowpass subbands in each dimension and  $(4^d - 2^d)/(2)$  directional subbands of  $(M)/(2^{dl})$  complex wavelet coefficients at level  $l$ , where  $M$  is the total size of the input data. The mechanics of the DT-CWT are not covered here. See [3] and [4] for a comprehensive explanation of the transform and details of filter design for the trees.

In two dimensions, the transform produces six directional subbands at each scale. We often describe processing on a local neighborhood or neighborhood window of wavelet coefficients. This refers to a group of “local” coefficients at nearby spatial locations and adjacent scales. Parent and child coefficients refer to coefficients in the next coarser and finer subbands respectively in the same directional subband and at the same spatial location (possibly interpolated).

### B. Derotated Coefficients

In [2], we introduce complex wavelet coefficients that are derotated by twice the phase of the parent coefficient, i.e., the coefficient at the next coarser scale at the same spatial location, such that their phase (in addition to the magnitude as described in [3]) becomes invariant at multiscale edges and ridges and is consistent within each subband for each type of feature. The derotated coefficients are shown to offer increased correlation at image edge and ridge discontinuities relative to standard wavelet coefficients and, therefore, the potential for improved estimation in additive noise. Equation (1) defines the new derotated coefficient  $w$ , where  $x$  is a DT-CWT coefficient and  $x_p$  the corresponding parent coefficient

$$\begin{aligned} |w| &= |x| \\ \angle(w) &= \angle(x) - 2\angle(x_p). \end{aligned} \quad (1)$$

For a neighborhood of wavelet coefficients  $\mathbf{x}$ , we can write the derotation as a matrix operation as shown in (2). In (2),  $w$  is a vector of derotated coefficients and  $A$  is a unitary rotation matrix which rotates the each coefficient’s phase by twice the phase of its parent wavelet coefficient

$$\mathbf{x} = \mathbf{A}\mathbf{w}. \quad (2)$$

### C. Statistical Image Modeling With Wavelets

An accurate model, whether implicit or explicit, is a critical component of nearly all image processing tasks. For statistical

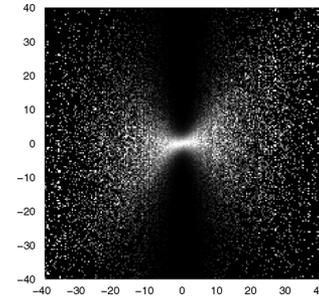


Fig. 2. “Bow tie” characteristic of wavelet coefficients of natural images. Conditional histogram of two spatially adjacent wavelet coefficients for the *Lena* image. Brightness corresponds to probability except that each column has been scaled to fill the range of available intensities. Although the coefficients are roughly second-order decorrelated, they are highly dependent. The standard deviation of a coefficient scales with the magnitude of the neighboring coefficient.

approaches the choice of a suitable stochastic model is vital. This section examines how to model wavelet coefficients to take advantage of the statistical characteristics common to many photographic images. Apart from the interscale phase properties introduced here, the key properties of wavelet representations that we wish to exploit are as follows.

- **Sparseness:** Good wavelet transforms have an energy compaction property and the transforms of natural signals tend to consist of only a few larger coefficients and many smaller coefficients.
- **Spatial clustering:** Strong dependencies in the form of spatial clusters exist between large magnitude wavelet coefficients in each subband, due to edges and areas of texture in the data.
- **Persistence across scale:** The magnitudes of wavelet coefficients are correlated across scale. If a parent coefficient is small, its children are more likely to be small.

It is widely known that the wavelet coefficients of images display highly nonstationary non-Gaussian statistics [5]. The marginal distributions are typically long tailed with high kurtosis (fourth moment divided by the variance squared). The shape, including the sharp peak at zero and the long tails, is the statistical manifestation of the sparseness property of wavelet coefficients. These distributions have previously been modeled using independent generalized Gaussian distributions also known as generalized Laplacian distributions [5], [6]. However, this framework fails to take advantage of the spatial clustering property or the persistence across scale of wavelet coefficients. Coefficients of natural photographic images exhibit marked higher order dependencies between neighboring coefficients. This form of dependency is illustrated in Fig. 2.

Fig. 2 shows a conditional histogram of two spatially adjacent wavelet coefficients typical of natural images. The coefficients were produced using the (DT-CWT) of the  $512 \times 512$  *Lena* image and are taken from a near horizontal subband at level 2. Brightness corresponds to probability except that each column has been scaled to fill the range of available intensities. This so called “bow tie” characteristic shape is found for wavelet coefficients at nearby spatial locations, as well as adjacent scales and directional subbands, for a wide range of natural images [7]–[9].

The wavelet coefficients of natural images display a self-reinforcing characteristic in that if one coefficient is large in magnitude, then other coefficients in its neighborhood are also likely to be large in magnitude. The intuitive explanation for this is that localized image structures such as edges have substantial power across many scales and nearby spatial locations at a given orientation. The wavelet coefficients that represent the image will also have large magnitudes at these scales, locations and orientation. However, the signs and relative magnitudes of these coefficients will depend on the exact shape, location, and orientation of the structure.

#### D. GSM Models for Wavelet Coefficients

A simple statistical model that has been used to model natural signals such as speech, and more recently to describe the nonstationary behavior of the wavelet coefficients of natural images, is given in (3). It assumes that each coefficient  $x(\mathbf{t})$  is specified by a stationary zero mean Gaussian process  $u(\mathbf{t})$  and a spatially fluctuating variance  $z(\mathbf{t})$ .  $\mathbf{t}$  is a position vector

$$x(\mathbf{t}) = \sqrt{z(\mathbf{t})}u(\mathbf{t}). \quad (3)$$

To model the self-reinforcing property of the coefficients,  $z(\mathbf{t})$  must be slowly varying in  $\mathbf{t}$  but need not be symmetric in all directions. It has been shown that for slowly varying  $z(\mathbf{t})$ , this model can successfully simulate the high kurtosis and longer tails of the marginal distributions in addition to the bow-tie shape of the marginal histograms of wavelet coefficients of natural signals [1], [7]. Note that the above model describes intrascale relationships only. A more comprehensive model would include a scale parameter  $l$  so that  $z = z(\mathbf{t}, l)$  to model magnitude persistence across scale.

The stationary portion of the model  $u(\mathbf{t})$  is Gaussian distributed over a small neighborhood of wavelet coefficients. It is generally assumed that  $z(\mathbf{t})$  varies slowly enough to be considered constant over that neighborhood of coefficients. Under this assumption the model is now a particular form of a spherically invariant random process called a GSM.

For a small neighborhood of coefficients at nearby spatial locations and scale, we now have a GSM vector  $\mathbf{x}$ , which is the product of two independent random variables: a positive scalar  $z$  referred to as the hidden multiplier or mixing variable and a Gaussian random vector  $\mathbf{u}$  distributed as  $\mathcal{N}(0, \mathbf{C}_u)$ . For each neighborhood of wavelet coefficients, this is written

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

It remains to specify the prior probability function  $p_z(z)$  for the multiplier  $z$ . Prior selection is covered in more detail in [1]. In the proposed algorithm, a Jeffrey's prior is used for the reasons stated below, at the end of Section II-E.

#### E. Wavelet Denoising

Image denoising involves finding an estimate  $\hat{\mathbf{x}}_s$  of a signal  $\mathbf{x}_s$  in noise  $\mathbf{n}_s$  given a noisy observation  $\mathbf{y}_s$ . This is summarized in (5)

$$\mathbf{y}_s = \mathbf{x}_s + \mathbf{n}_s. \quad (5)$$

The standard problem definition in much of the denoising literature is to assume zero mean Gaussian noise with covariance  $\mathbf{C}_n$ , so that  $\mathbf{n}_s$  is distributed as  $\mathcal{N}(0, \mathbf{C}_n)$ . In many cases the noise is assumed to be white so that  $\mathbf{C}_n = \sigma_s^2 \mathbf{I}$ . It is assumed that the variance  $\sigma_s^2$  or covariance  $\mathbf{C}_n$  of the noise is known. If it is not known, it will have to be obtained from an area of the image known to have little or no signal content or estimated using some other method, e.g., as in [10].

The common approach to wavelet based denoising is to transform the signal into the wavelet domain, denoise the detail coefficients and transform back to the image domain. Note that the lowpass scaling coefficients are not usually altered. Taking the forward transform of (5), we obtain (6)

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \quad (6)$$

In (6), the subscript  $s$  is dropped to indicate the variables are in the wavelet rather than the spatial domain and each vector is the complex wavelet transform of the corresponding vector in (5). The real and imaginary parts are considered as separate elements within each vector. In this paper we consider only one neighborhood of coefficients at a time. In this case, only the coefficients in the neighborhood are included in (6).

*Simple Thresholding Noise Removal:* Classical wavelet based denoising techniques employ straightforward nonlinear thresholding of noisy wavelet coefficients. They are motivated by the sparseness property of wavelet coefficients mentioned in Section II-C and ignore the clustering properties. Their objective is to suppress low amplitude coefficients which are more likely to constitute noise, and retain high amplitude values which contain the bulk of the desired signal. Two of the earliest thresholding operators were hard and soft thresholding [10], [11]. Various other operators have been suggested [6] and [12]–[15].

*GSM Techniques:* A significant block of wavelet denoising literature proposes variations on a common methodology. Although the reasoning used to derive each denoising scheme varies, many of the resulting methods can be described in terms of the GSM framework outlined in Section II-C. In general, the assumption of spatial and spectral locality is invoked and the estimate is based on a "local" neighborhood of wavelet coefficients at the same or adjacent spatial locations and scales. The inclusion of coefficients at the parent scale is often used to capture the "persistence across scale" characteristic.

Equation (4) can be combined with (6) to produce (7), where the vector of coefficients is a "local" neighborhood

$$\mathbf{y} = \sqrt{z}\mathbf{u} + \mathbf{n}. \quad (7)$$

Conditioned on the hidden multiplier for a neighborhood, the noisy observation is Gaussian distributed with zero mean and covariance  $z\mathbf{C}_u + \mathbf{C}_n$ , as given by (8) in which  $N$  is the size of the neighborhood

$$p_{\mathbf{y}|z}(\mathbf{y}|z) = \frac{(2\pi)^{-N/2}}{|z\mathbf{C}_u + \mathbf{C}_n|^{1/2}} \times \exp\left(\frac{-\mathbf{y}^T(z\mathbf{C}_u + \mathbf{C}_n)^{-1}\mathbf{y}}{2}\right). \quad (8)$$

As a consequence, for a given value of  $z$ , the minimum mean square error (MMSE) estimate for the original coefficients in the neighborhood is given by (9), which is an adaptive Wiener estimate with signal covariance  $z\mathbf{C}_u$ . See Jain [16, pp. 276–279] for further details of Wiener filtering.  $\mathbf{C}_u$  and  $\mathbf{C}_n$  are the covariance matrices for the vectors of wavelet coefficients of the Gaussian component of the mixture model and noise, respectively

$$\hat{\mathbf{x}} = z\mathbf{C}_u(z\mathbf{C}_u + \mathbf{C}_n)^{-1} \mathbf{y}. \quad (9)$$

In 2003, Portilla *et al.* proposed a somewhat new approach to estimation of the wavelet coefficients using a GSM framework and the Steerable Pyramid wavelet transform [1]. They resolved to calculate the Bayesian MMSE estimate for a particular coefficient in a neighborhood of coefficients considered to be a GSM. As for the aforementioned two-step procedure, this technique is implemented on a number of overlapping neighborhoods—one for each coefficient. The result is presented in (10), where  $\hat{x}_c$  is the estimate for the “central” complex coefficient

$$\begin{aligned} \hat{x}_c &= \mathbb{E}\{\mathbf{x}_c | \mathbf{y}\} \\ &= \int_0^\infty p(z | \mathbf{y}) E\{\mathbf{x}_c | \mathbf{y}, z\} dz. \end{aligned} \quad (10)$$

In a discrete implementation, this takes the form of (11), where  $K$  is the number of discrete values assigned to  $z$

$$\hat{x}_c = \sum_{k=1}^K p(z_k | \mathbf{y}) E\{x_c | \mathbf{y}, z_k\}. \quad (11)$$

Conditioned on the multiplier  $z$ , the MMSE estimate for the neighborhood of coefficients  $E\{\mathbf{x} | \mathbf{y}, z_k\}$  is given by (9). The posterior density  $p(z | \mathbf{y})$  is also required for (11). This can be calculated using Bayes formula, as shown in (12)

$$p(z | \mathbf{y}) = \frac{p(\mathbf{y} | z) p_z(z)}{\int p(\mathbf{y} | a) p_z(a) da}. \quad (12)$$

The density  $p(\mathbf{y} | z)$  is given by (8), but a decision must be made on the choice for the prior  $p_z(z)$ . In [1], Portilla *et al.* selected a Jeffrey’s prior, because it produced superior results to other the options implemented.

The results produced by this advance by Portilla and colleagues are impressive. A recent adjunct to this work attempts to use a spatially adaptive signal covariance matrix by assuming geometrically close areas have similar covariance statistics [17]. It is interesting to compare this approach with the algorithm presented here where the covariance information is adapted for multiscale features throughout the image using information from the next coarser scale to adjust for the type of feature and its particular angle.

#### F. Test Images

The images used in this article are well known test images that have been used in a variety of image estimation literature. However, most of the images are available in more than one version, with differences between them due to cropping, scanning, resizing, compression or conversion from color to gray level. For comparative purposes, the images used to test the proposed

TABLE I  
IMAGES USED TO TEST THE PROPOSED DENOISING ALGORITHMS

Image	Size
Barbara	512×512
House	256×256
Lenna	512×512
Peppers	256×256

denoising algorithm are those used in [1] which are available at [http://decsai.ugr.es/~javier/denoise/test\\_images/](http://decsai.ugr.es/~javier/denoise/test_images/).

However, the *Peppers* image was found to have a row and column of zero pixels at the top and left-hand edges of the image. These were filled in using the adjacent row and column, although for comparison purposes the original version is also used in the results section. Table I lists the images used.

### III. PROPOSED DENOISING ALGORITHM

The top-level denoising strategy used here is the same as most other wavelet based denoising algorithms: decompose the noisy image into  $L$  levels of six directional subbands and a set of (low-pass) scaling coefficients, denoise the complex wavelet coefficients in each subband except for the scaling coefficients and invert the transform to obtain the image estimate.

#### A. Modeling

Based on the investigations in [2], derotated wavelet coefficients are used to model structural features. Complex wavelets have been shown in recent denoising literature to be an effective method of representing images and have been shown in [2] to be particularly good, in terms of covariance information, as a basis in areas of periodic texture. Hence, standard wavelet coefficients are used to represent the image areas not near structural features. The two models are combined using a Bayesian model selection method.

Because the derotated coefficients have the same magnitude as standard DT-CWT coefficients we can retain the GSM modeling of wavelet coefficients used in [1] to capture the clustering characteristic and persistence across scale of wavelet coefficient magnitudes for both models.

If derotated coefficients are to be used, an accurate estimate of the phases of the parent coefficients will be required. This is afforded by the nature of multiscale denoising algorithms. Due to the spectral characteristics of typical natural images, wavelet coefficients of noisy images have a significantly higher signal to noise ratio at the parent scale relative to that of the child. If, in addition, the coefficient has been otherwise denoised, we can assume that a denoised parent coefficient is relatively noise free compared to a noisy child coefficient.

Because the parent subband is sampled at 1/4 of the density of the child subband, parent coefficients need to be interpolated from the next coarser subband of the same orientation. Effective interpolation of the parent coefficients is crucial in obtaining accurate phase information. The method used here is bandpass interpolation which unwraps the expected phase rotations within each subband before interpolation. Details of this method are provided in [18, Appendix C].

The modeling uses a neighborhood approach standard in wavelet based GSM algorithms. The wavelet coefficients are divided into overlapping groups of coefficients “neighboring”

one another in spatial location and scale. Denoising is performed on a “central” complex coefficient based on the model for the whole neighborhood.

A vector of observed wavelet coefficients in a given neighborhood  $\mathbf{y}$  can be written in terms of wavelet coefficients  $\mathbf{x}$  and  $\mathbf{n}$  representing the clean image and noise as given in (6). Within each neighborhood, the real and imaginary parts of complex coefficients are treated as separate variables. This is necessary to fully capture the statistics of the derotated coefficients and is discussed further in [18, Appendix D].

To integrate the specific modeling of structural image features with standard GSM modeling two different models are assumed for the neighborhoods of coefficients  $\mathbf{x}$  and a Bayesian framework is used to combine them. The models are presented in (13) and (14). As discussed in the Introduction, model 1 is intended to represent areas of texture using standard complex wavelet coefficients and model 2 to represent the major structural features of the image using derotated complex wavelet coefficients.

Model 1

$$\mathbf{x} = \sqrt{z} \mathbf{u}. \quad (13)$$

Model 2

$$\mathbf{x} = \mathbf{A}\mathbf{w} = \sqrt{z} \mathbf{A} \mathbf{q}. \quad (14)$$

Model 1 is the standard GSM modeling of a neighborhood of wavelet coefficients  $\mathbf{x}$  as described in Section II-E where  $z$  is the hidden or GSM multiplier and  $\mathbf{u}$  is a neighborhood of Gaussian variables with zero mean and covariance  $\mathbf{C}_u$ . In model 2,  $\mathbf{q}$  is a vector of Gaussian distributed random variables with covariance  $\mathbf{C}_q$ .  $\mathbf{A}$  is a unitary spatially varying inverse derotation matrix, which converts a set of derotated coefficients  $\mathbf{q}$  to the corresponding DT-CWT coefficients using the phase of the interpolated parent coefficients. Based on the assumption that an edge or ridge feature of a given polarity is equally likely to one of the opposite polarity,  $\mathbf{q}$  is assumed to have zero mean. See [2] for more background on derotated coefficients.

## B. Denoising

For each neighborhood of coefficients we wish to estimate a “central” coefficient  $x_c$  from the set of noisy coefficients in the neighborhood  $\mathbf{y}$ . To manage selection between the two models, we introduce a discrete model selection random variable  $m$  which can have values  $m_1$  and  $m_2$ . The Bayesian MMSE estimate is given in (15) derived in a similar manner to (10)

$$\begin{aligned} \hat{x}_c &= \mathbb{E}\{x_c | \mathbf{y}\} \\ &= \int x_c p(x_c | \mathbf{y}) dx_c \\ &= \int \int_0^\infty \sum_{b=1}^2 x_c p(x_c, z, m_b | \mathbf{y}) dz dx_c \\ &= \int_0^\infty \sum_{b=1}^2 p(z, m_b | \mathbf{y}) E\{x_c | \mathbf{y}, z, m_b\} dz. \end{aligned} \quad (15)$$

It remains to determine the expected value of the neighborhood’s central coefficient  $x_c$  for a given  $z$  for each model and an expression for the joint posterior probability for the model and multiplier variables given the observed noisy coefficients,  $p(z, m | \mathbf{y})$ .

A key advantage of the GSM framework is the tractability of the estimator  $E\{x_c | \mathbf{y}, z, m\}$ . Because the noise, as well as the vectors  $\mathbf{u}$  and  $\mathbf{q}$ , are Gaussian, for both models the expected value is a Wiener estimator conditioned on a value for  $z$  and in the case of model 2 the rotation matrix  $\mathbf{A}$ , with signal covariances  $z\mathbf{C}_u$  for model 1 and  $z\mathbf{A}\mathbf{C}_q\mathbf{A}^T$  for model 2. The estimators for each model are given in (16) and (17).

Model 1

$$E\{\mathbf{x} | \mathbf{y}, z, m_1\} = z\mathbf{C}_u(z\mathbf{C}_u + \mathbf{C}_n)^{-1} \mathbf{y}. \quad (16)$$

Model 2

$$E\{\mathbf{x} | \mathbf{y}, z, m_2\} = z\mathbf{A}\mathbf{C}_q\mathbf{A}^T(z\mathbf{A}\mathbf{C}_q\mathbf{A}^T + \mathbf{C}_n)^{-1} \mathbf{y}. \quad (17)$$

The joint posterior probability for the model and GSM multiplier are determined using Bayes formula as shown in (18) with the denominator defined in (19)

$$p(z, m | \mathbf{y}) = \frac{p(\mathbf{y} | z, m) p_{z,m}(z, m)}{p(\mathbf{y})} \quad (18)$$

$$p(\mathbf{y}) = \int \sum_{\mu=1}^2 p(\mathbf{y} | a, \mu) p_{z,m}(a, \mu) da. \quad (19)$$

In (18),  $p(\mathbf{y} | z, m)$  is Gaussian with zero mean and covariance  $z\mathbf{C}_u + \mathbf{C}_n$  for model 1 and  $z\mathbf{A}\mathbf{C}_q\mathbf{A}^T + \mathbf{C}_n$  in the case of model 2.

Model 1

$$\begin{aligned} p_{\mathbf{y}|z,m}(\mathbf{y} | z, m_1) &= \frac{(2\pi)^{-N/2}}{|z\mathbf{C}_u + \mathbf{C}_n|^{1/2}} \\ &\times \exp\left(-\frac{\mathbf{y}^T(z\mathbf{C}_u + \mathbf{C}_n)^{-1} \mathbf{y}}{2}\right). \end{aligned} \quad (20)$$

Model 2

$$\begin{aligned} p_{\mathbf{y}|z}(\mathbf{y} | z, m_2) &= \frac{(2\pi)^{-N/2}}{|z\mathbf{A}\mathbf{C}_q\mathbf{A}^T + \mathbf{C}_n|^{1/2}} \\ &\times \exp\left(-\frac{\mathbf{y}^T(z\mathbf{A}\mathbf{C}_q\mathbf{A}^T + \mathbf{C}_n)^{-1} \mathbf{y}}{2}\right). \end{aligned} \quad (21)$$

## C. Prior Probabilities

The prior  $p(z, m)$  is decomposed as shown in (22)

$$p(z, m) = p(z | m) p(m). \quad (22)$$

1) *Prior for the Hidden Multiplier  $z$* : Although the prior for the hidden multiplier  $p(z | m)$  could potentially be model dependent, currently a Jeffrey’s prior is used for both models, i.e.,  $p(z | m_1) = p(z | m_2) = 1/z$ . A Jeffrey’s prior was used in [1],

where it was found to produce superior denoising results compared to several other possibilities. It also has the advantage of simplicity, although it is an improper probability density and when implemented needs to be set to zero on an interval  $[0, \epsilon)$ , where  $\epsilon$  is a small positive number.

2) *Prior for the Model Selection Variable  $m$* : It is not logical to apply model 2 to a neighborhood if the neighborhood's dominant feature is not multiscale. That is, if the parent coefficients are not of sufficient magnitude, the phase of the derotated coefficients becomes meaningless. This issue is dealt with as follows: if the normalized weighted sum of the magnitudes of the parents of the coefficients in a neighborhood is less than the standard deviation of the noise on each complex coefficient then the parent phase is deemed "untrustworthy" and  $p(m_2)$  is set to zero for that neighborhood. Otherwise it is assumed we have no information about the likelihood of structural features in the target image and the models are given equal prior probabilities, i.e.,  $p(m_1) = p(m_2) = 0.5$ . This is described by (23) and (24)

$$p(m_2) = \begin{cases} 0, & \text{if } \sum_{n=1}^N \gamma_n |\widehat{x}_{p_n}| < \sigma \\ 0.5, & \text{otherwise} \end{cases} \quad (23)$$

$$p(m_1) = 1 - p(m_2). \quad (24)$$

In (23),  $\widehat{x}_{p_n}$  are the estimates for the parents of the coefficients in the neighborhood used for derotation and  $\sigma$  is the standard deviation of the noise on a complex coefficient in the parent subband. The positive weightings  $\gamma_n$ , which sum to unity, are defined in (25)

$$\gamma_n = \frac{|\tilde{C}_q(c, n)|}{\sum_{n=1}^N |\tilde{C}_q(c, n)|}. \quad (25)$$

In (25),  $c$  is the index of the complex coefficient that is to be denoised.  $\tilde{C}_q$  is the complex covariance matrix for the derotated coefficients obtained from the larger real covariance matrix  $C_q$ .  $C_q$  is given in (26), where  $\mathbf{q}_r$  and  $\mathbf{q}_i$  are the real and imaginary parts of  $\mathbf{q}$ .  $\tilde{C}_q$  is calculated using  $\tilde{C}_q = C_{rr} + C_{ii} + i(C_{ir} - C_{ri})$

$$C_q = E \left\{ \begin{bmatrix} \mathbf{q}_r \\ \mathbf{q}_i \end{bmatrix} \begin{bmatrix} \mathbf{q}_r^T & \mathbf{q}_i^T \end{bmatrix} \right\} = \begin{bmatrix} C_{rr} & C_{ri} \\ C_{ir} & C_{ii} \end{bmatrix}. \quad (26)$$

Note that using the adapted covariance  $\mathbf{A}C_q\mathbf{A}^T$  instead of  $C_q$  gives the same weights  $\gamma_n$ .

#### D. Calculation of Covariance Matrices

It is necessary to calculate covariance matrices for the noise  $C_n$  and the Gaussian components of both of the image models  $C_u$  and  $C_q$ . Separate matrices are calculated for each directional subband and level of the transform. All covariance matrices are real, treating the real and imaginary parts of complex coefficients separately but the matrices are constrained according to the guidelines in Appendix D of [18]. That is, for derotated image coefficients and level 1 wavelet coefficients, real covariance matrices are used where the real and imaginary parts are treated separately. All other coefficients are treated as having complex covariance matrices.

$C_n$  is estimated by generating noise data with the appropriate power spectrum and transforming this into the wavelet domain.

Provided the number of observations  $K$  is large, the sample covariance can be calculated using (27). In this case,  $\mathbf{v}_k$  are the vectors of wavelet coefficients of the noise in each neighborhood, i.e.,  $\mathbf{v}_k = \mathbf{n}_k$  and  $C_v = C_n$

$$C_v = \frac{1}{K} \sum_{k=1}^K \mathbf{v}_k \mathbf{v}_k^T. \quad (27)$$

Given  $C_n$ , the signal covariance  $C_u$  is computed from the observation covariance matrix  $C_y$  using  $C_u = C_y - C_n$  as in [1].  $C_y$  is calculated from the wavelet coefficients of the observed noisy signal using (27) where  $\mathbf{v}_k$  are now the neighborhoods of noisy coefficients  $\mathbf{y}_k$ . A similar procedure can be used to calculate  $C_q$ , as shown in (28) and (29)

$$\begin{aligned} C_{A^T y} &= E \left[ (\mathbf{A}^T \mathbf{y}) (\mathbf{A}^T \mathbf{y})^T \right] \\ &= \int_0^\infty C_{A^T y|z} p_z(z) dz \\ &= \int_0^\infty (z C_{A^T u} + C_{A^T n}) p_z(z) dz \\ &= E\{z\} C_q + C_{A^T n}. \end{aligned} \quad (28)$$

As done in [1], we set  $E\{z\}$  to unity, resulting in (29)

$$C_q = C_{A^T y} - C_{A^T n}. \quad (29)$$

$C_{A^T y}$  and  $C_{A^T n}$  are required for (29). To obtain the derotation phases used in the neighborhood dependent  $\mathbf{A}$  required to calculate these we have two options. We can use the noisy coefficients or use the coefficients resulting from  $E\{\mathbf{x}|\mathbf{y}, m_1\}$ , i.e., those denoised using model 1 only. If the latter option is chosen, the coefficients need to be projected into the range space of the wavelet transform by inverse transforming and transforming back into the wavelet domain, as this improves the accuracy of the estimate.  $C_{A^T y}$  and  $C_{A^T n}$  can then be calculated using (30) with  $\mathbf{v}_k$  as the noisy coefficients  $\mathbf{y}_k$  and noise coefficients  $\mathbf{n}_k$ , respectively

$$C_{A^T v} = \frac{1}{K} \sum_{k=1}^K \mathbf{A}_k^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{A}_k. \quad (30)$$

Finally, an eigenvector/eigenvalue decomposition of  $C_u$  and  $C_q$  is performed and any negative eigenvalues are set to zero to ensure that  $C_u$  and  $C_q$  are positive semidefinite.

#### E. Recalculation of Covariance Information

Ideally, the statistics for each of the models would be generated only from neighborhoods which they are intended to model. For example, when estimating the model 2 covariance matrix  $C_q$ , it is assumed that contributions from neighborhoods not close to discontinuities will be incoherent and  $C_q$  will be dominated by the statistics of the neighborhoods with edges and ridges as the dominant feature. Although this is true, other neighborhoods will effectively contribute noise to the estimation. This effect can be combatted by recalculating the covariance matrices after an initial denoising iteration and weighting the observations by the posterior probability for the model selection variable  $p(m|\mathbf{y})$ . For model 1, this is done using (31) to recalculate  $C_y$  where  $\mathbf{y}_k$  are the noisy coefficients

and  $p_k(m_1|\mathbf{y}_k)$  is the probability of model 1 for neighborhood  $k$

$$\mathbf{C}_{\mathbf{v},m_1} = \frac{1}{\sum_{k=1}^K p_k(m_1|\mathbf{y}_k)} \sum_{k=1}^K \mathbf{v}_k \mathbf{v}_k^T p_k(m_1|\mathbf{y}_k). \quad (31)$$

A similar procedure is applied for  $\mathbf{C}_{\mathbf{A}^T \mathbf{n}}$  and  $\mathbf{C}_{\mathbf{A}^T \mathbf{y}}$  for model 2 using (32) with  $\mathbf{v}_k = \mathbf{n}_k$  and  $\mathbf{v}_k = \mathbf{y}_k$ , respectively

$$\mathbf{C}_{\mathbf{A}\mathbf{v},m_2} = \frac{1}{\sum_{k=1}^K p_k(m_2|\mathbf{y}_k)} \times \sum_{k=1}^K \mathbf{A}\mathbf{v}_k \mathbf{v}_k^T \mathbf{A}^T p_k(m_2|\mathbf{y}_k). \quad (32)$$

Following this recalculation, the algorithm may be re-run with the updated covariance matrices. In this case, any phases required for derotation may be obtained from the (range-space projected) result of the previous denoising iteration.

#### F. Computational Considerations

We now examine the computational implications of including the proposed additional modeling of structural image features for the GSM algorithm.

Define  $\mathbf{C}_m$  as a model dependent covariance matrix such that  $\mathbf{C}_m = \mathbf{C}_u$  for model 1 and  $\mathbf{C}_m = \mathbf{A}\mathbf{C}_q\mathbf{A}^T$  for model 2. Note that in the case of model 1,  $\mathbf{C}_m$  is fixed for each subband whereas for model 2 it is dependent on the neighborhood. For (16), (17), (20), and (21), calculation of  $(z\mathbf{C}_m + \mathbf{C}_n)^{-1}$  is required. Without manipulation, this would require an inversion for each discretization of  $z$  for each neighborhood.

Let  $\mathbf{C}_n = \mathbf{S}\mathbf{S}^T$ , where  $\mathbf{S}$  is the symmetric square root of  $\mathbf{C}_n$ , which can be calculated from the eigenvector/eigenvalue decomposition of  $\mathbf{C}_n$ . Let  $\{\mathbf{Q}_m, \mathbf{\Lambda}_m\}$  be the eigenvector/eigenvalue decomposition of  $\mathbf{S}^{-1}\mathbf{C}_m\mathbf{S}^{-T}$ . Consider (33) and (34)

$$\begin{aligned} z\mathbf{C}_m + \mathbf{C}_n &= z\mathbf{C}_m + \mathbf{S}\mathbf{S}^T \\ &= \mathbf{S}(z\mathbf{S}^{-1}\mathbf{C}_m\mathbf{S}^{-T} + \mathbf{I})\mathbf{S}^T \\ &= \mathbf{S}\mathbf{Q}_m(z\mathbf{\Lambda}_m + \mathbf{I})\mathbf{Q}_m^T\mathbf{S}^T \quad (33) \\ (z\mathbf{C}_m + \mathbf{C}_n)^{-1} &= \mathbf{S}^{-T}\mathbf{Q}_m(z\mathbf{\Lambda}_m + \mathbf{I})^{-1}\mathbf{Q}_m^T\mathbf{S}^{-1}. \quad (34) \end{aligned}$$

Equations (33) and (34) show that for model 1, all of the inversions and eigenvector/eigenvalue decompositions are independent of  $z$  and need only be done once for each subband. Note that  $z\mathbf{\Lambda}_m + \mathbf{I}$  is diagonal and trivial to invert. However, for model 2, an eigenvector/eigenvalue decomposition of  $\mathbf{S}^{-1}\mathbf{C}_m\mathbf{S}^{-T} = \mathbf{S}^{-1}\mathbf{A}\mathbf{C}_q\mathbf{A}^T\mathbf{S}^{-T}$  needs to be calculated for each neighborhood, since  $\mathbf{A}$  is neighborhood dependent. Nevertheless, we have removed the dependence on the hidden multiplier, so the number of decompositions is independent of the number of values used in the discrete representation of  $z$ .

Apart from the necessary estimation of covariance matrices, the algorithm's computational bottleneck is the eigenvector/eigenvalue decomposition of  $\mathbf{S}^{-1}\mathbf{A}\mathbf{C}_q\mathbf{A}^T\mathbf{S}^{-T}$ . The time taken for the Matlab implementation of the algorithm to run on a  $512 \times 512$  image using a PC with a Pentium IV processor is approximately 2.5 min per denoising iteration compared to 45 s when model 2 is omitted from the algorithm, so the method is quite computationally demanding.

## IV. NEIGHBORHOOD SUPPORT

There is a tradeoff with regard to the size of neighborhood used for GSM based denoising. On one hand, it is desirable that the support of the Wiener filter is large to take advantage of the correlations between coefficients in the redundant transform. However, because the variance of wavelet coefficients can change over a relatively short distance, a larger window can result in an inaccurate estimate for  $p_z(z)$ . This tradeoff is also discussed by Mihçak *et al.* [19].

In recent literature, the neighborhood size is generally selected in an *ad hoc* manner. There are many variations in window selection for variance estimation and GSM based denoising methods. Voloshynovskiy *et al.* use different sized neighborhoods at different scales [20]. Strela *et al.* use differently shaped neighborhoods for different directional subbands with a more primitive form of GSM denoising in an attempt to capture the correlation structure along edges [21]. Mihçak *et al.* determine the size of the neighborhood dynamically for each neighborhood using a technique called the bootstrap method [19]. Finally, for GSM denoising using the steerable pyramid transform [1], Portilla *et al.* hand optimized the neighborhood structure and chose a “ $3 \times 3 + p$ ” neighborhood consisting of the coefficient to be denoised, the eight surrounding coefficients from the same directional subband and the parent coefficient at the same spatial location from the adjacent coarser scale. Note that the inclusion of parent coefficient in the neighborhood is consistent with the “persistence across scale” property described in Section II-C.

#### A. Optimal Fixed Neighborhood

For the denoising algorithm implemented, a number of different neighborhood windows were considered. The optimal fixed neighborhood was found to be a “ $+ + p$ ” neighborhood which contains the coefficient to be denoised, the four directly adjacent coefficients (which together form the shape of a “ $+$ ”) and the parent coefficient. It was found to produce marginally better results than the  $3 \times 3 + p$  neighborhood. A smaller neighborhood makes sense because the DT-CWT is less redundant than the steerable pyramid and has a complex coefficient at each spatial location, whereas the steerable pyramid coefficients are real, so the locations of DT-CWT coefficients are more sparsely populated. Hence, the variance of a DT-CWT coefficient is likely to become decorrelated over fewer coefficients than for the steerable pyramid.

The possibility of using smooth windows, i.e., windows where some coefficients are given less weight than others, has been investigated. This problem is nontrivial. Even if a system is devised to implement a smooth window, there is still the problem of how to assign weights to the coefficients in the neighborhood. The techniques developed are not included in the algorithm here due to their limited benefit, added complexity, and computational requirements.

#### B. Subband Dependent Neighborhoods

In the proposed algorithm, we would like to take advantage of the directional nature of structural features and some types of texture. In particular, we would like to take advantage of the

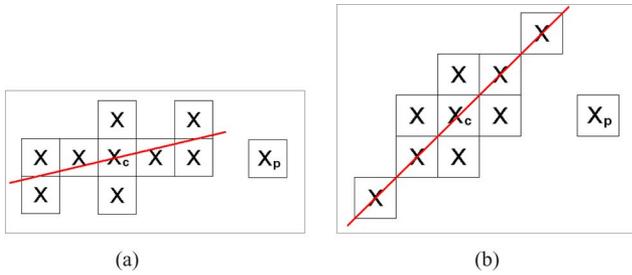


Fig. 3. GSM neighborhoods used at levels 1 and 2 of the proposed algorithm. The line indicates the direction of edges captured by that subband. (a) Near-horizontal subband. (b) Diagonal subband.

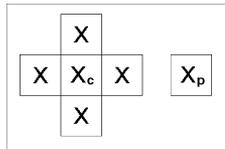


Fig. 4. GSM neighborhood used in the proposed algorithm at level 3 and higher for all directional subbands—“+ + p.”

correlation of derotated coefficients along the length of structural features demonstrated in [2]. We propose varying the size and shape of the neighborhood window depending on the subband orientation and scale. This is based on the assumption that coefficient magnitudes will display greater clustering in the direction of the subband’s orientation and at finer scales where the coefficients have a much smaller support relative to the size of the features. In addition, at finer scales a larger neighborhood assists in determining an accurate estimate for  $p(z)$  in the presence of noise.

The direction and scale dependent neighborhood window used was roughly optimized to improve SNR performance. The neighborhoods used are not exactly optimal but as will be shown in Section IV-E, the overall impact is not insignificant and this demonstrates the stronger relationships between wavelet coefficients in the direction of the subband’s orientation.

Figs. 3 and 4 show the coefficients included in the GSM neighborhoods. The line indicates the direction of edges captured by each of the subbands. At levels 1 and 2, the directional windows in Fig. 3, flipped or rotated appropriately, are used. At coarser scales, the directionally independent “+ + p” neighborhood shown in Fig. 4 is used. As in [1], at the coarsest level denoised the parent is omitted.

## V. ALGORITHM STRUCTURE

The following list summarizes the steps in the proposed denoising algorithm.

- 1) Decompose image using a multiscale transform.
- 2) Calculate model 1 covariance matrices  $C_u$  and  $C_n$  as detailed in Section III-D.
- 3) Calculate expected values and probabilities for model 1,  $E\{\mathbf{x}|\mathbf{y}, z, m_1\}$  and  $p_{\mathbf{y}_s|z, m_1}(\mathbf{y}|z, m_1)$ , using (16) and (20).
- 4) Obtain model 1 coefficient estimates using (15) with  $p(m_1) = 1$ .
- 5) Obtain phases for derotation by projecting model 1 estimates into the range space of the transform.

- 6) Calculate  $C_q$  for each subband as detailed in Section III-D using derotation phases from step 5.
- 7) Calculate expected values and probabilities for model 2,  $E\{\mathbf{x}|\mathbf{y}, z, m_2\}$  and  $p_{\mathbf{y}|z, m_2}(\mathbf{y}|z, m_2)$ , using (17) and (21).
- 8) Combine coefficient estimates using (15) and (18).
- 9) Project estimated coefficients into range space of transform.
- 10) Recalculate  $C_u$  and  $C_q$  as detailed in Section III-E.
- 11) Recalculate expected values and probabilities for both models using (16)–(18).
- 12) Obtain final estimate for coefficients using (15).
- 13) Reconstruct image from the estimated wavelet coefficients via the inverse wavelet transform.

## VI. RESULTS

Results have been obtained using 8-bit grayscale images corrupted with random Gaussian white noise. The algorithm detailed above was implemented with the Q-shift version of the DT-CWT with near-symmetric 13,19 tap filters at level 1 and Q-Shift 14,14 tap filters at higher levels [3]. The wavelet coefficients were denoised at five levels for the  $512 \times 512$  sized images and 4 levels for the  $256 \times 256$  images so that the denoised subbands had at least  $16 \times 16$  coefficients. The image was decomposed to a further 2 levels so that parent and grandparent coefficient phases were available for use in derotation. 20 discrete values were used to define the prior for  $z$ .

To illustrate the effects of the novel components—directionally dependent neighborhoods and the dual model framework—comparisons are made with a standard GSM algorithm (model 1 only and a fixed neighborhood size) as described in [1] but implemented with the DT-CWT and with all other parameters, such as the prior for  $z$  and covariance estimation, identical to the dual model algorithm. Comparisons are made with the SNR results given in [1] to ensure our GSM implementation is adequate and to allow comparison with the best published results.

### A. Model Selection Results

First, we verify that the model selection framework is operating as intended and that the Bayesian framework is indeed selecting model 2 for neighborhoods near structural features. Fig. 5 shows the value of  $p(m_2|\mathbf{y})$  after the first denoising iteration (after step 9) and second denoising iteration for a single wavelet subband of the *Lena* image. We see that at major structural features aligned with the subband’s orientation the value of  $p(m_2|\mathbf{y})$  is close to unity meaning model 2 is selected as intended. Notice that in flat areas the criterion in (23) forces  $p(m_2|\mathbf{y})$  to zero.

### B. Error Analysis

Next, we illustrate the improvement achieved by the new modeling and verify that the dual model algorithm offers improvement in areas near structural discontinuities as it was designed. To do this, we plot two images displaying the improvement and degradation of the proposed algorithm relative to standard GSM denoising. We quantify the improvement of the algorithm proposed here over standard GSM denoising

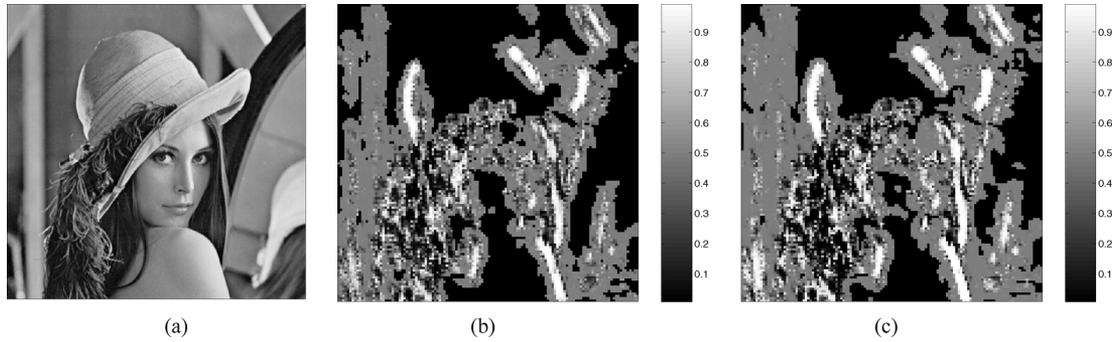


Fig. 5. Model selection results for the *Lena* image. Value of  $p(m_2|y)$  for directional subband 4 ( $105^\circ$ ) at level 2. The noise standard deviation is 20. (a) Original image. (b) *A posteriori* model 2 probabilities after first iteration. (c) *A posteriori* model 2 probabilities after second iteration.

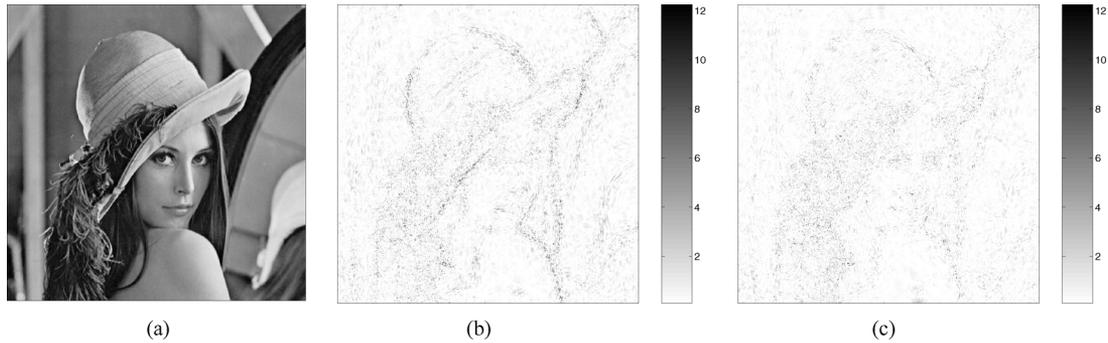


Fig. 6. Error analysis of proposed algorithm compared to standard GSM denoising for the *Lena* image. The noise standard deviation is 25. (a) *Lena* image. (b) Improvement of proposed algorithm over standard GSM denoising. (c) Degradation of proposed algorithm compared to standard GSM denoising.

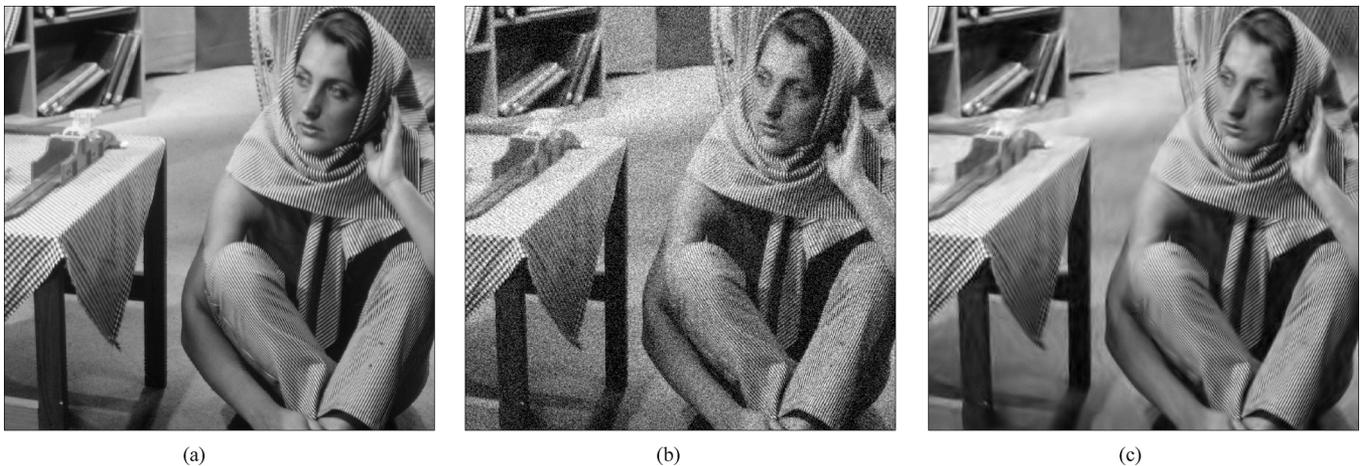


Fig. 7. Denoising results for the *Barbara* image. The noise standard deviation is 25. (a) Clean image. (b) Noisy image. (c) Denoising algorithm proposed here.

as the difference in absolute error with negative values set to zero, i.e.,  $\text{imp} = \max(|e_s| - |e_p|, 0)$ , and the degradation as  $\text{deg} = \max(|e_p| - |e_s|, 0)$ .  $e_p$  is the error of the dual model algorithm proposed here and  $e_s$  is the error for standard GSM denoising implemented with the DT-CWT using only model 1 and a fixed neighborhood size. Fig. 6 shows these comparisons for the *Lena* image plotted as an inverted grayscale image. The proposed algorithm performs better at structural features than the standard GSM method, particularly in areas directly next to discontinuities. Note that it is not better for every pixel as the improvement is based only on improved statistics.

### C. Visual Analysis

Fig. 7 shows clean, noisy, and denoised versions of the *Barbara* image using the proposed algorithm. Fig. 8 shows a close up comparison of the proposed algorithm with that implemented using standard GSM denoising. Ringing artifacts are reduced and edges tidied.

Some noise suppression algorithms have the tendency to smooth discontinuities. The specific modeling of discontinuities in the proposed algorithm reduces ringing artifacts near discontinuities as well as sharpening edges. This is demonstrated

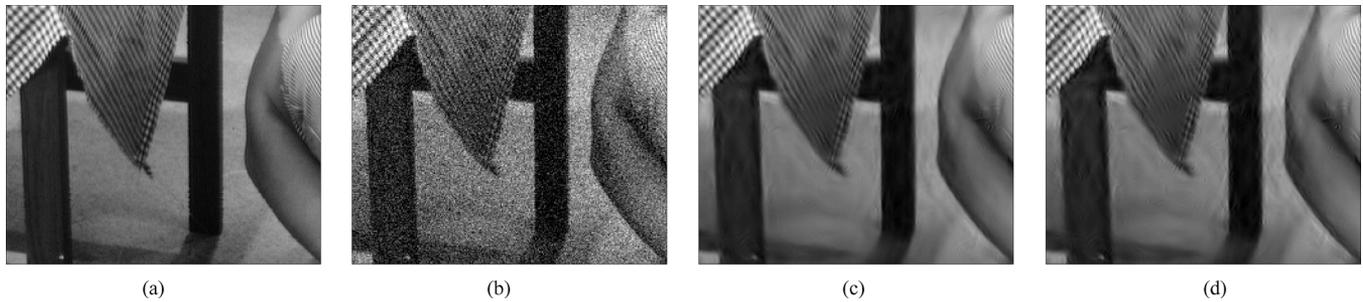


Fig. 8. Comparative denoising results for the *Barbara* image. (a) Clean image. (b) Noisy image. (c) Denoising algorithm proposed here. (d) Standard GSM denoising.

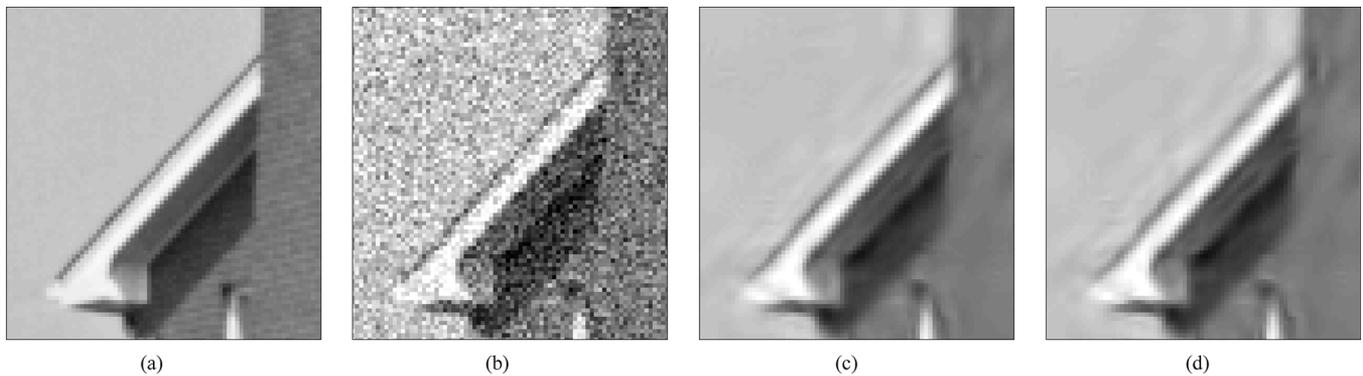


Fig. 9. Comparative denoising results for the *House* image. The noise standard deviation is 25. (a) Clean image. (b) Noisy image. (c) Denoising algorithm proposed here. (d) Standard GSM denoising.

in Fig. 9 where the algorithm incorporating interscale phase relationships produces a sharper image than standard GSM for the *House* image corrupted with noise of standard deviation 25. Note the relative widths of the edge and ridge features that make up the roof in the estimates in Fig. 9(c) and (d).

In most cases, ringing is not eliminated entirely but it is significantly reduced and importantly the improvement comes with the sharpening of edges rather than softening. The improvement at edges is subtle and difficult to see on paper. A demonstration of the results is provided at <http://www-sigproc.eng.cam.ac.uk/~ngk/denoise.zip>.

#### D. SNR Results

In this section, a numerical assessment of the algorithm's performance is presented and compared to other state-of-the-art approaches. Note that improvement in traditional SNR terms will be limited as the algorithm focuses primarily on improving structural features, which do not always constitute a large proportion of the image area. For this reason, an adapted measure of SNR improvement focussing on the areas affected by the algorithm is also provided in this section.

Results were obtained for four 8-bit grayscale images each corrupted with random white noise at different levels. This was repeated for eight different noise samples for each image and noise level. The images used to obtain these results are identical to those used to produce the results obtained by Portilla and colleagues and additional results are given for the *Peppers* image without zero edge pixels as detailed in Section II-F. Input PSNR is defined as  $10 \log(255/\sigma_s)$  dB, where  $\sigma_s$  is the standard de-

viation of the noise. The noise standard deviations used ranged from 5 to 50.

*Standard SNR Improvement Measure:* The results of the standard GSM modeling implemented using the DT-CWT are very similar to those published by Portilla *et al.* in [1], as would be expected. The differences may be due to the different basis functions used including the lower number of oriented subbands, handling of edges and the discrete parameterization of the hidden multiplier. The results are presented in Fig. 10. The proposed algorithm provides consistent improvement on standard GSM denoising implemented with the DT-CWT and that implemented by Portilla *et al.* Table II gives results for the algorithm proposed here for a broader range of PSNR inputs.

*Adapted SNR Improvement Measure:* SNR improvement is a measure of the fractional reduction in error energy achieved by the algorithm. Therefore, in evaluating the effects of the proposed algorithm on structural features, it is not appropriate to consider areas of the image not altered by the new algorithm. Consider that if only part of an image is restored using the new approach, the SNR improvement for the whole image will never drop below a certain level dictated by the areas remaining unchanged.

The DT-CWT is an energy preserving transform. The image domain error energy is equal to that of the wavelet coefficients, provided they are in the range-space. To better evaluate the improvements of the proposed modeling at discontinuities we measure SNR improvement of range-space projected wavelet coefficients multiplied by the weight given to the new modeling (model 2) for each coefficient, i.e.,  $p(m_2|\mathbf{y})$ . Scaling coefficients, which are left unaltered before range-space projection,

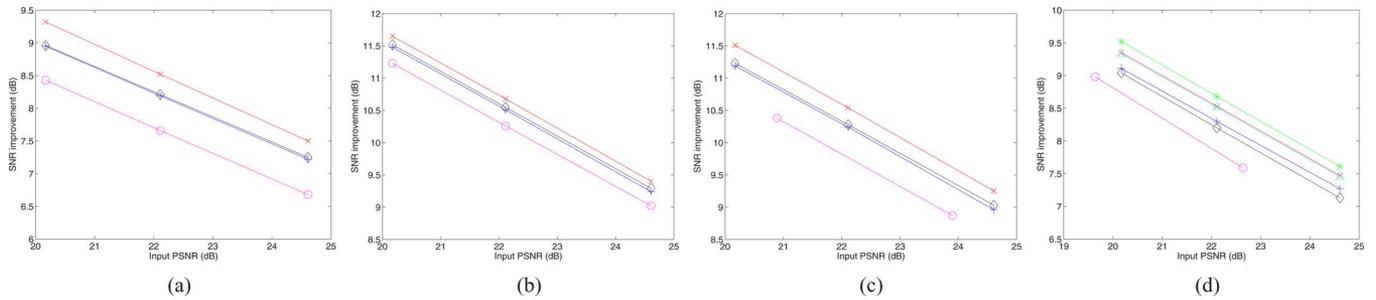


Fig. 10. SNR improvement versus input SNR compared to the best known wavelet based denoising methods. Crosses (x): Proposed denoising algorithm. Pluses (+): Standard GSM denoising using the DT-CWT. Diamonds ( $\diamond$ ): GSM denoising using Steerable Pyramid as published in [1]. Circles ( $\circ$ ): Best of the comparisons published in [1] from [22] for *Barbara* and *Lena* and [23] for *House* and *Peppers*. For *Peppers*, we also plot the results for the original image without zero edge pixels. Stars (\*): Proposed denoising algorithm. Triangles ( $\Delta$ ): Standard GSM denoising using the DT-CWT. (a) *Barbara*. (b) *Lena*. (c) *House*. (d) *Peppers*.

TABLE II  
OUTPUT PSNR OF PROPOSED DENOISING ALGORITHM (DECIBELS)

Image	Standard deviation of noise / Input PSNR (dB)					
	5/34.15	10/28.13	15/24.61	20/22.11	25/20.17	50/14.15
Barbara	37.86	34.20	32.11	30.63	29.49	26.00
House	38.81	35.48	33.84	32.65	31.68	28.46
Lenna	38.46	35.67	34.00	32.79	31.82	28.75
Peppers	37.68	34.22	32.22	30.79	29.69	26.32

TABLE III  
SNR IMPROVEMENT OVER STANDARD GSM DENOISING  
USING ADAPTED SNR METRIC (DECIBELS)

Image	Standard deviation of noise / Input PSNR (dB)					
	5/34.15	10/28.13	15/24.61	20/22.11	25/20.17	50/14.15
Barbara	0.20	0.43	0.55	0.60	0.62	0.60
House	0.32	0.49	0.57	0.60	0.61	0.62
Lenna	0.14	0.26	0.31	0.34	0.35	0.37
Peppers	0.14	0.44	0.32	0.37	0.40	0.44

TABLE IV  
EFFECTS OF VARIOUS CHANGES TO PROPOSED ALGORITHM

Change	++p window	Single iteration	Oracle derotation phase
Average SNR difference	-0.15dB	-0.05dB	+0.04dB

are not included. The SNR improvements over standard GSM denoising obtained using this metric are displayed in Table III.

### E. Miscellaneous Results

We now look at the effect of various changes to the proposed algorithm. The first change is using a ++p neighborhood at all levels and directional subbands. The second change is to restrict the algorithm to a single denoising run so that the covariance matrices are not recalculated using the posterior neighborhood probabilities, i.e., steps 10–13 in Section V are skipped. The final change considered is that of using “oracle” phases, i.e., the phases of the clean coefficients, to derotate the coefficients for model 2. This should indicate how much of the power of derotated coefficients is lost by using an approximation of the parent phase in our denoising algorithm.

Table IV contains the average impact on SNR of each of the changes averaged over the *Barbara*, *House*, *Lena*, and *Peppers* images for a noise standard deviation of 25. Note that the average SNR improvement offered by the proposed algorithm over standard GSM denoising with a directionally independent neighborhood for these images at this noise level is 0.26 dB.

The effect of using orientation and level dependent windows is a considerable 0.15 dB, meaning the directionally dependent neighborhood is key to realising the full potential of the algorithm. However, this figure overstates the independent contribution of this component. If the directionally dependent neighborhood is used with the standard, single model GSM denoising algorithm, the effect is an average improvement of only 0.07 dB. The improvement is much larger for the *Barbara* image, which has large areas of periodic texture, and is on average only 0.03 dB for the other images. The insight gained here is that the phase invariance offered by derotated coefficients allows the dual model algorithm to take advantage of correlations with neighboring coefficients further from the central coefficient at finer scales. The high dependence of the phase of DT-CWT coefficients on the relative position of nearby edges means this is much less true for standard DT-CWT coefficients near multiscale features.

The effect of performing the second denoising iteration is small but useful. Surprisingly little improvement is lost by using an approximation for the parent phases for derotation.

## VII. CONCLUSION

A denoising method based on GSM modeling of the wavelet coefficients of a shift-invariant, directionally selective transform has been presented. The proposed algorithm reduces artifacts near edges while maintaining edge sharpness.

Image processing algorithms designed specifically to perform well at discontinuities often suffer from impaired performance in areas of images not suited to the algorithm. Characterization using derotated wavelet coefficients is not suited to all images or all regions of images but the adaptive model selection framework used in the denoising method proposed here ensures that it does not degrade areas dominated by features more suited to characterization using standard complex wavelet coefficients while offering good improvement near discontinuities.

## VIII. FUTURE WORK

It may be possible to use information obtained at coarser scales to help define the model probabilities at finer scales. The presence of an edge feature at a given scale should increase the probability of a ridge at a finer scale. Similarly the presence of an edge should indicate the presence of two ridges at finer scales.

This information should result in more accurate model probability estimates.

Potentially, the most promising area for future work using interscale phase relationships is in deconvolution and other inverse problems. Deblurring algorithms commonly encounter problems in reconstructing discontinuities as the observations have often lost high frequency information and there is usually a trade-off to be made between over-smoothing of edges and the amplification of noise and ringing artifacts. By encouraging the correct relationships between wavelet coefficients at adjacent levels, it may be possible to constrain the inversion such that ringing is reduced and edges sharpened as has been done for denoising.

Finally, the model selection framework developed has potential for integrating better models for other parts of the images with different statistical characteristics. For example, it may be possible to distinguish between areas of periodic texture and areas where the texture is more random.

## REFERENCES

- [1] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using Gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [2] M. A. Miller and N. G. Kingsbury, "Image modeling using interscale phase properties of complex wavelet coefficients," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1491–1499, Sep. 2008.
- [3] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *J. Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.
- [4] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
- [5] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [6] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sep. 1996, vol. I, pp. 379–382.
- [7] L. C. Parra, C. Spence, and P. Sajda, "Higher-order statistical properties arising from the non-stationarity of natural signals," *Adv. Neural Inf. Process.*, vol. 13, pp. 786–792, Dec. 2000.
- [8] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Comput. Harmon. Anal.*, vol. 11, pp. 89–123, 2001.
- [9] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc. SPIE, 44th Annu. Meet.*, Denver, CO, Jul. 1999, vol. 3813, pp. 188–195.
- [10] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [11] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
- [12] H.-Y. Gao and A. Bruce, "Waveshrink with firm shrinkage," *Statist. Sin.*, vol. 7, pp. 855–874, 1997.
- [13] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Roy. Soc. Lond.*, vol. 357, pp. 2543–2560, Sep. 1999.
- [14] H.-Y. Gao, "Wavelet shrinkage denoising using the non-negative garrote," *J. Comput. Graph. Staist.*, vol. 7, pp. 469–488, 1998.
- [15] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, Sep. 2001.
- [16] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [17] J. A. Guerrero-Colon and J. Portilla, "Two level adaptive denoising using Gaussian scale mixtures in overcomplete oriented pyramids," in *Proc. IEEE Int. Conf. Image Processing*, Genova, Italy, Sep. 2005, pp. 105–108.
- [18] M. A. Miller, "Multiscale techniques for imaging problems," Ph.D. dissertation, Dept. Eng., Cambridge Univ., Cambridge, U.K., Sep. 2006.
- [19] M. K. Mihçak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Phoenix, AZ, Mar. 1999, vol. 6, pp. 3253–3256.
- [20] S. Voloshynovskiy, O. Koval, and T. Pun, "Wavelet-based image denoising using non-stationary stochastic geometrical image priors," in *Proc. ISIT/SPIE's Annu. Symp. Electronic Imaging 2003: Image and Video Communications and Processing V*, Santa Clara, CA, Jan. 2003, pp. 675–687.
- [21] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising using a local Gaussian scale mixture model in the wavelet domain," in *Proc SPIE 45th Annual Meeting*, San Diego, CA, Jul. 2000, pp. 363–371.
- [22] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 438–441, Nov. 2002.
- [23] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field *a priori* model," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 549–565, Apr. 1997.



**Mark Miller** (M'06) received an honors degree in electrical and electronic engineering from the University of Canterbury, New Zealand, in 2002, and the Ph.D. degree in multiscale image processing from Trinity College, Cambridge, U.K.

He worked as a Research Assistant at the University of Canterbury developing a synthetic aperture sonar motion correction algorithm before completing the Ph.D. degree.



**Nick Kingsbury** (M'86) received the honors degree and the Ph.D. degree in electrical engineering from the University of Cambridge, Cambridge, U.K., in 1970 and 1974, respectively.

From 1973 to 1983, he was a Design Engineer and, subsequently, a Group Leader with Marconi Space and Defence Systems, Portsmouth, U.K., specializing in digital signal processing and coding as applied to speech coders, spread spectrum sat-comms, and advanced radio systems. Since 1983, he has been a Lecturer in communications systems and image processing at the University of Cambridge and a Fellow of Trinity College, Cambridge. He was appointed to a readership in signal processing in 2000 and to Professor of signal processing in 2007. He is now Head of the Signal Processing Group, Department of Engineering, Cambridge University.