Adaptive Probabilistic Models of Wavelet Packets for the Analysis and Segmentation of Textured Remote Sensing Images*

Karen Brady, Ian Jermyn, Josiane Zerubia
Ariana, joint research group CNRS/INRIA/UNSA
INRIA, 2004 Route des Lucioles, B.P.93
06902 Sophia Antipolis Cedex, France
Firstname.Lastname@sophia.inria.fr

Abstract

Remote sensing imagery plays an important role in many fields. It has become an invaluable tool for diverse applications ranging from cartography to ecosystem management. In many of the images processed in these types of applications, semantic entities in the scene are correlated with textures in the image. In this paper, we propose a new method of analysing such textures based on adaptive probabilistic models of wavelet packets. Our approach adapts to the principal periodicities present in the textures, and can capture long-range correlations while preserving the independence of the wavelet packet coefficients. This technique has been applied to several remote sensing images, the results of which are presented.

1 Introduction

The automatic detection of regions of interest in remote sensing images is an important problem, with a wide variety of possible applications, including cartography, land use monitoring in agricultural areas, and assessment of environmental damage caused by natural disasters. Texture is frequently a salient cue in the analysis of such images, often being strongly correlated with semantic entities in the scene. Distinguishing a grass covered field from one which has been planted with trees, for example, is a task to which texture analysis is naturally suited.

Over the years numerous approaches have been proposed to tackle the problem of texture description. For a full overview, see [12] and [11]. One approach that has been extremely active in recent years is the use of wavelets [10, 9]. An example is the Hidden Markov Tree technique developed by [4], which describes the interscale dependencies of a standard wavelet decomposition. Others have investigated the use of wavelet packets as a texture analysis tool. In [8], packets were used in a classification experiment on natural

*This work was partially supported by European Union project MOUMIR, HP-99-108 (www.moumir.org). The authors would like to thank the French Mapping Agency (IGN) for the donation of some of the images used in this paper, and Space Imaging (www.spaceimaging.com) for permission to use others.

BMVC 2003 doi:10.5244/C.17.59

textures. Since [5], attempts have been made to adapt the wavelet (or wavelet packet) decomposition to the underlying structure of the texture, for example [3] and [1], but these methods have not been developed within a coherent probabilistic framework.

In [2], we introduced new, probabilistic models of texture, which were successfully applied to the segmentation of Brodatz mosaics. In this paper, we describe further developments of these texture models and a new segmentation method, as well as presenting results on remote sensing images. As stated, we address the issue of texture description within a probabilistic framework. Starting from a probability measure on the infinite texture, which we assume here to be Gaussian, we derive the measure for the texture on a finite region. This leads naturally to a class of adaptive wavelet packet models that capture the structure of a given texture, for example its principle periodicities, in a manner analogous to the Wold decomposition [7]. A simple classification rule enables pixelwise classification of the image while retaining the advantages of more complicated prior models. We consider supervised segmentation because it is methodologically well-defined.

The paper is organized as follows. We introduce our texture model in Section 2 and provide an outline of its development. In section 3, we describe how model parameters are estimated, and discuss the adaptive nature of our technique. Section 4 describes the segmentation procedure used. In section 5, we present results on several remote sensing images. Finally, in section 6, we conclude and discuss future work.

2 Modelling Planar Parallel Texture

One of the characteristics of planar texture, perhaps its defining characteristic, is that it is infinitely extendable, so that images of the texture are functions on an infinite (or at least very large) domain D_∞. A model of such a texture is given by a probability measure on the space of such images. We denote such a measure by:

$$\Pr(I|k \equiv l, K_l) \tag{1}$$

where I is the infinite image; K_l is the set of parameters of the model of texture $l \in L$, the set of textures; and $k : D_\infty \to L$ is the class map, which here takes every pixel to class l.

In practice however, when we analyse and segment textures, we work with images which are defined on finite regions. This means that one needs, not the distribution on the space of infinite images, but that on the space of images defined on a finite region. We thus need to marginalise equation (1) over the values of the pixels outside this region.

In this paper, as mentioned in section 1, we choose to model textured images using a Gaussian measure. In abstract notation, this can be expressed as:

$$\Pr(I|\cdot) = |F/\pi|^{1/2} e^{-\langle I-\nu|F|I-\nu\rangle} \tag{2}$$

where $\langle I|J\rangle$ is the inner product of the functions $|I\rangle$ and $|J\rangle$ in the space of images; $F|I\rangle$ means F acting on $|I\rangle$; ν is the mean; and $|F|$ is the determinant of the operator F.

Given a region $R \subset D_\infty$, we can split the operator F up into four parts: F_{RR}, $F_{R\bar{R}}$, $F_{\bar{R}R}$, and $F_{\bar{R}\bar{R}}$; these correlate R with itself, R with $\bar{R}$, its complement, $\bar{R}$ with R, and $\bar{R}$ with $\bar{R}$ respectively. Similarly, the mean can be split into two parts, ν_R and $\nu_{\bar{R}}$. Marginalising over the values of the pixels in $\bar{R}$ gives the following probability measure for the image I_R on the finite region R:

$$\Pr(I_R|\cdot) = |G_R/\pi|^{1/2} e^{-\langle I_R-\nu_R|G_R|I_R-\nu_R\rangle} \tag{3}$$

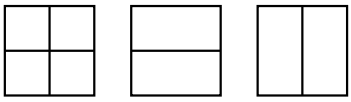

Figure 1: Fourier domain splits corresponding to quad and binary vertices.

where $G_R \equiv F_{RR} - F_{R\bar{R}} \left(F_{\bar{R}\bar{R}}\right)^{-1} F_{\bar{R}R}$. The computational complexity of evaluating this operator requires that we diagonalise G_R.

2.1 Diagonalisation of the operator G_R

If we can find a set $B = \{|a\rangle : a \in A\}$, where A is an index set, of functions on the region R such that:

1. The set $\{i_R|a\rangle : a \in A\}$ of infinite images that are equal to $|a\rangle$ in R and to zero in $\bar{R}$ are eigenfunctions of the operator F (with eigenvalues f_a);

2. The set B forms an orthonormal basis for functions on R;

then we can diagonalise the operator G_R. The first condition implies that the support of $Fi_R|a\rangle$ lies in the region R. Thus the second term in G_R is zero, and the first term becomes $f_a|a\rangle$. Hence $\langle a|G_R|a\rangle = f_a\langle a|a\rangle$. The second condition then means that G_R is diagonalised by B, allowing us to write our measure as:

$$\Pr(I_R|\cdot) = \prod_{a\in A} (f_a/\pi)^{1/2} e^{-\sum_{a\in A} f_a \langle I_R - \nu_R|a\rangle\langle a|I_R-\nu_R\rangle} \tag{4}$$

How do we find such a set B?

2.2 Using Wavelet Packets

As the texture could appear in the image arbitrarily translated, our measure must be translation invariant. This condition has two consequences. The first is that the mean ν must be constant, so that its Fourier transform is a delta function: $\nu(k) \propto \delta(k)$. The second is that F must be diagonal in the Fourier basis, $F(k,k') = f(k)\delta(k,k')$, which means that our measure is now characterized by a function f on the Fourier domain:

$$\Pr(I|\cdot) = |F/\pi|^{1/2} e^{-\sum_k f(k)(I^*(k)-\nu^*(k))(I(k)-\nu(k))} \tag{5}$$

For an arbitrary function $f(k)$, it is very hard to find a set B that satisfies the conditions in section 2.1. We thus want to choose a set of functions f that is varied enough to capture the structure present in the texture, but limited enough that we can satisfy the conditions. To this end, consider the set $\mathscr{T}$ of rooted trees in which each vertex has either four children, or two children and a label in the set $\{H,V\}$. Each such tree corresponds to a dyadic partition of the Fourier domain, where the leaves correspond to the elements of the partition. A vertex with four children corresponds to a standard quad tree split, while a vertex with two children corresponds to a two-way split in either the x (H) or the y (V) direction, but not both. The three possible splits are shown in figure 1.

Based on the set $\mathscr{T}$, we define a set of functions $\mathscr{F} = \bigcup_{T\in\mathscr{T}} \mathscr{F}_T$, where $\mathscr{F}_T = \{f : f \text{ is piecewise constant on } T\}$. Given a tree $T \in \mathscr{T}$, and a mother wavelet, we can define a wavelet packet basis B_T, each leaf of the tree corresponding to one subband. Each element of this basis has frequency support that lies approximately in the element of the partition T corresponding to its subband. The basis elements are thus approximate eigenfunctions of the operators defined by the functions in $\mathscr{F}_T$. Those basis elements whose support lies in R thereby satisfy condition 1 to a certain approximation.

How we address condition 2 is ultimately determined by the shape of the region R. If it is dyadic in shape, then our task of completing the set of wavelets inside the region, and in doing so forming a basis for it, is relatively straightforward. This is the chosen approach for training our texture models and is discussed further in section 3. For the case of arbitrarily shaped regions however, condition 2 is considerably harder to satisfy. It more or less forces the use of iterative optimization algorithms to perform the segmentation because, at each stage, we need a hypothesis for the class map k in order to know the regions and compute a basis. In this paper, we use a much simpler (but heuristic) pointwise classification method that does not require us to consider condition 2 for arbitrarily shaped regions.

3 Parameter Estimation

Since we are performing supervised segmentation, we will learn the parameters of the texture models corresponding to various semantic entities in the scene (*e.g.* 'forest', 'water', 'crops') in advance from training data. We thus collect a number of sample image patches for each entity considered. We have the freedom to choose the shape and size of these patches and so we select dyadic patches that are large enough to capture any periodicities in the texture. The advantage of dyadic patches is that condition 2 is approximately satisfied, at least for this training stage.

The unknown parameters of a texture model are: the wavelet packet basis to be used, or equivalently the dyadic partition, T; and the values of the means ν_α and inverse variances, f_α, $\alpha \in T$ labelling the subbands. Note however that the constancy of the mean ν, and in consequence the region mean ν_R, means that the wavelet packet coefficients of ν_R are all zero except for the scaling coefficient subband, which we assign $\alpha = 0$. We will thus use ν_α to denote the mean of subband α, with the understanding that it is zero unless $\alpha = 0$. From now on, f and ν will denote the maps from subbands to parameter values $f(\alpha) = f_\alpha$ and $\nu(\alpha) = \nu_\alpha$.

Following the argument in section 2.1, the measure for each dyadic training patch then takes the form:

$$\Pr(I_R|\nu, f, T) = \prod_\alpha \left[\left(\frac{f_\alpha}{\pi} \right)^{N_\alpha/2} e^{-f_\alpha \sum_{i\in\alpha} \left(w_{\alpha,i} - \nu_\alpha\right)^2} \right] \tag{6}$$

where i is an index for the individual wavelets within each subband α (strictly speaking, this must vary with α, a change we choose not to notate); $w_{\alpha,i}$ is the $\langle\alpha, i\rangle$ wavelet coefficient of the dyadic image; and N_α is the number of coefficients in subband α.

3.1 Adapting The Wavelet Decomposition

To find the optimal parameters for a given texture we examine the probability

$$\Pr(f,T|d) \propto \Pr(d|f,T)\Pr(f|T)\Pr(T) \tag{7}$$

where d is the training data samples used for a given texture. We assume $\Pr(f|T)$ to be uniform, and choose the prior, $\Pr(T)$, to penalize large decompositions:

$$\Pr(T) = Z^{-1}(\beta)e^{-\beta|T|} \tag{8}$$

where $|T|$ is the number of elements in the partition. The probability $\Pr(d|f,T)$ is given by equation (6). Differentiating with respect to f_α and ν_α gives us the maximum *a posteriori* (MAP) estimates of these quantities for fixed T:

$$\hat{f}_\alpha = \frac{N_\alpha}{2\sum_{i\in\alpha}\left(w_{\alpha,i}-\nu_\alpha\right)^2} \qquad \hat{\nu}_\alpha = \tfrac{2}{N_\alpha}\sum_{i\in\alpha} w_{\alpha,i} \tag{9}$$

Note how the independence of the subbands is critical here: without independence the MAP estimates would be harder to compute. The independence of the subbands further implies that we can use a depth-first search through the space $\mathscr{T}$ to find the exact MAP estimate for T.

Figure 2 contains some optimal decompositions. The images show the logarithm of f, in order to make clear the decomposition. Most of the f values are zero when scaled and quantized to eight bits. Also shown for comparison are the magnitudes of the Fourier coefficients.

3.2 Interpretation

The above models could equally well be expressed in terms of a standard wavelet basis, and considering what this means throws light on what the models achieve. Although the inverse covariance operator of a model is (approximately) diagonal in the adapted wavelet packet basis chosen by the above procedure, it will not in general be diagonal in the standard wavelet basis. Mathematically, this is clear. Intuitively, the reason is the following. Many textures possess periodic correlations between pixels at intermediate frequencies that extend over many wavelengths ([6]). The standard wavelet basis, however, links the notions of scale and frequency, so that the range of correlations is of the order of the inverse of the frequency of the wavelet (i.e. one wavelength). Correlations between pixels over a range corresponding to many wavelengths can thus only be captured in a standard wavelet basis by introducing interactions between coefficients. This is typically achieved through the use of tree models, in which coefficients interact via their parents, grand-parents, and so on.

In contrast, the adaptive bases presented here capture these long-range correlations using wavelet packet bases that can represent correlations between pixels over many wavelengths while remaining independent, with no need for interactions between the coefficients themselves. For Gaussian models, this type of description is always possible, since the inverse covariance can always be diagonalised, at least in principle. The problem is that the basis elements may not have compact spatial support, which renders them unsuitable for segmentation. The above procedure can thus be thought of as finding the

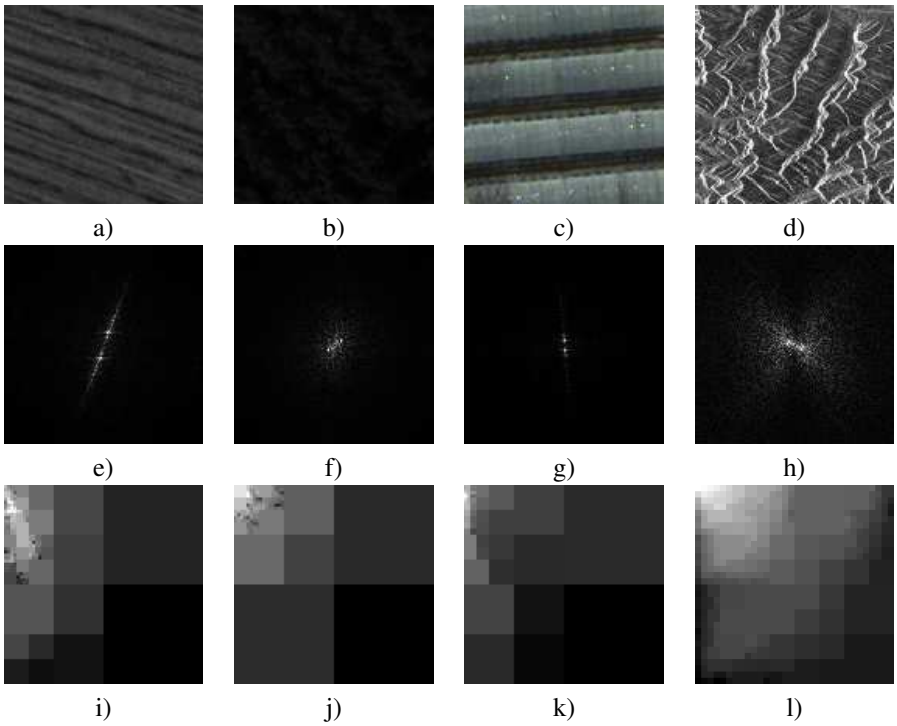

Figure 2: Row 1: Original textures: a) – d). Row 2: The corresponding Fourier magnitude images: e) – h). Row 3: The log display of their optimal decompositions: i) – l).

basis that diagonalises an inverse covariance that must simultaneously be learnt, subject to probabilistic restrictions on its form that ensure the more or less compact spatial support of the basis elements.

4 Classification

Having estimated our texture parameters, we move on to the problem of classifying images of scenes that contain multiple entities. Unlike the training stage of our procedure where we were at liberty to choose convenient dyadic training patches, the image regions corresponding to a given entity are arbitrarily shaped. Our dyadic wavelet packets of section 3 no longer form a basis for the region to be analysed. The full solution of this problem implies we must deal with condition 2, which we do not yet know how to satisfy. Instead, in this paper, we use a simple and heuristic classification method. This method has one major advantage: it is faster than the iterative optimization algorithms that would be necessary if condition 2 were to be satisfied.

First note that each texture model has a largest effective filter size, corresponding to the smallest elements in the Fourier domain partition. This filter size is dyadic. Given a set of texture models then, there is a largest effective filter size among them all. To classify pixel x in the image domain, we examine an image patch centred on x of size

equal to this largest filter size among all models. We can calculate the probability of the image restricted to this dyadic patch for each trained texture model using equation (6). Assuming that the prior on the class map k is uniform on L, these data probabilities can be normalized to give the posterior probabilities for the class of the patch. The MAP estimate of the class of the patch is then given, tautologically, by the class with maximum posterior probability. Rather than assign this class to every pixel in the patch however, we heuristically assign this class only to the pixel x at the centre of the patch. We then repeat this procedure for every pixel in the image.

The danger of such a pointwise procedure is that the class map k can be very irregular. One way to attack this problem is to use a non-trivial prior on k in conjunction with an optimization algorithm. Having avoided these algorithms in conjunction with condition 2 however, we choose not to introduce them now. Instead we use another heuristic procedure that has a similar effect to a regularizing prior: it smooths the class map and shortens boundaries. This is done by combining the posterior probability distributions on L at each pixel, π_x, in the following way:

$$\hat{k}(x) = \arg\max_{l \in L} \prod_{x' \in V(x)} \pi_{x'}(l) \tag{10}$$

where $V(x)$ is a set of neighbours of pixel x, which includes x itself. Examples of neighbourhoods can be seen in figure 3. This rule has a similar effect to a regularizing prior

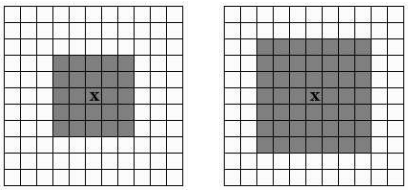

Figure 3: a) Neighbourhood-5 and b) Neighbourhood-7 schemes for pixel x

(*e.g.* the Potts prior), but it still allows a pixelwise classification because it uses the data at the neighbours of a pixel but not their unknown classes. In consequence, one can use larger neighbourhoods with little extra penalty.

5 Results

The method described above has been tested on a number of real aerial and satellite optical and SAR images. Examples of the results obtained are shown in Figure 4. Kappa values for these segmentation results are given in table 1. These values are based on the manually created ground truth images presented in the second row of figure 4. The classes used to generate each ground truth image are shown in table 2.

A more insightful look into the nature of the segmentations can be gained from examining the entropy $S(\pi_x)$ of the distributions π_x. In figure 5, the first row shows the entropy which lies in the interval $[0, |L|]$, rescaled to $[0, 255]$. The second row of this figure shows an entropy-weighted misclassification map. In these images, the value of a

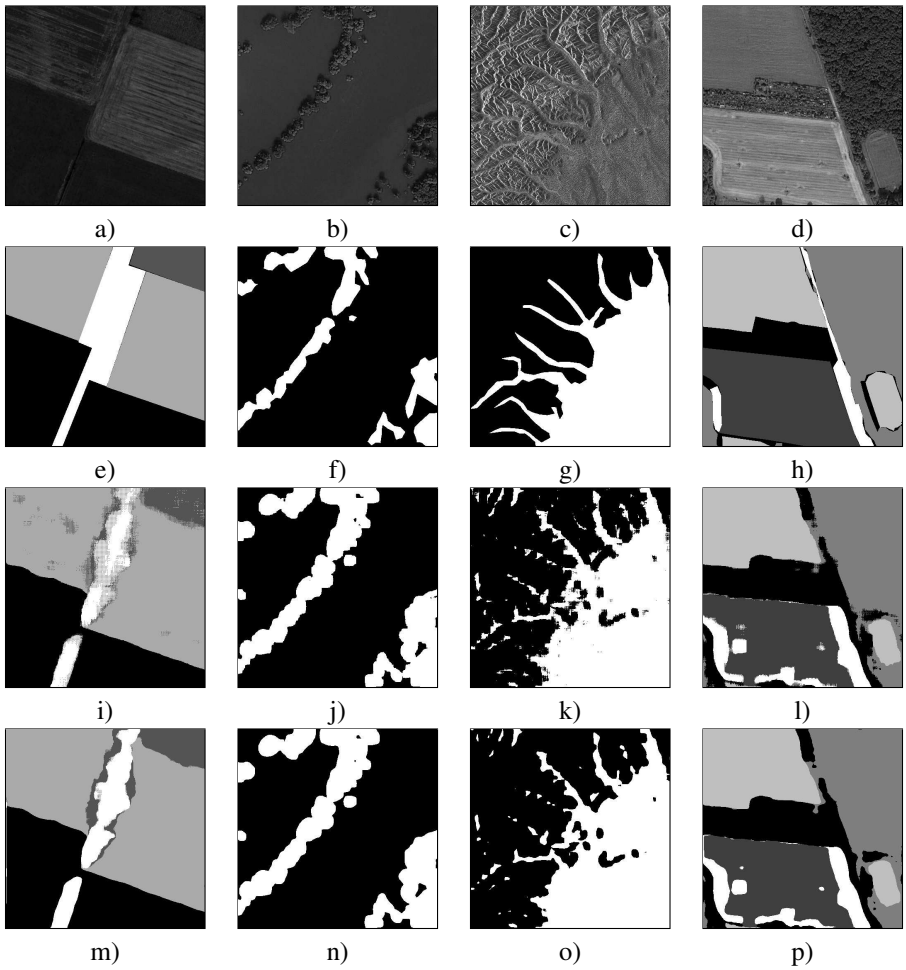

Figure 4: Original images: a) Aerial, 1m resolution ©IGN, b) IKONOS satellite, 1m resolution ©Space Imaging, c) SAR, d) IKONOS satellite, 1m resolution ©Space Imaging; Manually created ground truth images: e) – h); Neighbourhood-1 classifications: i) – l); Neighbourhood-7 classifications: m) – p).

	Figure 4(a)	**Figure 4(b)**	**Figure 4(c)**	**Figure 4(d)**
Neighbourhood-1	0.81	0.70	0.74	0.78
Neighbourhood-7	0.81	0.68	0.75	0.78

Table 1: Kappa values for the segmentation results presented in figure 4.

	Classes (in order of increasing grey level)
Figure 4(a)	unploughed field, scrubby field, ploughed field, field border
Figure 4(b)	flood, trees
Figure 4(c)	mountains, plain
Figure 4(d)	scrub, ploughed field, woodland, unploughed field, field border

Table 2: Classes defined for the ground truth images in figure 4.

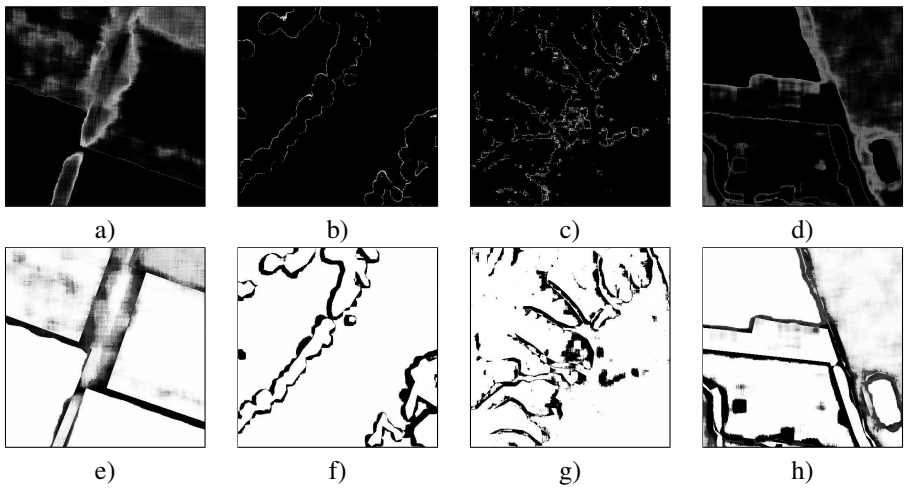

Figure 5: Entropy images for the neighbourhood-1 classifications of figure 4: q) – t); Corresponding entropy-weighted misclassification maps: u) – x).

pixel is given by $e^{-S(\pi_x)}[2\delta(k(x),k^*(x))-1]$ where k is the algorithm output, and k^* is the manually created ground truth. (Note that e^S is analogous to the number of classes that were confused by the algorithm.) Thus values greater than 128 represent 'correct', with certainty increasing with increasing intensity, while values less than 128 represent 'incorrect', with certainty increasing with decreasing intensity. These images show that the method is uncertain about its segmentation primarily at the boundaries between entity regions, as would be expected. In the interior of entity regions, the entropy is generally extremely low.

6 Conclusions

Texture is an important attribute of the appearance of many semantic entities of interest in remote sensing images. In this paper, we have described a new adaptive probabilistic model for texture description and applied it to the supervised segmentation of such images. Wavelet packet bases, which arise naturally within our probabilistic framework, allow the model to adapt to an individual texture class, and in doing so to capture its structure, and in particular the long-range correlations that create its principal periodicities. The method works well on a variety of textures, despite its simplicity. Two outstand-

ing issues to be resolved are the estimation of the parameter β in the prior on the set of trees, and the choice of a neighbourhood size. In addition, there are two sets of issues that point to future work. First, in some cases the models are too simple, and second, the segmentation method is too rough. We discuss these points in turn.

The presence of a narrow band of frequencies that runs throughout a texture means that the statistics of the wavelet packet coefficients with support in that band are no longer clustered around zero (although they have mean zero), in contrast to the statistics reported for standard wavelet coefficients [10]. Gaussian models for such coefficients are thus not accurate for modelling the statistics of adaptive bases.

The models described in this paper are not rotation invariant, although in principle this is just as much of a requirement as translation invariance. One can impose rotation invariance on a Gaussian distribution, but the results are not that interesting since the inverse covariance must now be a function only of the magnitude of the frequency. Non-trivial rotation invariance can be achieved by mixtures over rotations or by otherwise increasing the complexity of the model.

Finally, translation-invariant Gaussian models have no way of taking into account the relative phase of Fourier components, and yet phase is critical for texture structure.

All three of these modelling issues will be treated in future work by extending the Gaussian model through the addition of quartic terms to the exponent.

The segmentation method used above also possesses a drawback. For small neighbourhoods the results are noisy, while for larger neighbourhoods the smoothing effect distorts the shape and size of entity regions. Future work will focus on the satisfaction of condition 2, and the use of iterative optimization algorithms with regularizing priors to improve the accuracy of the segmentations.

References

[1] M. Acharyya and M. K. Kundu. Adaptive basis selection for multi texture segmentation by M-band wavelet packet frames. In *Proc. IEEE ICIP*, pages 622–625, 2001.

[2] K. Brady, I. Jermyn, and J. Zerubia. Accepted to *IEEE ICIP*, 2003., September 2003.

[3] T. Chang and C.C.J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. IP*, 2(4):429–441, 1993.

[4] H. Choi and R. Baraniuk. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Trans. IP*, 10(9):1309–1321, 2001.

[5] R. R. Coifman and M. V. Wickerhauser. Entropy-based methods for best basis selection. *IEEE Trans. Info. Th.*, 38(2):713–718, 1992.

[6] G. Van der Wouwer. *Wavelets for Texture Analysis*. PhD thesis, University of Antwerp, 1998.

[7] J. M. Francos, A. Z. Meiri, and B. Porat. A unified texture model based on a 2-d wold like decomposition. *IEEE Trans. SP*, 41:2665–2678, 1993.

[8] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Trans. PAMI*, 15(11):1186–1190, 1993.

[9] S. Livens, P. Scheunders, G. Van de Wouwer, and D. Van Dyck. Wavelets for texture analysis, an overview. In *Proc. ICIP and its Applications*, pages 581–585, 1997.

[10] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[11] T. Randen and J. H. Husoy. Filtering for texture classification: A comparative study. *IEEE Trans. PAMI*, 21(4):291–310, 1999.

[12] M. Tuceryan and A. K. Jain. *Texture Analysis*, chapter 2.1. World Scientific, 1998.