MULTISCALE CLASSIFICATION USING COMPLEX WAVELETS

Justin Romberg, Hyeokho Choi, Richard Baraniuk

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

Nicholas Kingsbury

Dept. of Engineering
University of Cambridge, Trinity College
Cambridge, CB2 1TQ, UK

ABSTRACT

Multiresolution signal and image models such as the hidden Markov tree (HMT) aim to capture the statistical structure of smooth and singular (textured and edgy) regions. Unfortunately, models based on the orthogonal wavelet transform suffer from shift-variance, making them less accurate and realistic. In this paper, we extend the HMT modeling framework to the complex wavelet transform, which features near shift-invariance and improved angular resolution compared to the standard wavelet transform. The model is computationally efficient (featuring linear-time computation and processing algorithms) and applicable to general Bayesian inference problems as a prior density for the data. In this paper, we develop a simple multiscale maximum likelihood classification scheme based on the complex wavelet HMT that outperforms those based on traditional real-valued wavelet transforms. The resulting classification can be used as a front end in a more sophisticated multiscale segmentation algorithm.

1. INTRODUCTION

The success of wavelet-based algorithms for image processing is due mostly to the compact representation of images provided by the discrete wavelet transform (DWT). This compact representation leads to an easier characterization of image structure, and results in simple yet effective wavelet-domain image modeling techniques. Modeling lies at the core of statistical image processing; better models yield better solutions to problems such as compression, estimation, detection, and classification.

The wavelet transform manifests image structure in the following two key properties [1]:

2Populations: Smooth signal/image regions are represented by small wavelet coefficients, while edges, ridges, and other singularities are represented by large coefficients.

Persistence: Large and small wavelet coefficient values cascade along the branches of the wavelet tree.

Interpreting these properties statistically, 2Populations implies that wavelet coefficients have non-Gaussian marginal statistics, with a large peak at zero (many small coefficients) and heavy tails (a few large coefficients), while Persistence implies that wavelet coefficient values are statistically dependent along the branches of the wavelet tree.

The hidden Markov tree (HMT) captures these statistical properties in a model of the joint probability density function (pdf) $f(\mathbf{w})$ for the wavelet coefficients $\mathbf{w}$ of an image. To capture the 2Populations property, we model each wavelet coefficient w_i with a two-state Gaussian mixture; a controlling hidden state s_i records the magnitude of the coefficient as either S (small) or L (large). To capture the Persistence property, we connect the hidden states across scale (down the wavelet tree) in a Markov-1 chain in such a way that the unbroken progressions L-L-L-L and S-S-S-S are more likely than broken ones such as L-L-S-L and S-S-L-S. These latter two progressions lead to edge blurring and ripple artifacts, respectively.

When employed in a Bayesian framework as a prior density on the wavelet coefficients of the data, the HMT model plays the rôle of a statistical *grammar* that makes unbroken strings of L and S states preferable to broken strings. Numerous experiments with real data have verified that the HMT is an accurate and effective model for a large class of real-world signals and images [1, 2].

Unfortunately, due to the strong shift-variance of the orthogonal DWT [3], the 2Populations and (most notably) Persistence properties are strongly affected by the placement of the smooth and edgy regions in the data. In particular, the coefficient magnitudes near an edge can swing wildly as the alignment between the edge and the wavelet basis functions changes slightly (even by one sample). This weakens the assumption of persistence, resulting in a less accurate model. To attack this problem, we need to counter the shift-variance of the wavelet transform. Traditionally, this is done by using a shift-invaraint (redundant) wavelet transform, but HMT models in this domain [4] can be computationally burdensome.

Supported by NSF grants MIP–9457438 and CCR–9973188, ONR grant N00014–99–1–0813, DARPA/AFOSR, grant F49620–97–1–0513, and Texas Instruments.

Supported by the Rice-Trinity Exchange Program and the UK Engineering and Physical Sciences Research Council (EPSRC).

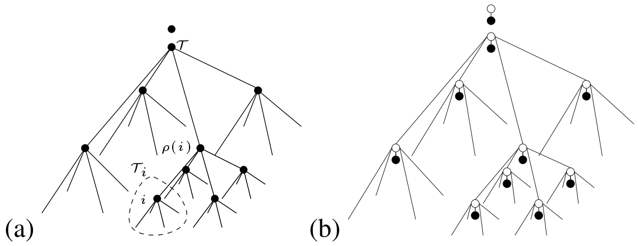


Fig. 1. *(a) Quad-tree of wavelet coefficients w_i (black nodes) for one subband of the 2-d DWT. (b) 2-d HMT model (white nodes represent hidden state variables s_i).*

In this paper we take an approach based on the dual tree *complex wavelet transform* (CWT) [5]. The CWT expands a signal or image in terms of a set of complex wavelets with complementary real and imaginary parts. The most important property for the present study is the near shift invariance of the CWT coefficient *magnitudes*, which leads to strong persistence of large magnitudes around edges, independent of their alignment. As an added bonus, in 2-d the CWT offers six angle-selective subbands rather than the standard DWT's three. Thus the transform is less sensitive to the angular orientation of the edges. Finally, the computational cost of the CWT remains $O(N)$. In addition, the HMT modeling framework is easily extended to the CWT (see Section 4).

To demonstrate the effectiveness of the complex wavelet HMT (CHMT), we will develop a multiscale texture classification algorithm that outperforms that based on the standard HMT model. Multiscale texture classification, descibed in Section 5, is a way of dividing up an image into regions containing similar textures. More complicated segmentation algorithms use multiscale classifiers as front ends, and the segmentation quality relies on the effectiveness of the multiscale classification.

After reviewing the HMT and CWT in Sections 2 and 3, we derive the new model in Section 4. In Section 5, we demonstrate a multiscale classification algorithm based on the new model. Section 6 concludes with a discussion of other potential applications of the new model, including denoising, segmentation, and synthesis.

2. WAVELET HMT MODEL

For concreteness, we will emphasize images in this paper, though the ideas apply to data of any dimensionality. The 2-d orthogonal DWT consists of three subbands of coefficients [6], each of which has a quad-tree structure like that shown in Fig. 1(a). Denote the *parent* of wavelet coefficient w_i as $w_{\rho(i)}$.

To statistically match the 2Populations property of the wavelet transform, we consider the collection of small/large wavelet coefficients as outcomes of a Gaussian pdf with small/large variance. This approximates the marginal pdf $f(w_i)$ of each wavelet coefficient as a *Gaussian mixture*. To each wavelet coefficient w_i, we associate a discrete hidden state s_i that takes on values $m = \mathsf{S}, \mathsf{L}$ with probability mass function (pmf) $p(s_i)$. Conditioned on $s_i = m$, w_i is Gaussian with mean $\mu_{i,m}$ and variance $\sigma^2_{i,m}$. Thus, its overall marginal pdf is given by[1]

$$f(w_i) = \sum_{m=\mathsf{S},\mathsf{L}} p(s_i = m)\, f(w_i | s_i = m), \qquad (1)$$

with $f(w_i|s_i = m) \sim N(\mu_{i,m}, \sigma_{i,m})$. While each wavelet coefficient w_i is conditionally Gaussian given its state variable s_i, overall it has a non-Gaussian density due to the randomness of s_i.

To match the Persistence property of large and small coefficients, we set up a Markov-1 dependency structure on the hidden states across scale (a hidden Markov model) [1]. Using links to denote dependencies, the HMT model takes the form of the tree in Fig. 1(b). The state transition probabilities between the connected states model the persistence of large/small coefficient magnitudes across scale. The overall HMT model consists of an HMT for each of the three DWT subbands.

The HMT model parameters consist of the Gaussian mixture means and variances $\mu_{i,m}, \sigma^2_{i,m}$, the transition probabilities $p(s_i|s_{\rho(i)})$, and the pmf $p(s_0)$ for the root state s_0. Group these into the parameter vector $\mathbf{\Theta}$. In practice, we usually share the same parameters within each scale band [1]. This process, called *tying*, makes the model more robust against overfitting.

The HMT can be trained to capture the wavelet-domain features of the image(s) of interest using the iterative expectation-maximization (EM) algorithm [1]. For a given set of training data, the trained model $f(\mathbf{w}|\mathbf{\Theta})$ approximates the joint pdf of the wavelet coefficients. See [1, 2] for more details.

3. COMPLEX WAVELET TRANSFORM

The dual tree CWT comprises two parallel wavelet filter bank trees that contain carefully designed filters of different delays that minimize the aliasing effects due to decimation [5]. Each tree produces a valid set of *real* DWT coefficients u_i and v_i; together they form the *complex coefficients* $c_i = u_i + jv_i$.[2] The CWT is near shift invariant and, in 2-d, provides good directional selectivity with six oriented subbands per scale [5].

[1] The number of states can be increased beyond two to match the marginal distribution with arbitrary precision. However, we have found that a model with two states provides a fine accuracy/complexity balance.

[2] The CWT can be interpreted as a wavelet tight frame with a redundancy factor of two [6].

In 2-d, the CWT basis functions closely approximate modulated complex exponentials (Gabor functions) of the form

$$h(x, y) = a(x, y)\, e^{j(\omega_x x + \omega_y y)} \quad (2)$$

with $a(x, y)$ a slowly varying Gaussian-like real window function centered on $(0, 0)$, and (ω_x, ω_y) the center frequency of the corresponding subband. The real and imaginary parts of the CWT basis functions are symmetric and antisymmetric respectively about $(0, 0)$ and are therefore orthogonal. This means that the real and imaginary parts of each CWT coefficient are approximately statistically uncorrelated.

Furthermore, since $a(x, y)$ is slowly varying, the *magnitude* $|c_i| = \sqrt{u_i^2 + v_i^2}$ of each CWT coefficient is insensitive to small image shifts. It therefore forms a more accurate estimate of image activity at a given location and scale than the corresponding coefficient of a real DWT, which suffers from shift variance.

The directional properties of the CWT arise from the fact that $h(x, y)$ has constant phase along lines such that $\omega_x x + \omega_y y$ is constant, giving filters selective to edges parallel to this direction. For the six subbands at each scale, $\omega_y/\omega_x \approx \{-3, -1, -\frac{1}{3}, \frac{1}{3}, 1, 3\}$, which produces bands oriented at approximately all odd multiples of $15°$.

4. COMPLEX WAVELET HMTS

Due to its near shift invariance, the progressions of large and small magnitudes in the CWT coefficients $\mathbf{c}$ are more reliable than in either the real part $\mathbf{u}$ or imaginary part $\mathbf{v}$ alone. In this section, we will leverage this reliability into a robust HMT model for the CWT.

In the simplest method, we associate with each CWT coefficient c_i a single hidden state q_i taking value S,L depending on whether $|c_i|$ is small or large. Consider each complex coefficient c_i as a 2-d random vector (u_i, v_i), and recall that since the real and imaginary filter banks are orthogonal, the real and imaginary parts are approximately uncorrelated random variables. Thus we approximate the marginal density $f(c_i)$ as a two-state, 2-d Gaussian mixture

$$f(c_i|q_i = m) = \frac{1}{2\pi\sigma_{i,m}^2} \exp\left(-\frac{u_i^2 + v_i^2}{2\sigma_{i,m}^2}\right). \quad (3)$$

Factoring (3) into

$$f(c_i|q_i = m) = \frac{1}{2\pi\sigma_{i,m}^2}\, e^{-u_i^2/2\sigma_{i,m}^2}\, e^{-v_i^2/2\sigma_{i,m}^2} \quad (4)$$

reveals that we model the real and imaginary parts as two-state, 1-d Gaussian mixtures with *shared state variable* q_i. This is quite reasonable: Small $|c_i|$ result only when both $|u_i|$ and $|v_i|$ are small simultaneously (smooth region). This is captured by $q_i = \mathsf{S}$. On the contrary, large $|c_i|$ result when either $|u_i|$ is large, $|v_i|$ is large, or both are large (edge region). This is captured by $q_i = \mathsf{L}$. (Note that, as in the DWT HMT case, small outcomes from the L-state Gaussian mixture density are entirely possible – and in fact likely – since this component has zero mean.)

This approach is equivalent to modeling each $|c_i|$ as a two-state Rayleigh mixture model. Theformulation (3)–(4) (coupled Gaussian mixtures for the real and imaginary parts) is preferable, however, as direct Rayleigh modeling leads to unmanageable posterior distribution expressions.

The CWT HMT (CHMT) corresponding to (3) with scale-to-scale Markov transitions has a structure identical to the real DWT HMT of Section 2 and [1]. The only differences will be the substitution of (3) for (1) and the use of six subband trees instead of three. The result is an HMT that supports $O(N)$ coefficient computation and training and inference algorithms and features better angle selectivity.

As an alternative to the above, we could model the $\mathbf{u}$ and $\mathbf{v}$ DWTs separately using standard 2-state, 1-d HMTs. The result is a 4-state, 2-d HMT for each $|c_i|$ that depends on the four combinations of large/small u_i and v_i. Tying the $\mathbf{u}$ and $\mathbf{v}$ HMTs node by node by forcing them to share the same mixture means and variances (but not hidden state values) effectively models $\mathbf{u}$ and $\mathbf{v}$ as independent realizations of the same random process. Thus, we can apply the training and inference algorithms for multiple data observations described in [1]. We will not pursue this approach further in this paper.

5. APPLICATION: MULTISCALE CLASSIFICATION

With a wavelet-domain HMT model in hand, many statistical image processing tasks, such as estimation, classification, and segmentation, are straightforward [1, 2]. In this section, we will use the CHMT to model the hypotheses (textures) in a multiscale maximum likelihood classification algorithm.

Consider a standard texture *classification problem*: given probability models $f(\mathbf{x}|H_i)$ for a set of textures $\{H_i\}$ and a data set $\mathbf{x}$ of size n, we decide which of the texture models best describes $\mathbf{x}$. This decision is based on the texture model that yields the highest likelihood, i.e. we choose the H_i that makes $f(\mathbf{x}|H_i)$ the greatest. As n gets larger, the decision becomes more reliable.

A more interesting problem is the *supervised segmentation problem*: given the $f(\mathbf{x}|H_i)$ and an image comprised of a montage of homogenous texture regions, we affix labels to the regions in the image that correspond to each texture. The solution to this problem is less straightforward since we don't have any information about the *shape* of the regions. If we are given shape information (and a model for texture in each shape), we can use the maximum likelihood technique described above to classify each region independently.

Multiscale classification acts as a bridge between these two problems. At a scale $j = 0, 1, \ldots$, we divide the $N \times N$ image into dyadic squares with side length $\frac{N}{2^j} \times \frac{N}{2^j}$, and then classify each dyadic square using the maximum likelihood technique. The result is a set of "raw segmentations" of the image, where we have assumed that the shapes of the texture regions are dyadic squares. The raw segmentations at course scale (small values of j) are based on many data points, and hence the decisions are more reliable than those at fine scales (large j). However, the raw segmentations at fine scales have better resolution; the smaller dyadic squares more accurately approximate the true shape of the texture regions. Given this set of raw segmentations, we can develop an algorithm [2, 7] that combines them into a final segmentation.

Wavelet based texture models, such as the CHMT, are ideally suited for the multiscale classification problem. A subtree of wavelet coefficients, with its root at a scale j, contains information about the image inside a dyadic square of side length $\sim 2^{-j}$. By examining all subtrees of the wavelet coefficients of an image, we can assign a label to each dyadic square depending on which texture the subtree of wavelet coefficients best describes.

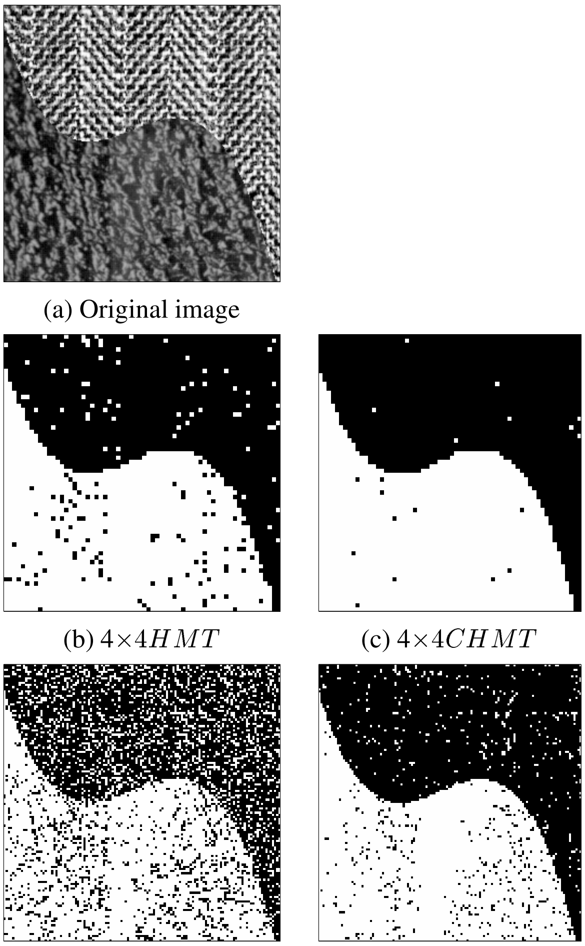
Wavelet HMT based methods approach the problem with two steps [2]:

1. Train HMT models for the wavelet coefficients of each texture. This requires samples (and hence a priori knowledge) of the textures found in the image.

2. Compute the likelihood of the image wavelet coefficients under each model at multiple scales, and assign the label corresponding to the model that yielded the highest likelihood.

HMT-based multiscale classification results using real and complex wavelets are shown in Fig. 2. The algorithm based on complex wavelets outperforms (gives fewer errors than) that based on real wavelets, which leads us to believe that complex wavelets capture more of the texture's key structure.

We can use the multiscale classification as a front end to a supervised segmentation algorithm [2]. The segmentation algorithms use a priori information about the shape of the texture regions to combine the raw segmentations into a final segmentation map. For example, [2] presents an algorithm based on a context model for shape, while [7] uses one based on description length. The multiscale classifier based on the complex wavelet tree can be inserted into either of these algorithms without modification. Since the multiscale classification is more accurate using the CHMT, we expect that there will be improvement in the end segmentation results.



(a) Original image

(b) $4{\times}4HMT$

(c) $4{\times}4CHMT$

(d) $2{\times}2HMT$

(e) $2{\times}2CHMT$

6. CONCLUSIONS

The ability of multiresolution signal and image models to capture the structures of signals and images is often hampered by the shift variance of the underlying wavelet transform. In this paper, we have seen that using complex wavelets, we can design statistical models that focus on the salient signal features (persistence across scale of edges and so forth) rather than artifacts due to shift variance. As our simple classification example has demonstrated, the CHMT outperforms methods based on traditional wavelet transforms, due to the parsimony and better angular resolution of its underlying transform.

While we have chosen to illustrate the application of the CHMT with a classification problem, this new model is generally applicable to Bayesian inference problems, including denoising, synthesis, compression and fusion. The near shift invariance and improved angular resolution of the CWT should benefit these applications as well.

7. REFERENCES

[1] 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.

[2] H. Choi and R. Baraniuk, "Multiscale texture segmentation using wavelet-domain hidden Markov models," in *Proc. 32st Asilomar Conference*, Pacific Grove, CA, Nov. 1-4 1998.

[3] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells Jr, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Sig. Proc. Letters*, vol. 3, no. 1, pp. 10–12, Jan. 1996.

[4] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *Submitted to IEEE Trans. on Image Proc.*, 2000, Available at www.dsp.rice.edu/publications.

[5] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, vol. 357, pp. 2543–2560, September 1999, Special issue for the discussion meeting on "Wavelets: the key to intermittent information?".

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

[7] Our, ," SPIE 00 paper on shape.