MULTISCALE EDGE GRAMMARS FOR COMPLEX WAVELET TRANSFORMS

Justin K. Romberg, Hyeokho Choi, and Richard G. Baraniuk

Dept. of Electrical and Computer Engineering
Rice University
Houston, TX 77005, USA

ABSTRACT

Wavelet domain algorithms have risen to the forefront of image processing. The power of these algorithms is derived from the fact that the wavelet transform restructures the image in a way that makes statistical modeling easier. Since the edge singularities in an image account for the most important information, understanding how edges behave in the wavelet domain is the key to modeling. In the past, wavelet-domain statistical models have codified the tendency for wavelet coefficients representing and edge to be large across scale. In this paper, we use the complex wavelet transform to uncover the phase behavior of wavelet coefficients representing an edge. This allows us to design a hidden Markov tree model that can discriminate between large magnitude wavelet coefficients caused by texture regions in the signal, and ones caused by edges.

1. INTRODUCTION

Over the past few years, wavelet-domain algorithms have redefined the state-of-the-art in statistical image processing. The power of these algorithms is rooted in the fact that wavelets provide a sparse representation for real-world images. A sparse representation make it easier to describe images; the wavelet transform restructures an image in a way that makes it easier to characterize statistically.

The sparsity of the wavelet coefficients of real-world images comes from the fact that (loosely speaking) large wavelet coefficients are caused only by singularities in the signal, and singularities make up only a small portion of an image.

However, the edge structure plays the most important role in our perception of an image [1]. That is, even though there are relatively few large wavelet coefficients representing edges, they carry most of the important information. Understanding how edges behave in the wavelet domain is key to understanding how images behave in the wavelet domain.

In this paper, we examine the *multiscale edge grammar* that describes the relationships between wavelet coefficients representing an edge. This grammar describes the interrelationships required of the wavelet coefficients to make up a "valid" edge. The keys to this grammar are the concepts of *persistence of magnitudes* and *coherence of phase* across scale, which we will describe in the complex wavelet domain.

The quintessential wavelet-domain algorithm, used widely because of its power and simplicity, is denoising by soft-thresholding. In fact, thresholding has certain optimality properties [2] if the noise-free function lies in certain smoothness spaces. In practice, though, wavelet domain thresholding produces unpleasant "ringing" artifacts around the edges in an image. The reason for this ringing is that an edge is represented by wavelet coefficients at multiple scales following the aforementioned edge grammar; if these coefficients are treated independently, then the grammar could be violated, resulting in "unclean" edges. To get results with "clean edges", we need algorithms that exploit the statistical dependence between the wavelet coefficients caused by the edge structure in the image.

In the past, edge structure has been exploited very loosely in wavelet domain models such as the Hidden Markov Tree [3] (HMT). The (simple) concept that these methods are based on is *persistence*: if a parent wavelet coefficient is large/small, then its child tends to be large/small.[1] The rationale behind small value persistence is that wavelet coefficients are small in smooth regions, and that a smooth region is subdivided into more smooth regions at finer scales. The rationale for large value persistence is that large wavelet coefficients are caused by edges, and if a region contains an edge, some of the subdivisions of that region will also contain the same edge.

The HMT model captures this property by assigning a hidden state to each wavelet coefficient signifying whether the coefficient is small (the corresponding basis function has its support in a smooth region) or large (an edge lies in the support of the basis function) [3, 4]. Dependencies between wavelet coefficients across scale are then introduced by making the state of the children coefficients depend on the state of the parent. Given the states, the parent and child are otherwise independent. HMT image processing algorithms (see [4] for denoising and [5] for segmentation) show a marked improvement over standard techniques, but the results still suffer from ringing and smoothed out edges.

The HMT models edge structure with chains of large coefficients across scale. Although true in spirit (under the commonly used heuristic of wavelet basis functions as local edge detectors), Figures 1(a) and 2(b) show that it is quite possible for the magnitude of a wavelet coefficient representing an edge to be arbitrarily small. We can address this problem by using complex wavelets (see [6] and Section 2). Since the complex wavelet transform is approximately shift-invariant, the persistence of large values across scale becomes a more valid assumption (see Figure 1(b) and 2(c)), and a more effective model results [7].

Not only is the magnitude of the complex wavelet coefficients more amenable to HMT modeling, but the *phase* also shows important relationships. Figure 2(d) shows that the phase of the complex wavelet coefficients along the edge in an image have significant structure. In Section 3, we discuss

Research supported by NSF grant CCR–9973188, ONR grant N00014–99–1–0813, DARPA, and Texas Instruments.
Email: {jrom, choi, richb}@ece.rice.edu; Internet: dsp.rice.edu

[1]The wavelet coefficients can be naturally ordered on a tree. The *children* coefficients analyze the signal at one scale finer than the *parent*.

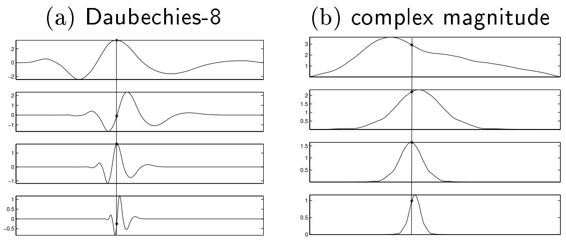

Figure 1: *Step response of Daubechies-8 and complex wavelets at several scales. The vertical line passes through wavelet coefficient values for one particular location of an edge. For D8, the chain of wavelet coefficient magnitudes representing an edge at this location are small at some scales and large at others. In contrast, the complex magnitudes show a steady progression of large values from one scale to the next.*

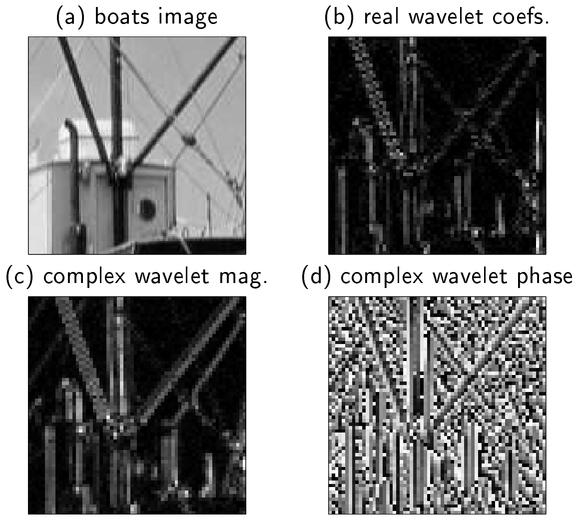

Figure 2: *(a) section of the "boats" image, (b) the real wavelet coefficients at one scale, (c) the magnitude of the complex wavelet coefficients at one scale, and (d) the phase of the complex wavelet coefficients. Notice the behavior of the wavelet coefficients along an edge in the image. The real wavelet coefficients oscillate between large and small magnitude values, while the complex magnitudes remain relatively constant. The phase of the complex coefficients is also structured along the edge.*

how the phase of a complex wavelet coefficient representing and edge gives us direct information about the location of the edge. Since the location of the edge does not change from scale to scale, the phases of the wavelet coefficients representing the edge exhibit *coherence* across scale; as an immediate consequence, the phases of wavelet coefficients at fine scales can be predicted from the phases at course scales. When there are many edges inside the support of a wavelet basis function, such as in a texture region in an image, their phase effects *interfere* with one another, and the resulting phases are incoherent across scale.

In Section 4, we extend the HMT model to include these phase relationships. Algorithms based on this model can discriminate between the large magnitude wavelet coefficients representing edges and the ones representing texture regions and treat them accordingly.

In this paper, the model in constructed for and demonstrated on 1D image slices. The 2D extension of the model is a topic of current research, and as such is only summarily discussed in the conclusion.

2. COMPLEX WAVELET TRANSFORM

The complex wavelet transform (CWT), like the usual real wavelet transform, expands a signal or image in terms of a set of complex wavelet basis functions $\{\psi_{j,k}(t)\}$, where the $\psi_{j,k}(t)$ are shifted and dilated versions of a mother wavelet $\psi(t)$; $\psi_{j,k}(t) = 2^{-j/2}\psi(2^j t - k)$. The expansion coefficients $c_{j,k}$ are calculated by taking the inner product of the signal with the basis function $\psi_{j,k}$. The $c_{j,k}$ can be arranged naturally in a tree structure (quad-tree for 2D), with each "parent" coefficient at scale j giving rise to two (four) "children" coefficient at scale $j+1$. We will denote the parent of $c_{j,k}$ as $c_\rho(j,k)$.

The basis functions have complementary real and imaginary parts: $\psi = \psi^{\mathsf{r}} + i\psi^{\mathsf{i}}$, with both ψ^{r} and ψ^{i} meeting the conditions for real-valued mother wavelets.[2] The expansion coefficients $c_{j,k}$ for a discrete signal can be computed with the same $O(N)$ efficiency as the real wavelet transform using the dual-tree technique in [6].

If we make ψ^{r} and ψ^{i} a Hilbert transform pair (algorithms for designing such wavelets can be found in [9]), the $c_{j,k}$ become approximately shift-invariant [10]. Since the envelope $|\psi_{j,k}(t)|$ is a slowly varying function of time, the magnitude $|c_{j,k}|$ of each wavelet coefficient is insensitive to small signal shifts. It therefore forms a more accurate estimate of signal activity at a given location and scale than the corresponding coefficient of a real wavelet transform, which will suffer from shift variance. Specifically, an edge in a signal causes a chain of large magnitude wavelet coefficients across scale (see Figure 1(b)), the first component in our multiscale grammar.

The complex wavelet basis set used in this paper (filter coefficients are given in [6]) can be thought of as a discrete approximation to the Cauchy (or Klauder) wavelet [11]

$$\psi(t;\alpha) = \frac{\Gamma(\alpha+1)}{2\pi}(1-it)^{-(1+\alpha)}. \tag{1}$$

These functions are optimally concentrated in time and scale (analogous to the Gaussian being optimally concentrated in time and frequency). The magnitude and phase of the Cauchy basis functions are given by

$$|\psi(t;\alpha)| = (1+t^2)^{-(1+\alpha)/2} \tag{2}$$

[2]The CWT can be interpreted as a wavelet tight frame with a redundancy factor of two [8].

and

$$\arg \psi(t;\alpha) = (1+\alpha)\arctan t. \tag{3}$$

Cauchy wavelets have two key properties. The first is that for moderately large α (say $\alpha \approx 10$), the phase is approximately linear for the region around $t=0$ with large magnitude. The second is that the integral of a Cauchy wavelet basis function is another Cauchy wavelet basis function

$$\int_{-\infty}^{t} \psi(\tau;\alpha)d\tau = -i\psi(t;\alpha-1). \tag{4}$$

In the next section, we will use these properties to describe the phase behavior of edges across scale, which will comprise the second part of our multiscale edge grammar.

3. PHASE PROPERTIES OF EDGES

The value of a wavelet coefficient representing a perfect step edge, calculated (in 1D) using

$$c_{j,k} = h\int_{-\infty}^{\ell} \psi_{j,k}(t)dt, \tag{5}$$

depends explicitly on both the height h and the location ℓ of the jump discontinuity. As such, a given value of a wavelet coefficient analyzing an edge could have been caused by any number of height/location combinations. However, the *phase* of the wavelet coefficient is independent of the jump height and therefore gives us only information about the location of the singularity.

To calculate the phase of the step response for the Cauchy wavelet $\psi(t;\alpha)$ given in (1) above, we combine (3) and (4):

$$\phi_{j,k}(\ell) := \arg c_{j,k} = \alpha \arctan(2^j\ell - k) - \frac{\pi}{2}. \tag{6}$$

We can, of course, solve (6) for the edge location ℓ. The solution is non-unique, since the phase of the wavelet coefficient is only known modulo 2π, but we can fully recover the edge location by looking at the phase of the wavelet coefficients at several scales.

As we move from scale to scale, the location of the edge remains constant. Given that we are analyzing a step edge, we can use (6) to predict the phase of children wavelet coefficients from the phase of the parent coefficient (in [12], another method for predicting phase from scale to scale is presented). For example, let c_{j,k_ρ} be the parent coefficient of c_{j+1,k_c}. We have

$$\phi_{j,k_\rho}(\ell) \approx \lambda(\ell - k_\rho) - \frac{\pi}{2} \tag{7}$$

$$\phi_{j+1,k_c}(\ell) \approx 2\lambda(\ell - k_c) - \frac{\pi}{2} \tag{8}$$

where $\lambda \sim 2^j\alpha$ is the slope of the linear approximation to the phase at scale j. Combining the above equations, results in an expression for the child phase in terms of the parent phase:

$$\phi_{j+1,k_c} = 2\phi_{j,k_\rho} + \frac{\pi}{2} \pm C \tag{9}$$

where $C = 2\lambda(k_\rho - k_c)$ is a constant, since $(k_\rho - k_c) \sim 2^{-j}$ if c_{j+1,k_c} is the left child of c_{j,k_ρ} and $(k_\rho - k_c) \sim -2^{-j}$ if c_{j+1,k_c} is the right child of c_{j,k_ρ}.

We now have the second element of our multiscale edge grammar: not only do edges causes chains of large magnitude coefficients across scale (persistence in magnitudes), the phase relationships between these coefficients is governed by (9) (coherence of phase).

When the signal contains many discontinuities inside the support of a wavelet basis function, in a textured region of an image, for example, the corresponding wavelet coefficients still have large magnitudes. However, their phases will interfere with each other, and thus will not be coherent across scale.

We can draw a loose comparison between complex wavelet analysis and radar. In radar, when the object of interest is large compared to the analyzing wavelength, the signal has coherent phase. If the object of interest is small compared to the analyzing wavelength, the signal has incoherent phase and results in "speckle". In complex wavelet analysis, the object of interest is an edge, and a chain of wavelet coefficients have coherent phase if the edge is *isolated*, meaning it is the only feature at each scale.

This phase coherence allows us, using the machinery presented in the next section, to discriminate between chains of large magnitude coefficients across scale caused by "texture" regions, and chains of large coefficients caused by edges.

4. INCORPORATING PHASE INTO THE HMT

The hidden Markov tree wavelet model, introduced in [3], operates on two principles: 1) that most of the wavelet coefficients of real world signals are very small, with a few being large; and 2) small/large values tend to persist across scale.

To match the first property, the HMT models the marginal probability distribution (pdf) of the wavelet coefficients at each scale as a Gaussian mixture. To each wavelet coefficient c_n,[3] we associate a discrete hidden state s_n that takes values $q = \mathsf{S}, \mathsf{L}$ with probability mass function $p(s_n)$. Conditioned on $s_n = q$, the real and imaginary parts of c_n are independently Gaussian with variance $\sigma^2_{n,q}$. The overall marginal density function is given by [7]

$$f(c_n) = \sum_{q=\mathsf{S},\mathsf{L}} p(s_n = q) f(c_n | s_n = q) \tag{10}$$

with

$$f(c_n|s_n = q) = \frac{1}{2\pi\sigma^2_{n,q}} \exp\left(-\frac{(u_n^2 + v_n^2)}{2\sigma^2_{n,q}}\right) \tag{11}$$

where u_n and v_n are the real and imaginary parts of c_n.

The HMT matches the second property, the persistence of large/small values across scale, by setting up a Markov-1 dependency structure on the hidden states across scale (a hidden Markov model), as shown in Figure 3(a). The state transition probabilities between the parent and child hidden states model the persistence property.

The HMT parameters, which we will collect into one parameter vector Θ, consist of the Gaussian mixture variances $\sigma^2_{n,q}$, the transition probabilities $p(s_n|s_{\rho(n)})$, and the probability mass function $p(s_0)$ for the root state s_0. In practice, all the wavelet coefficient parameters at a scale j share the same HMT parameters; this is know as *tying*. As with all hidden Markov models, there are three problems associated with HMTs [13]; 1) given Θ, find the likelihood of a given set of c_n; 2) given Θ and a set of c_n, calculate

[3]For the sake of clarity, we will use one number n to index the wavelet coefficients in this section instead of the usual double index $\{j,k\}$.

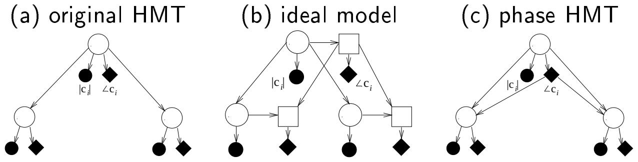

Figure 3: *(a) shows the graphical model for the standard HMT. The states (represented by the open circles) controls the magnitude of the wavelet coefficients (represented by the filled circles). In (b), the edge location, which governs the phase, is modeled as a separate hidden state represented by a square and is inherited down through scale. Part (c) shows the model used in the paper; the current state depends on the state of the previous wavelet coefficient and the actual phase of the previous wavelet coefficient.*

the most likely state sequence s_n; and 3) given a set of c_n, calculate the most likely set of parameters Θ to generate the c_n. These three problems all have elegant, $O(N)$ computational complexity solutions, presented in [3].

To modify the HMT model to account for the phase coherence of edges across scale, we will separate the "large" L state into a state for large magnitude coefficients with incoherent phase, we will label the state T for "texture", and a state for large magnitude wavelet coefficients with coherent phase across scale, labelled E for "edge". We now have a model with three states, $\{\mathsf{S}, \mathsf{T}, \mathsf{E}\}$. The conditional distributions for the S and T states remain the same as the usual HMT:

$$f(c_i|s_n = \mathsf{S}) \sim N(0, \sigma^2_{n,\mathsf{S}}) \quad (12)$$

$$f(c_i|s_n = \mathsf{T}) \sim N(0, \sigma^2_{n,\mathsf{T}}). \quad (13)$$

The pdfs (12)–(13) can also be thought of as Rayleigh distributions (with parameter $\sigma_{n,q}$) on the magnitude $|c_n|$ and an independent, uniform (between 0 and 2π) distribution on the phase $\arg c_n$.

For the E state, the magnitude of the wavelet coefficients is distributed the same as in the T state,

$$f(\ |c_n|\ |s_n = \mathsf{E}) \sim \mathrm{Rayleigh}(\sigma_{n,\mathsf{T}}). \quad (14)$$

Given the edge location, the model the phase as being tightly distributed around $\phi_n(\ell)$:

$$f(\arg c_n|s_n = \mathsf{E}) \sim \beta_{2\pi}(\phi_n(\ell), p, p), \quad (15)$$

a symmetric Beta density on the circle centered at $\phi_n(\ell)$.

Obviously the associated edge location ℓ for a wavelet coefficient in state E is unknown a priori. One approach would be to adopt ℓ as another, continuous valued hidden state variable. The value is inherited from scale to scale, and is used to generate the phase if the wavelet coefficient is in state E (see Figure 3(b) for the graphical description). Developing algorithms for such a model is difficult, however, since they require finding the best state sequence in a continuous space.

Instead, we will model the phase as being distributed around a value $\mathcal{P}(\arg c_{\rho(n)})$ predicted, using (9), from the phase of the parent coefficient $c_{\rho(n)}$:

$$f(\arg c_n|s_n = \mathsf{E}) \sim \beta_{2\pi}(\mathcal{P}(\arg c_{\rho(n)}), p, p). \quad (16)$$

If we consider the predicted phase value as part of the state, we have the graphical model shown in Figure 3(c). We no longer have a true hidden Markov structure, since the state of a wavelet coefficient depend on both the state of the parent and the actual value of the parent, but we can still derive efficient algorithms to solve the three problems of hidden Markov models listed above.

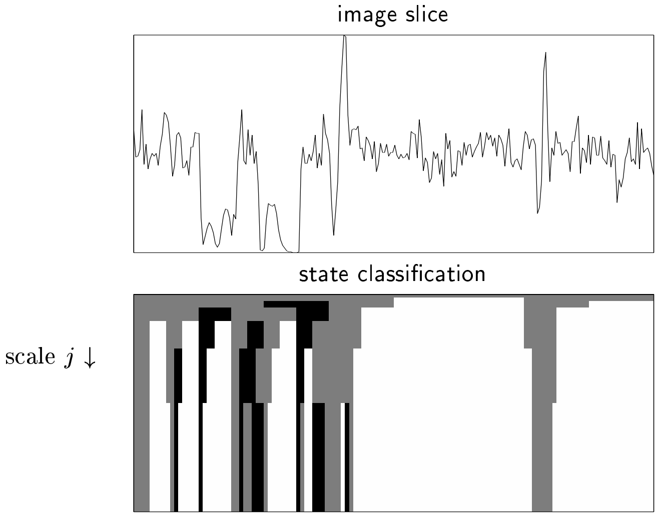

Figure 4: *State classification results using the model presented in Section 4. The figure on bottom is a multiscale display of the states of the wavelet coefficients. White areas correspond to coefficients in the S state, gray areas to T, and black areas to E. Note that all of the major edges in the slice are resolved at the finest scale with the exception of the pair on the right that are separated by a distance smaller than the width of the finest scale wavelet basis function.*

To show the effectiveness of this model in discriminating between large magnitude wavelet coefficients cause by texture and ones cause by edges, we can use the Viterbi algorithm (the solution to HMT problem 2 above) to segment the wavelet coefficients into the most likely set of states. Figure 4 shows the results on a 1D slice of the popular "cameraman" test image. At course scales, almost all of the wavelet coefficient are classified as texture, since many edges reside inside the support of each of the basis functions. As we move to finer and finer scales, we see edges slowly being resolved; at the finest scale, all edges except the two on the right located extremely close together (meaning, perhaps, that we could not reasonably expect to separate them into different entities) are represented by unique wavelet coefficients.

Although this toy segmentation algorithm is the only example we give in this paper, this new model provides a mechanism for specifying a *prior distribution* on the wavelet coefficients of real world images, and could conceivably be used (like the standard HMT has) to solve many other problems in statistical signal and image processing, e. g., denoising, classification, and compression.

5. CONCLUSIONS

Characterizing edge structure in the wavelet domain lies at the heart of image modeling. Traditionally, efforts to capture edge structure have revolved around modeling the dependencies between the magnitudes of wavelet coefficients from scale to scale. In this paper, we have seen that the phase also plays a key role in the behavior of edges in the wavelet domain. Using a few simple properties of complex wavelets, we developed a more comprehensive description of this behavior and showed how it can be incorporated into existing signal and image models.

Along with the across scale phase behavior, we believe that in 2D the behavior within scale will play a crucial role. As we can see in Figure 2(d), the phases of complex wavelet coefficients at the same scale exhibit a significant amount of structure along edges in image. This is a topic of current research for the authors.

REFERENCES

[1] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco, 1982.

[2] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. on Info. Theory*, vol. 41, no. 3, May 1995.

[3] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[4] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *To appear in IEEE Trans. on Image Proc.*, July 2001.

[5] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *To appear in IEEE Trans. Image Proc.*, 2001.

[6] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, vol. 357, pp. 2543–2560, September 1999.

[7] H. Choi, J. K. Romberg, R. G. Baraniuk, and N. G. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. of ICASSP 00*, Istanbul, Turkey, June 2000.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.

[9] I. W. Selesnick, "The design of Hilbert transform pairs of wavelet bases via the flat delay filter," *preprint*, 2000.

[10] F. C. Fernandes, *Directional, Shift-Insensitive, Complex-Wavelet Transforms with Controllable Redundancy*, Ph.D. thesis, Rice University, 2001.

[11] M. Holchneider, *Wavelets: An Analysis Tool*, Oxford Science, 1995.

[12] T. H. Reeves and N. G. Kingsbury, "Prediction of coefficients from coarse to fine scales in the complex wavelet transform," *preprint*.

[13] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.