Natural image modeling using complex wavelets

Andre Jalobeanu, Laure Blanc-Feraud and Josiane Zerubia

aUSRA/RIACS, NASA Ames Research Center, Moffett Field, CA, USA
bCNRS/INRIA/UNS, INRIA Sophia Antipolis, France

ABSTRACT

We propose to model satellite and aerial images using a probabilistic approach. We show how the properties of these images, such as scale invariance, rotational invariance and spatial adaptivity lead to a new general model which aims to describe a broad range of natural images. The complex wavelet transform initially proposed by Kingsbury is a simple way of taking into account all these characteristics. We build a statistical model around this transform, by defining an adaptive Gaussian model with interscale dependencies, global parameters, and hyper-priors controlling the behavior of these parameters. This model has been successfully applied to denoising and deconvolution, for real images and simulations provided by the French Space Agency.

Keywords: Complex wavelets, image modeling, hierarchical Bayesian inference, denoising, deblurring

1. INTRODUCTION

There are two topics presented here, the modeling of natural remote sensing images, and their application to solve ill-posed inverse problems. The images we deal with are highly complex and we do not attempt to model them completely. They are defined over high-dimensional spaces which are difficult to approach. Through experimental study, it is possible to project these images onto lower-dimensional spaces, making them more convenient to handle. The modeling process can be seen as finding the best projection, in the sense that it provides us with a good understanding of the observed phenomena, and therefore a good representation. However, this projection also has to provide the shortest description, so that the results of the projection are accessible (it is not realistic to model each pixel individually). Furthermore, we should not forget our final goal, which is denoising in the present work. Thus the model we construct is well suited for the problem to solve.

In Sect. 2, we first recall how to build a complex wavelet transform. In Sect. 3 we study the properties of natural images, leading to a few general consequences on image modeling. After explaining why we choose to use complex wavelets in Sect. 5, we show how to build a statistical model of the subbands. Finally, in Sect. 7, we propose a new hierarchical multiscale model, combining global parameters with the subband model to form a general image model. In Sect. 8, this is applied to noise removal, and we show how to build complex wavelet packets to also deal with blur, then we give the details and results of the new multiscale deblurring technique.

2. THE COMPLEX WAVELET TRANSFORM

2.1. Implementation

To build a complex wavelet transform (CWT), Kingsbury has developed a quad-tree algorithm, by noting that an approximate shift invariance can be obtained with a real biorthogonal transform by doubling the sampling rate at each scale. This is achieved by computing 4 parallel wavelet transforms, which are differently subsampled. Thus, the redundancy is limited to 4, compared to real shift invariant transforms.

At level $j=1$, the CWT it is simply a non-decimated wavelet transform (using a pair of odd-length filters $h^o$ and $g^o$) whose coefficients are re-ordered into 4 interleaved images by using their parity. This defines the 4 trees $T=A$, $B$, $C$ and $D$. If $a$ and $d$ denote approximation and detail coefficients ($a_0 = X$, the input image), we have:

<table>
<thead>
<tr>
<th>Tree T</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{1/2}$</td>
<td>$a_0$</td>
<td>$a_0 * h^o h^o$</td>
<td>$a_0 * h^o h^o$</td>
<td>$a_0 * h^o h^o$</td>
</tr>
<tr>
<td>$T_{1/2}$</td>
<td>$a_0 * g^o h^o$</td>
<td>$a_0 * g^o h^o$</td>
<td>$a_0 * g^o h^o$</td>
<td>$a_0 * g^o h^o$</td>
</tr>
<tr>
<td>$T_{1/2}$</td>
<td>$a_0 * h^o g^o$</td>
<td>$a_0 * h^o g^o$</td>
<td>$a_0 * h^o g^o$</td>
<td>$a_0 * h^o g^o$</td>
</tr>
<tr>
<td>$T_{1/2}$</td>
<td>$a_0 * g^o g^o$</td>
<td>$a_0 * g^o g^o$</td>
<td>$a_0 * g^o g^o$</td>
<td>$a_0 * g^o g^o$</td>
</tr>
</tbody>
</table>

E-mail: ajalobea@riacs.edu, blancf@sophia.inria.fr, zerubia@sophia.inria.fr
For all other scales \((j > 1)\), the transform involves an additional pair of filters, even-length, denoted \(h^e\) and \(g^e\). There must be a half-sample shift between the trees to achieve the approximate shift invariance. Therefore, different length filters are used for each tree, i.e. we need to combine \(h^e, g^e\) with \(h^o, g^o\), the 4 possible combinations corresponding to the 4 trees:

\[
\begin{array}{cccc}
\text{Tree } T & \text{A} & \text{B} & \text{C} & \text{D} \\
(a^{j+1}_{T+1})_{x,y} & (a^j_A * h^e h^c)_{2x,2y} & (a^j_B * h^e h^o)_{2x,2y+1} & (a^j_C * h^o h^e)_{2x+1,2y} & (a^j_D * h^o h^o)_{2x+1,2y+1} \\
(a^{j+1}_{T-1})_{x,y} & (a^j_A * g^e h^c)_{2x,2y} & (a^j_B * g^e h^o)_{2x,2y+1} & (a^j_C * g^o h^e)_{2x+1,2y} & (a^j_D * g^o h^o)_{2x+1,2y+1} \\
(a^{j+1,2}_{T+1})_{x,y} & (a^{j+1}_A * h^e g^c)_{2x,2y} & (a^{j+1}_B * h^e g^o)_{2x,2y+1} & (a^{j+1}_C * h^o g^e)_{2x+1,2y} & (a^{j+1}_D * h^o g^o)_{2x+1,2y+1} \\
(a^{j+1,3}_{T+1})_{x,y} & (a^{j+1}_A * g^e g^c)_{2x,2y} & (a^{j+1}_B * g^e g^o)_{2x,2y+1} & (a^{j+1}_C * g^o g^e)_{2x+1,2y} & (a^{j+1}_D * g^o g^o)_{2x+1,2y+1} \\
\end{array}
\]

The trees are processed separately, as in a real transform. The combination of odd and even filters depends on each tree. The transform is achieved by a fast filter bank technique, of complexity \(O(N)\). The reconstruction is done in each tree independently, by using the dual filters. To obtain \(a^0\) the results of the 4 trees are averaged. This ensures the symmetry between them, thus enabling the desired shift invariance.

The complex coefficients are obtained by combining the different trees together. If we index the subbands by \(k\), the detail subbands \(d^{j,k}\) of the parallel trees \(A, B, C\) and \(D\) are combined to form complex subbands \(z^{j,k}_+\) and \(z^{j,k}_-\), by the linear transform:

\[
\begin{align*}
z^{j,k}_+ &= (d^{j,k}_A - d^{j,k}_D) + i (d^{j,k}_B + d^{j,k}_C) \\
z^{j,k}_- &= (d^{j,k}_A + d^{j,k}_D) + i (d^{j,k}_B - d^{j,k}_C)
\end{align*}
\]

### 2.2. Invariance properties

The main property of the CWT is the shift invariance, as shown by Kingsbury, \(^2\) i.e. the magnitudes \(|z_\pm|\) are nearly invariant to shifts of the input image. The shift invariance is perfect at level 1, and approximately achieved beyond this level: the transform algorithm is designed to optimize this property.

The impulse responses and the related partitioning of the frequency space are given in Fig. 1. This demonstrates the ability to separate 6 different directions. Compared to real separable transforms, which only define two directions (rows and columns, the diagonals are mixed), it provides near rotational invariance and gives a selectivity which better represents strongly oriented textures.

Image: Left: rough partition of the frequency space corresponding to the CWT and labeling of the different subbands. Right: impulse responses related to the subbands, showing the directional selectivity.

The redundancy of the CWT is 4, and this is independent of the transform depth. The transform remains fully invertible, which makes it usable for modeling tasks.

However, this is not really a complex transform, since it is not based on a continuous complex mother wavelet. Nevertheless, the quad-tree transform has in practice the same properties as a complex transform w.r.t. shifting of the input image.
3. PROPERTIES OF NATURAL IMAGES

In this section, we present axioms and properties of images justified by processing observed natural images. Structural properties are defined, enabling us to describe image contents with general statistical rules. In the next sections we will show that these properties have precise consequences on image modeling.

3.1. Axiom A1: self-similarity

Natural phenomena are often self-similar, from a statistical point of view. As a result, natural images display scale invariance. This means that changing the scale of a signal (by studying only a small part of it for instance) does not change the statistical properties of the signal. When studying a statistic (e.g. the expectation of the power spectrum), the spatial shape of this statistic is not affected by scaling. The self-similarity can be associated with the notion of fractal, widely used to model objects such as mountains or coast lines.

Scale invariance combined with statistical modeling lead to describe the expectation of the spectrum magnitude as decaying like $r^{-q}$, where $r$ is the radial frequency and $q$ a positive parameter. There are no other simple functions that could be used to model the spectrum decay and that are scale invariant at the same time.

To study the scale invariance property, let us plot the decay of the spectrum magnitude with a log-log scale. Any power law should appear as a straight line, with a slope equal to $-q$. Fig. 2 illustrates the agreement between the data and the proposed spectrum model. For an image $X$ this model can be written as follows:

$$F[X]_{uv} = N(0, 1) w_0 r^{-q}$$

where $F$ denotes the Fourier transform and $N(0, 1)$ a white Gaussian noise of variance 1. $q$ is called the fractal parameter, and $w_0$ the energy parameter. For a broad class of natural images, $q$ is ranging from 0.9 to 1.5. It corresponds to describing an image by a Fractional Brownian Motion (FBM).

3.2. Axiom A2: spatial adaptivity

Natural images generally exhibit a strong spatial adaptivity, since they are composed of different textures, homogeneous areas, sharp edges and small features (see Fig. 3 for an illustration). The importance of having a spatially variant model is shown by the failure of stationary approaches to correctly model images, especially when dealing with inverse problems such as denoising or deblurring.

However, taking into account the space-varying characteristics of a natural scene is a difficult task, since it requires to define (and then to estimate) a large number of additional parameters, at least one per pixel. This can inevitably lead to overfitting issues, since the nonstationary parameters can adapt more to the noise than to the underlying signal, if no constraint is put on them. We will show how the use of a prior probability density helps to constrain these parameters and take advantage of the spatial adaptivity.

3.3. Property P1: interscale persistence

When dealing with scale invariant data, it is natural to use a multiresolution decomposition (e.g. wavelets) to analyze them. This enables us to get a sparser representation, and to separate the scale invariance from other properties. There is a strong resemblance between the images of Fig. 3, which represent different subbands of same orientation and different scale (and this remains valid whatever decomposition is used). The underlying spatial structure of the image persists through the different scales, i.e. high/low value of the coefficient magnitudes often lead to high/low values at the next scale. This can be seen as the inability of any multiscale transform to perfectly decorrelate natural images, since such a transform exhibits some interscale redundancy.

3.4. Property P2: intrascale dependencies

Now we consider two different subbands of the same scale, but with different orientations (they are generated by linear filtering with the same filter w.r.t. size and shape, but oriented differently). Fig. 4 illustrates the dependence within the same scale, showing that homogeneous areas and details are found approximately at the same spatial location in the different subbands. Isotropic details, such as smooth areas and round features, are naturally found at all orientations. This shows again that multiscale transforms do not perfectly decorrelate the information contained in natural images.
Figure 2. Log-log plot of the power spectrum as a function of the radial spatial frequency, to illustrate the self-similarity of natural images. First row: city of Amiens © IGN. Second row: two images of Toulouse (left) and two images of Vannes (right) © CNES. All images are 512 × 512 pixels, approximate resolution 1m-2m.

Figure 3. From left to right: image of Amiens (1024 × 1024 pixels) – © IGN, horizontal detail subbands of a real wavelet transform (Daubechies-4) at scales 1, 2 and 3. This illustrates the interscale persistence.
4. CONSEQUENCES ON IMAGE MODELING

4.1. Wavelet representation

Combining Axioms A1 and A2 (self-similarity and nonstationarity) leads to the following image model, which is the simplest that can account for these two constraints:

\[ X \sim (s^0 \ N(0, 1)) \ast \mathcal{F}^{-1} \left[ w_0(\theta) \ r^{-q} \right] \]  

(3)

where \( s^0 \) is a spatially adaptive parameter field and \( N(0, 1) \) a white Gaussian noise of variance 1. This nonstationary FBM is a Gaussian process with a covariance matrix \( \Sigma \). The best basis to represent such a process is the Karhunen-Loeve (K-L) basis. It diagonalizes \( \Sigma \), therefore the image coefficients in this basis are independent. It is optimal, since it provides the minimum mean error when approximating an image by its projection on \( m \) orthogonal basis vectors chosen a priori (this relates to image denoising by thresholding).

For stationary processes, the K-L basis is the Fourier basis. However, the Fourier basis is not well suited to the nonstationarity of natural images. In the case of natural images a wavelet basis is the closest approximation to a K-L basis, since it provides both a frequency and a spatial representation. The fact that it is only an approximation is illustrated by the properties P1 and P2 (residual dependencies). The aim is not only to diagonalize a Gaussian process, but also to provide the shortest description of the useful signal. Modeling can be seen as finding a simple description of nature, and in this case it can be related to image compression. It is well-known that wavelets provide a compact representation of natural images, and are widely used in coding. For all these reasons, we choose to do the modeling in the wavelet domain.

4.2. Heavy-tailed subband distribution

An interesting consequence of the self-similarity is that high order statistics are scale invariant. It means that they obey a function \( f \) of the scale, and the shape of \( f \) is not affected by scaling (it is only multiplied by a scale factor). Eqn. (2) shows how the second order moment decays as \( r^{-q} \) where \( r \) is the inverse of the scale. A generalization of this law is the scale invariance of the wavelet subband histograms (see Fig. 5): they relate to the full coefficient distribution instead of its order \( n \) moments. Then, the marginal distribution of a wavelet coefficient \( \xi_{k,l} \) (where \( k \) is the subband index and \( l \) a spatial index) is modeled as a Generalized Gaussian (GG) distribution,\(^{14, 15} \) which has the remarkable property to be scale invariant:

\[ P(\xi_{k,l} \mid \alpha_k, p) = \frac{p}{2\pi \alpha_k \Gamma(2/p)} e^{-|\xi_{k,l}/\alpha_k|^p} \]  

(4)

where \( p \) is the shape parameter and \( \alpha_k \) the scaling parameter. The smaller the shape parameter, the stronger the nonstationarity. It is possible to prove that \( p = 2 \) corresponds to a simple shift invariant Gaussian model.\(^7, 16\)

It is possible to account for this heavy-tailed marginal when modeling the wavelet coefficients, while still using a spatially adaptive model. We choose to use an adaptive Gaussian model (i.e. with space-varying variances) such that its marginal is a GG distribution, as it will be shown in Sect. 6.1.
4.3. Residual dependencies

Properties P1 and P2 clearly show the inability of any wavelet basis to fully decorrelate the covariance matrices of nonstationary self-similar images. Therefore the residual dependencies have to be taken into account. We can take advantage of these correlations when trying to estimate the true image coefficients from noisy observations, since the noise coefficients are decorrelated if the noise is white and stationary. We propose to use graphical models, such as trees, to do the modeling in a way that allows for a fast inference (therefore we avoid Markov Random Fields). There are various ways of introducing the correlations, as we will explain further.

5. CHOICE OF THE WAVELET BASIS

5.1. Compactness and optimal shape detection

As noticed previously, the image representation needs to be compact. All wavelet transforms are not equivalent from this point of view. The redundancy of the transform has to be minimum. Therefore the decimated transforms are obviously more efficient than the non-decimated, shift invariant ones.

Furthermore, the smoothness properties and the spatial extent of the wavelet are also important – the larger the spatial localization, the worse the frequency selectivity. For instance, a wavelet providing a good spatial, but poor frequency selectivity, will fail to correctly diagonalize a FBM process. Meanwhile, a frequency selective wavelet will not handle highly nonstationary processes very well.

Reducing the dependencies to improve the compactness of the representation is not only achieved by space-frequency trade-offs. The compactness is also related to optimal shape detection. To concentrate the energy of a given feature over the smallest number of wavelet coefficients, the wavelets have to provide optimal detection of this type of feature. To detect a Gaussian for instance, the Gaussian-shaped wavelets are optimal. Fig. 7 shows the approximation error (sum of squared difference between approximated and true images) for two kinds of images, using two sorts of wavelets: Symmlet wavelets are smooth and suitable for natural images, whereas Haar wavelets best represent piecewise constant images. Fig. 6 shows how Haar, Symmlet and Complex wavelets perform locally to detect small features such as oriented edges or textures. Complex wavelets perform best to detect oriented features and textures, thanks to their improved directional selectivity.

Figure 5. Histograms of the details subbands of a wavelet transform (wavelet Symmlet-8, same image as in Fig. 3). Points = data, solid lines = Generalized Gaussian model a) scale 1 b) scale 2 c) scale 3. The histograms for diagonal, vertical and horizontal details are superimposed at each scale.

Figure 6. Areas extracted from vertical detail subbands at scale 2 for different wavelet transforms (cf. Fig 3): a) Haar, b) Symmlet-8, c) Complex wavelets.
Figure 7. Approximation error $E$ as a function of the number of non-null wavelet coefficients (log-log plot): Haar and Symmlet-8 wavelets. The dashed line corresponds to the asymptote $E \propto N^{-1/2}$. a) satellite image of Amiens: the error is bigger with Haar, b) synthetic piecewise constant image: the error is smaller with Haar.

Figure 8. a) Areas extracted from the image of Amiens (cf. Fig 3) shifted 0, 1 and 2 pixels to the right. Images reconstructed only from scale 2 of different wavelet transforms: b) Haar, c) Symmlet-8, d) Complex wavelets.

5.2. Shift invariance

Fig. 8 shows the sensitivity of different wavelet transforms to shifts of the input image. To perform an optimal detection of any feature, a transform should not “miss” the feature because of such shifts. Yet it is the case for real wavelets, since they are decimated: the alignment between feature and wavelet plays a central role in the detection, and because of the decimation, a correct alignment is not always possible. The magnitude of the
complex wavelet coefficients is not sensitive to shifts, which makes the CWT a very good choice when it comes to modeling.

5.3. Rotational invariance

Rotational misalignments can have dramatic consequences on the detection of a smooth edge, as illustrated by Fig. 9. A non-redundant real wavelet transform is not rotation invariant, since rotating the input image produces different coefficient magnitudes, leading to visible artifacts. As with shifts, we do not want the detection quality of the transform to depend on rotations, therefore we choose the complex wavelets.

![Figure 9. a) Area extracted from image of Amiens cf. Fig 3. Images reconstructed only from scale 2 of different wavelet transforms: b) Haar, c) Symmlet-8, d) Complex wavelets.](image)

6. SUBBAND MODELING

Now that we chose the CWT for the multiscale transform, we have all the elements required to build a precise image model. In the following sections, we show how to take into account the properties mentioned above to construct a spatially adaptive model of the complex wavelet subbands, which allows the information to propagate through the scales. The main idea is to use Gaussian mixtures to encode both nonstationarity and interscale dependencies.

6.1. Spatial adaptivity

Using a Generalized Gaussian distribution is certainly not an efficient way of modeling the nonstationarity of images. It usually does not provide analytic expressions when combined with a Gaussian when doing a Bayesian inference. Therefore, we prefer to use Gaussian mixtures, since we can build marginal subband distributions that are very close to the observed ones, whether we use a continuous or discrete mixture.

Within a hierarchical approach, we create hidden variables that the wavelet coefficients depend upon. These hidden variables encode the spatial adaptivity, i.e. their realization is spatially dependent. They are simply the variance of the wavelet coefficients, we denote them $\mathbf{s}_{k,l}$, and $\mathbf{s}$ is the set of variances. We have:

$$P(\xi_{k,l} \mid s_{k,l}) = \frac{1}{2\pi s_{k,l}^2} e^{-|\xi_{k,l}|^2/2s_{k,l}^2}$$

so the CWT coefficients $\xi$ are Gaussian when $s$ are fixed, and their marginal can be heavy-tailed, depending on the spatial distribution of $s$. This is an explicit way of taking into account the spatial adaptivity. Thus, within a hierarchical Bayesian framework, the nonstationary distribution of the variances is estimated first, providing an efficient data-dependent adaptive model of the subbands. This kind of model is more informative than the marginal, which only accounts for the nonstationarity without really taking advantage of it.

We have to define a model on $s$ such that the marginal $P(\xi)$ is a GG with an exponent $p$. It is possible to design a continuous density for each $s_{k,l}$ such that the marginal be close enough to a GG, but it does not lead to analytic expressions when doing inference, and it is not compatible with the interscale dependencies. Therefore we use a discrete Gaussian mixture defined by the parameters $p$ and $\nu$:

$$P(s_{k,l} \mid \rho_k, \nu_k) = \sum_m \rho_k^m \delta(s_{k,l} - \nu_k^m)$$

$\rho$ and $\nu$ are the parameters of the mixture.
Depending on the value of $p$, between 1 ($p = 2$) and 5 ($p = 0.5$) components are needed to ensure a K-L divergence smaller than 0.01 between the mixture and GG distributions. A method to compute the mixture parameters $\rho$ and $\nu$ as functions of the parameters $\alpha$ and $p$ of the GG can be found in Ref. 7, it is based on the EM$^{17}$ estimation method.

### 6.2. Interscale persistence and causality

By observing the joint histograms of detail subbands at two successive scales, we notice that some information propagates through scales. What kind of information really persists? The global shape of the subband magnitudes is conserved, but the coefficients change locally. The phase seems poorly correlated, except for some particular features such as clear geometric shapes, not a major component of natural images.

If we consider the mixture model presented above, we find that the states $s$ propagate through scales (in the sense that they are correlated), while the coefficients $\xi$ are independent given $s$. Fig. 10 shows the difference between direct interscale dependence, which we believe is not appropriate, and dependence through the hidden variables $s$. This is a Hidden Markov Tree (HMT)$^{12,18}$ model, defined by the joint density of the variances at the nodes of the tree corresponding to the wavelet decomposition. The quadtree structure results from the subsampling at each scale, linking a coefficient at scale $j$ to its 4 children at scale $j - 1$.

It is convenient to parametrize the model by transition matrices $\varepsilon$ for each subband. If we denote $\pi(k)$ the parent of the subband $k$, we have:

$$
\varepsilon_{kn}^k = P(s_{k,l} = \nu_{\pi(k)}^n | s_{\pi(k),l} = \nu_{\pi(k)}^n)
$$

(7)

Now we can write the joint variance density given $\varepsilon$, $\nu$ and $\rho$, as well as the joint CWT density given the variances (excluding the root node):

$$
P(s | \varepsilon, \nu, \rho) = \prod_{k,l} \varepsilon_k^k P(s_{k,l} | \rho_k, \nu_k) \quad \text{and} \quad P(\xi | s) = \prod_{k,l} P(\xi_{k,l} | s_{k,l})
$$

(8)

where the conditional densities are defined by (6) and (5).

![Figure 10. a) explicit interscale dependence, c) interscale dependence through hidden variables.](image)

### 6.3. Intrascale dependencies

How do we take into account the property P2, which is related to the dependence between coefficients within the same scale? Again we study the joint histograms of two subbands, and discover that the explicit dependence is not appropriate, while the hidden variable approach leads to the best results. The main drawback of the explicit way of linking the variables together is that it breaks the tree structure, leading to cycles, which prevents from using fast single-pass estimation techniques defined on trees. An alternative is to group coefficients of the same scale and different orientations together, thus forming a vector, and defining the tree structure on these vectors.

We have summarized in Fig. 11 the different ways of handling this dependence. To keep the compatibility with the approaches presented in the previous sections, we propose to model the intrascale dependence by hidden variables. But instead of using different hidden variables within the same scale, linked by dependency relationships, we strongly simplify the model by sharing the same state variable among coefficients of various orientations at same scale and position. This enables us to keep the model previously defined, keeping in mind that $s_{k,l}$ refers to the same variable for all the subbands $k$ of a scale $j$. 

![Figure 11.](image)
7. A HIERARCHICAL BAYESIAN MULTISCALE IMAGE MODEL

The image is described as the result of a hierarchical random process, illustrated by Fig. 12. Each parameter of the model is assumed to be a random variable, therefore we call it a hierarchical Bayesian model. We have described the models for the CWT coefficients, their variances, transition probabilities and the parameters they depend upon. The graphical model helps writing the full joint density of the model:

\[
P(X, \xi, s, \nu, \varepsilon, w_0, q, p) = P(w_0) P(q) P(p) P(X | \xi) P(\xi | s) P(s | \nu, \rho, \varepsilon) P(\varepsilon) P(\nu, \rho | \varepsilon, w_0, q, p)
\]

\[
\text{with } P(w_0) = U_{w_0}[0, w_{0,\text{max}}], P(q) = U_q[0.9, 1.5], P(p) = U_p[0.5, 2]
\]

- \(X\) is the image, obtained by inverse CWT of \(\xi\) (\(P(X | \xi)\) is a Dirac distribution);
- \(\xi\) represents the CWT coefficients following a conditional Gaussian distribution with variances \(s\);
- \(s\) are the coefficient variances following a Markov tree model with parameters \(\varepsilon, \nu\) and \(p\);
- \(w_0\) and \(q\) are the fractal spectrum model parameters;
- \(p\) is the nonstationarity parameter (shape parameter of the subband GG marginals).

We assume uniform priors on the parameters on which we do not have sufficient prior knowledge, such as the transitions. The bounds for the global parameters \(p, w_0\) and \(q\) are derived from experiments on real images.

8. APPLICATIONS TO REMOTE SENSING

8.1. Image denoising

The proposed hierarchical model is very useful when performing image denoising. Accurately modeling images is key to efficient denoising, i.e. easily differentiating the useful signal from noise. The direct problem can be written as \(Y = X + N(0, \sigma^2)\) where \(Y\) is the observed data and \(N\) is a white zero-mean Gaussian stationary
noise, of variance \( \sigma^2 \) for each pixel. Since the CWT is a linear transform, we can write this equation in the wavelet domain:

\[
o_{k,l} = \xi_{k,l} + N(0, \sigma_k^2)
\]

(10)

where \( o \) denotes the noisy wavelet coefficients. The CWT is normalized, therefore we have \( \sigma_k = \sigma \).

Eqn. (10) is the observation equation, representing the direct problem. We have to solve the inverse problem, i.e. getting an estimate of the true coefficients from corrupted coefficients.

We use an empirical Bayesian approach to estimate all the parameters of the model. This is an efficient alternative to existing methods such as EM, since it is not iterative: we follow the graphical model of Fig. 12 to estimate the different parameters in their dependence order. However, we do not fix the values of the key parameters \( s \), we estimate their probability density function instead, which leads to more precise results than fixing the values of \( s \).

The estimator used to compute the denoised wavelet coefficients is the posterior mean (PM), known for its properties well suited to image processing (it is better than the Maximum A Posteriori for instance\(^7\)).

Our approach can be summarized as follows: 1) We estimate \( w_0, q \) and \( p \) from the observed data. 2) We get the Gaussian mixture parameters \( \nu \). 3) We use an independent subband model and \( \hat{\nu} \) to compute \( P_{\text{ind}}(s \mid o) \), which helps estimating \( \hat{\varepsilon} \). 4) The HMT approach with \( \hat{\varepsilon} \) is used to estimate \( P(s \mid o) \). 5) Finally, these marginals are taken as an adaptive prior model of \( s \) to denoise \( o \), thus giving the PM estimate \( \hat{\xi} \).

The denoising algorithm consists of the following steps:

- **Direct CWT:** \( o = \text{CWT}[Y] \)
- **Initialization:**
  - Estimate \( \hat{p} \) (using first and second order moments of the coefficients magnitude, for several subbands\(^7\)).
  - Then compute the parameters \( \hat{\nu}_0 \) using \( \hat{p} \).
  - Estimate \( \hat{\alpha}_k \) for each subband \( k \) (using the first order moment only\(^7\)).
  - Multiply the vector \( \hat{\nu}_0 \) by \( \hat{\alpha}_k \) to obtain the estimates of \( \nu \) for each subband, i.e. \( \hat{\nu}_k \).
- **Estimate the transition matrices:**
  - First compute the probabilities \( P_{\text{ind}}(s_{k,l} = \hat{\nu}_k^m \mid o_{k,l}) \) as though the subbands were independent.
  - Using Bayes rule we have \( P(s_{k,l} = \hat{\nu}_k^m \mid \xi_{k,l}, o_{k,l}) = P(o_{k,l} \mid \xi_{k,l}) P(\xi_{k,l} \mid s_{k,l}) P_{\text{ind}}(s_{k,l} = \hat{\nu}_k^m \mid o_{k,l}) \),
  - then this joint density is integrated over \( \xi_{k,l} \).
  - Then use \( P_{\text{ind}} \) to compute \( \hat{\varepsilon} \):

\[
\hat{\varepsilon}_k^{\text{pm}} = \frac{\sum_{i \in T_k} P_{\text{ind}}(s_{k,i} = \hat{\nu}_k^m \mid o_{k,i}) P_{\text{ind}}(s_{\pi(k),i} = \hat{\nu}_{\pi(k)}^n \mid o_{k,i})}{\sum_{i \in T_k} P_{\text{ind}}(s_{\pi(k),i} = \hat{\nu}_{\pi(k)}^n \mid o_{k,i})}
\]

(11)

- **State probability estimation step:**
  - Estimation of \( P(s_{k,l} = \hat{\nu}_k^m \mid o_{k,l}) \) using the backward-forward (Baum-Welch\(^7\)) algorithm.
  - We use \( \hat{\varepsilon}, \hat{\nu} \) and we set \( \xi \equiv o \) (“complete data” approximation).
- **Denoising step:**
  - We use the posterior mean (PM) estimate \( \hat{\xi}_{k,l} = E[\xi_{k,l} \mid o_{k,l}, \hat{\nu}_k] \)
  - To compute this estimate, we need to integrate \( P(\xi_{k,l} \mid o_{k,l}, \hat{\nu}_k) \) over \( s_{k,l} \), which can be rewritten using Bayes rule twice, as \( P(o_{k,l} \mid \xi_{k,l}) P(\xi_{k,l} \mid s_{k,l}) P(s_{k,l} \mid \hat{\nu}_k) \)
  - To take advantage of the spatial adaptivity, we use \( P(s_{k,l} \mid o_{k,l}, \hat{\nu}_k) \) instead of \( P(s_{k,l} \mid \hat{\nu}_k) \):

\[
\hat{\xi}_{k,l} = o_{k,l} \sum_m P(s_{k,l} = \hat{\nu}_k^m \mid o_{k,l}) \frac{(\hat{\nu}_k^m)^2}{(\hat{\nu}_k^m)^2 + \sigma_k^2}
\]

(12)

- **Inverse CWT:** \( \hat{X} = \text{CWT}^{-1}[\hat{\xi}] \)
8.2. Image deconvolution and complex wavelet packets

The degradation model is now represented by the equation $Y = h \star X + N(0, \sigma^2)$. The $\star$ represents a circular convolution. The Point Spread Function (PSF) denoted by $h$ is a positive and normalized kernel.

A few authors, such as Donoho et al., proposed to deblur signals in a wavelet-vaguelette basis. Rougé, then Mallat and Kalifa, proposed to denoise images after a deconvolution without regularization. All these techniques consist of two steps: deconvolution by a pseudo-inverse or generalized inverse filter, then denoising in a wavelet packet basis. This type of method is not iterative and provides a very fast implementation.

To denoise the image deblurred without regularization, a sparse representation has to be chosen in order to separate the signal from the noise as well as possible. A representation is said to be sparse if it approximates the signal with a small number of parameters. We have shown previously that the CWT provides such a sparse decomposition.

The noise amplified by the deconvolution process is colored. Furthermore, the coefficients of this noise are not independent in a standard wavelet basis. Thus, the basis must adapt to the covariance properties of the noise. The covariance matrix should be "nearly diagonal" in the basis, to decorrelate the noise coefficients as much as possible. The Fourier basis achieves such a diagonalization, but it does not provide a sparse representation and therefore is not suitable for any thresholding method. A good compromise is to use a wavelet packet basis, since it nearly realizes the two following essential conditions: the signal representation is sparse, and the noise covariance operator is nearly diagonalized.

In the next section, we show how to build a complex wavelet packet transform which enables us to represent both signal and deconvolved noise correctly.

8.2.1. Implementation

We have extended the original CWT by applying the filters $h$ and $g$ on the detail subbands, thus defining the complex wavelet packet transform (CWPT). This enables us to preserve the properties of the original complex wavelet transform.

The subbands are now indexed by $(p,q)$ for each tree $T$. We have:

$$
\begin{array}{cccc}
\text{Tree } T & A & B & C & D \\
(d_T^{p+1,2p+1}h^{e})_{x,y} & (d_A^{p,q} \ast h^{e}h^{o})_{2x,2y} & (d_B^{p,q} \ast h^{o}h^{e})_{2x,2y+1} & (d_C^{p,q} \ast h^{o}h^{e})_{2x+1,2y} & (d_D^{p,q} \ast h^{o}h^{o})_{2x+1,2y+1} \\
(d_T^{p+1,2p+1}g^{e})_{x,y} & (d_A^{p,q} \ast g^{e}h^{o})_{2x,2y} & (d_B^{p,q} \ast g^{e}h^{o})_{2x,2y+1} & (d_C^{p,q} \ast g^{e}g^{o})_{2x+1,2y} & (d_D^{p,q} \ast g^{o}g^{o})_{2x+1,2y+1} \\
(d_T^{p+1,2p+1}g^{o})_{x,y} & (d_A^{p,q} \ast g^{o}g^{o})_{2x,2y} & (d_B^{p,q} \ast g^{o}g^{o})_{2x,2y+1} & (d_C^{p,q} \ast g^{o}g^{o})_{2x+1,2y} & (d_D^{p,q} \ast g^{o}g^{o})_{2x+1,2y+1} \\
\end{array}
$$

8.2.2. Invariance properties

Compared to the original complex wavelet transform, which only separates 6 directions, the directional selectivity is highly improved. With the chosen tree, up to 26 directions can be separated, as illustrated in Fig. 13 which shows the impulse responses related to the different subbands and the induced frequency plane partitioning.

Instead of dividing an approximation space which does not define any new orientation, the wavelet packet decomposition processes the detail subbands, which are strongly oriented. Each detail subband at level 1 isolates an area of the frequency space defined by a mean direction and a dispersion, enabling one to select a range of directions around a given orientation. If the subband is decomposed into 4 new subbands, it means that the corresponding frequency area is split into 4 new areas, which define new orientations.

The shift invariance can be demonstrated in the following way: we perform the CWP transform of a step function (1D signal), we keep only one wavelet packet subband at level 2 (the other subbands are set to zero), and we do the inverse transform. This is done for 4 possible shifts of the input signal, and we remark the approximate shift invariance of the output, for both wavelet packet subbands (see Fig. 14).
8.2.3. The proposed deblurring method

We assume that the CWPT provides an efficient near-diagonalization of the deconvolved noise, i.e. the noise wavelet coefficients are independent. Then we consider the observation equation (10), and we can simply apply the denoising method described above by setting the correct value of the noise variance $\sigma^2$ for each subband.

We compute $\sigma^2_k$ by using the impulse responses of the subbands, denoted $W^k$. To obtain $W^k$, we take the CWPT of a discrete Dirac and we keep only the subband $k$, all the others being set to zero, then we perform the inverse transform. The mean power spectrum of the deconvolved noise is given by $|1/F[h]|^2$ where $F$ denotes the Fourier transform. Then, for subband $k$ we have:

$$\sigma^2_k = \sigma^2 \sum_{ij} \left| \frac{F[W^k]_{ij}}{F[h]_{ij}} \right|^2 \quad (13)$$

The proposed deblurring technique consists of the following steps:

- **Deconvolution without regularization:**
  Deblur the image $Y$ using a pseudoinverse or generalized inverse filter.

- **Computation of the noise variances:**
  Use Eqn. (13) to compute the variance of the deconvolved noise for each subband.

- **Denoising step:**
  Use the algorithm defined in Section 8.1.

The images presented in this paper are degraded by an exponential transfer function, equivalent to the following PSF model $h_{ij} = [1 + (i^2 + j^2)]^{-1}$. The observed image is blurred and corrupted by noise of variance $\sigma^2 = 2$ and is shown in Fig. 15 a. The deblurred image using the new algorithm is shown in the same figure.

The following table shows the SNR improvement (in dB) between observed and restored images for different models using complex wavelet packets. 1: Simple model with a single mixture component, i.e. Gaussian model.
2: Independent coefficients with a 3-component mixture. 3: The full proposed model with interscale dependencies through a HMT. This one always performs the best.

<table>
<thead>
<tr>
<th>noise variance</th>
<th>1. Gauss</th>
<th>2. Independent</th>
<th>3. Dependent (HMT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2 = 2$</td>
<td>5.35</td>
<td>5.40</td>
<td>5.68</td>
</tr>
<tr>
<td>$\sigma^2 = 8$</td>
<td>3.24</td>
<td>3.75</td>
<td>3.90</td>
</tr>
</tbody>
</table>

Figure 15. a) Aerial image of the city of Rennes ©IGN, artificially blurred and corrupted by noise. b) The result of the deconvolution algorithm using the hierarchical multiscale model.

9. CONCLUSION

We have proposed a new probabilistic image model, which combines the invariance properties of the complex wavelet transform with the hierarchical Bayesian approach to provide an efficient description of natural images. This model is based on properties derived by analyzing statistics of remote sensing data.

Applied to image deblurring, the proposed method is superior to other competing algorithms on satellite images, since it is computationally more efficient, more accurate and fully automatic. We use a posterior mean to estimate the true image coefficients in the wavelet domain. The model parameters are determined by empirical Bayesian estimation, starting from global parameters and ending with spatially adaptive ones, which allows for a nonstationary approach. Both inter- and intrascale dependencies are taken into account, which helps computing the spatially adaptive variance of the wavelet coefficients while increasing the robustness to the noise.

The proposed model could be generalized further by introducing multiple wavelet bases, thus improving the representation of small features not well detected by complex wavelets. The hierarchical Bayesian framework is suitable for such an extension, and an empirical approach could be used to perform the denoising with different bases, then estimating an optimal model map, and finally combining all the results together.

ACKNOWLEDGMENTS

The authors would like to thank Nick Kingsbury (from the Signal Processing Group, Dept. of Eng., University of Cambridge) for the complex wavelets source code and collaboration, the French Mapping Institute (IGN) and the French Space Agency (CNES) for providing the images.
REFERENCES