TERRAIN CLASSIFICATION IN POLARIMETRIC SAR USING WAVELET PACKETS

Nirmal Keshava and José Moura *

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

POL-SAR data acquired from the two 1994 flights of the SIR-C/X-SAR platform has illustrated the variability of measurements due to seasonal, spectral, and angular changes. Consequently, statistical techniques for terrain classification make robust, unsupervised classification problematic. We present an algorithm for classifying terrain that accounts for variability in terrain signatures by deriving a single representative process for each terrain from a family of stochastic scattering models. A best-basis search through a wavelet packet tree, using the Bhattacharyya coefficient as a cost measure, determines the optimal unitary basis of eigenvectors for the representative process and offers a scale-based interpretation of the scattering phenomena. The associated eigenvalues and means are determined through iterative algorithms. The technique is illustrated with a simple example.

1. INTRODUCTION

One useful application for data collected from polarimetric synthetic aperture radar (POL-SAR) platforms is the classification of vegetated terrain. The basis for classification rests on the fact that different terrains (e.g., forest, desert, bodies of water) scatter incident radar signals differently, but predictably. The statistics of the scattering reflect the underlying phenomonology and serve as a "signature" for that terrain. Classifying a scene imaged by a POL-SAR platform requires knowledge of the scattering statistics for every category of terrain and a decision test that uses the signatures to select the terrain best described by the input data. Recent efforts at terrain classification using POL-SAR data [1, 2, 3] have made significant progress in classifying vegetated terrains using statistical methods.

The phenomonology described by the statistics belonging to a terrain, however, is not invariant. Factors such as the incidence angle of the platform, atmospheric effects, and seasonal changes lead to different statistical models for the same terrain. The net result is that the design of an optimal classification test for a specific terrain type becomes problematic, due to blurred signatures.

POL-SAR data of the Earth was collected in 1994 from the SIR-C platform [4] aboard the space shuttle at two frequencies, 1250 MHz (L-band) and 5300 MHz (C-band) during two ten-day flights in April and September/October of 1994. Measurements were collected during the second mission at numerous incidence angles providing multiple observations of the same terrain. The by-product of temporal, spectral, and angular variation is a family of statistics that each reflect a unique radar-terrain phenomenology. Recent efforts at terrain classification [5] have acknowledged the effect of temporal changes in terrain statistics, but no comprehensive algorithm for robust classification exists at this time.

In this paper we develop a framework for robust terrain classification that creates a single *representative* statistical description of a terrain from a family of empirically-derived statistics. Wavelet packet bases serve as building blocks for the representative description. The advantage of such a classifier is the ability to correctly categorize terrain even when its scattering statistics have migrated from the original signature, while at the same time minimizing the dimensionality of the algorithm. This flexibility permits the algorithm to operate on a broader spectrum of input data than less robust algorithms that require frequent supervision.

2. PROBLEM FORMULATION

The scattering statistics of a terrain are frequently organized in a polarization covariance matrix (PCM), which relates the four channels of co-polar and cross-polar radar returns, $X = [x_{HH}\ x_{HV}\ x_{VH}\ x_{VV}]$, recorded in each resolution cell. A terrain covariance matrix (TCM) is a spatial extension of a PCM that relates four-channel data from one or more resolution cells. For example, a TCM, K, for four channels of POL-SAR data obtained from two adjacent cells is defined as:

$$K = E\{(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{X} - \mu_{\boldsymbol{X}})^H\} \tag{1}$$

with $\mu_{\boldsymbol{X}}$ being the mean vector for $\boldsymbol{X}$, which is

$$\boldsymbol{X} = [\ X(1)\quad X(2)\ ]^H \tag{2}$$

where

$$X(1) = [\ x_{HH}(1)\quad x_{HV}(1)\quad x_{VH}(1)\quad x_{VV}(1)\ ] \tag{3}$$

$$X(2) = [\ x_{HH}(2)\quad x_{HV}(2)\quad x_{VH}(2)\quad x_{VV}(2)\ ] \tag{4}$$

We construct a representative process from a family of processes by *maximizing* a scalar measure of similarity, the Bhattacharyya coefficient, between the representative process and each mean vector and covariance matrix in the family of terrain descriptions. The result is a wavelet-based process which maximizes an aggregate measure of stochastic distance. The wavelet-based TCM is constructed by inserting the eigendecomposition of its covariance matrix, as well as that of the original processes, into the analytic expression for the sum of the individual Bhattacharyya coefficients. Consequently, the maximization reduces to the optimization of the two defining quantities of the wavelet-based TCM: its unitary matrix of eigenvectors and the associated eigenvalues. The unitary matrix containing the eigenvectors is one of any orthonormal wavelet packet bases derived from a tree spawned by a single mother wavelet, and the eigenvalues are any set of acceptable values satisfying an

*This work was partially supported by DARPA through AFOSR grant F49620-96-O436.

0-8186-7919-0/97 $10.00

overall power constraint. Finally, the representative mean vector is derived in a subsequent optimization utilizing the representative TCM.

Once representative processes are constructed for each target terrain, they are inserted into a decision test that classifies radar data on a pixel-by-pixel basis.

3. MATHEMATICAL FOUNDATIONS

For two equally probable, real-valued, N-dimensional, Gaussian processes, $N(m_1, \Sigma_1)$ and $N(m_2, \Sigma_2)$, the Bhattacharyya coefficient, ρ, is given by [6]:

$$\rho(m_1, \Sigma_1, m_2, \Sigma_2) = e^{-\mu}. \tag{5}$$

Letting $\Delta m = m_1 - m_2$, μ is defined by:

$$\mu = \frac{1}{8}(\Delta m)^H [\frac{\Sigma_1 + \Sigma_2}{2}]^{-1}(\Delta m) + \frac{1}{2}\ln \frac{\mid \frac{\Sigma_1+\Sigma_2}{2} \mid}{\sqrt{\mid \Sigma_1 \mid \mid \Sigma_2 \mid}}. \tag{6}$$

ρ must necessarily be between 0 and 1, and a higher value of ρ indicates increasing similarity between the two processes.

Consider a set, (m, Σ), of Q equally probable, real-valued, N-dimensional, Gaussian, stochastic processes:

$$(m, \Sigma) = \{(m_1, \Sigma_1), (m_2, \Sigma_2), \ldots, (m_Q, \Sigma_Q)\}. \tag{7}$$

Let $(\hat{m}, \hat{\Sigma})$ be the wavelet-based process that represents the means and TCMs in (m, Σ). The members of (m, Σ) can be thought to represent the multiple descriptions of the same terrain whose statistics have been perceived differently due to a change in observation or a change in the underlying behavior of the original process.

If the pairwise Bhattacharyya coefficient, $\rho(m_i, \Sigma_i, \hat{m}, \hat{\Sigma})$, indicates the similarity between (m_i, Σ_i) and $(\hat{m}, \hat{\Sigma})$, then

$$\rho(m, \Sigma, \hat{m}, \hat{\Sigma}) = \frac{1}{Q}\sum_{i=1}^{Q} \rho(m_i, \Sigma_i, \hat{m}, \hat{\Sigma}) \tag{8}$$

represents the overall measure of similarity to be maximized.

4. THE MATCHING ALGORITHM

In [7], an algorithm is presented for optimally matching a Gaussian, wavelet-based process to one arbitrary process when the means of both processes are equal. The algorithm is extended here to match a wavelet-based process to a family of Q processes when the means are unequal. The task is addressed by first assuming the means are all equal in order to find a wavelet-based $\hat{\Sigma}$ to match $\Sigma_1, \ldots, \Sigma_Q$, and then finding an optimal $\hat{m}$ based on this result.

4.1. Equal Means: A Simpler Optimization

$\hat{\Sigma}$ can be expanded into an eigendecomposition:

$$\hat{\Sigma} = \hat{U} \cdot \hat{S} \cdot \hat{U}^H \tag{9}$$

and, for $1 \leq i \leq Q$, similarly for Σ_i:

$$\Sigma_i = U_i \cdot S_i \cdot U_i^H \tag{10}$$

where $\hat{S}$ and S_i are diagonal matrices and $\hat{U}$ and U_i are unitary matrices.

Substituting the eigendecompositions from (9) and (10) into (8) yields:

$$\rho(\Sigma, \hat{\Sigma}) = 2^{-\frac{N}{2}} \sum_{i=1}^{Q} \mid \hat{S} \mid^{\frac{1}{4}} \mid S_i \mid^{\frac{1}{4}} \mid \hat{S} + \hat{U}^H \Sigma_i \hat{U} \mid^{-\frac{1}{2}}. \tag{11}$$

Letting $\Omega(\Sigma, \hat{\Sigma}) = (\frac{1}{\rho(\Sigma, \hat{\Sigma})})^2$ and retaining only relevant terms,

$$\Omega(\Sigma, \hat{\Sigma}) = \mid \hat{S} \mid^{-\frac{1}{2}} \frac{1}{(\sum_{i=1}^{Q} \mid S_i \mid^{\frac{1}{4}} \mid \hat{S} + \hat{U}^H \Sigma_i \hat{U} \mid^{-\frac{1}{2}})^2}. \tag{12}$$

The expression for $\Omega(\Sigma, \hat{\Sigma})$ in (12) is the expression to *minimize* with respect to $\hat{U}$ and $\hat{S}$ that will result in maximizing $\rho(\Sigma, \hat{\Sigma})$.

The minimization of $\Omega(\Sigma, \hat{\Sigma})$ with respect to $\hat{U}$ and $\hat{S}$ can be viewed as the combination of two independent algorithms: 1) the minimization of $\Omega(\Sigma, \hat{\Sigma})$ with respect to $\hat{S}$ when $\hat{U}$ is fixed, and 2) the minimization of $\Omega(\Sigma, \hat{\Sigma})$ with respect to $\hat{U}$ when $\hat{S}$ is fixed.

4.2. Optimal Eigenvalues: Fixed-Point Algorithm

Consider the constrained problem of minimizing $\Omega(\Sigma, \hat{\Sigma})$ when a unitary basis matrix, $\hat{U}$, has been fixed. What remains is to find the matrix, $\hat{S}$ that contains the eigenvalues, $g_1, \ldots, g_N$, for the wavelet-based process, $\hat{\Sigma}$, that maximize $\Omega(\Sigma, \hat{\Sigma})$, subject to the constraint $g_1 + g_2 + \ldots + g_N = P$, where P is the average trace of the matrices, $\Sigma_1, \ldots, \Sigma_Q$.

In (12), let $G = \mid \hat{S} \mid^{-\frac{1}{2}}$, and allow V to be:

$$V = \frac{1}{(\sum_{j=1}^{Q} \mid S_j \mid^{\frac{1}{4}} \mid V^j \mid^{-\frac{1}{2}})^2} \tag{13}$$

where

$$V^j = \hat{S} + \hat{U}^H \Sigma_j \hat{U}. \tag{14}$$

Then, partial derivatives of $\Omega(\Sigma, \hat{\Sigma})$ may be taken with respect to g_i and set equal to zero:

$$\frac{\partial \Omega}{\partial g_i} = \frac{\partial G}{\partial g_i} V + G \frac{\partial V}{\partial g_i} \tag{15}$$

$$= -\frac{1}{2}\frac{1}{g_i} G V + G \frac{\sum_{j=1}^{Q} \mid S_j \mid^{\frac{1}{4}} \mid V^j \mid_{ii} \mid V^j \mid^{-\frac{3}{2}}}{(\sum_{k=1}^{Q} \mid S_k \mid^{\frac{1}{4}} \mid V^k \mid^{-\frac{1}{2}})^3} \tag{16}$$

$$= 0, \tag{17}$$

where $\mid V^j \mid_{ii}$ represents the i-th principal minor of $\mid V^j \mid$. Rearranging (15)-(17) yields:

$$g_i = \frac{z}{z_i} \tag{18}$$

where

$$z = \sum_{j=1}^{Q} \mid S_j \mid^{\frac{1}{4}} \mid V^j \mid^{-\frac{1}{2}} \tag{19}$$

$$z_i = 2\sum_{k=1}^{Q} \mid S_k \mid^{\frac{1}{4}} \frac{\mid V^k \mid_{ii}}{\mid V^k \mid^{\frac{3}{2}}}. \tag{20}$$

Both z and z_i, $i = 1, \ldots, N$, are functions of $g_1, \ldots, g_N$. The expression in (18) is a fixed-point algorithm for g_i, $i = 1, \ldots, N$. Letting $(\cdot)^n$ denote the n-th iteration value, and inserting a normalizing constant, p^{n-1}, to enforce the power constraint on the eigenvalues, the expression in (18) can be rewritten in vector form as:

$$\begin{bmatrix} g_1^n \\ \vdots \\ g_N^n \end{bmatrix} = p^{n-1} z^{n-1} \begin{bmatrix} \frac{1}{z_1^{n-1}} \\ \vdots \\ \frac{1}{z_N^{n-1}} \end{bmatrix}. \tag{21}$$

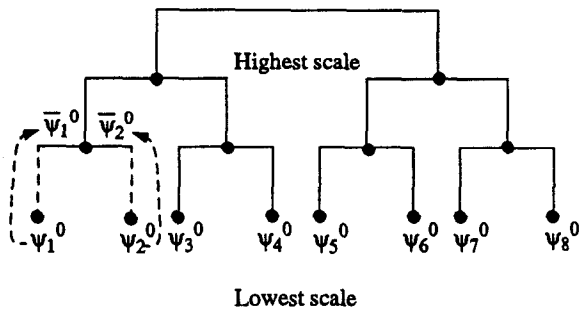

Figure 1. The wavelet packet tree

The matrix expression in (21) may be iterated to yield values for g_i, $i = 1, \ldots, N$, and hence, $\hat{S}$, that minimize $\Omega(\Sigma, \hat{\Sigma})$ when $\hat{U}$ is fixed.

4.3. Optimal Eigenvectors: Basis Migration Algorithm

Alternately, minimizing $\Omega(\Sigma, \hat{\Sigma})$ when $\hat{S}$ is fixed requires a technique to find the unitary matrix, $\hat{U}$, from a wavelet packet tree, such as the one in Figure 1. The Bhattacharyya coefficient, unfortunately, is not an additive cost function [8], i.e., the branches of the wavelet packet tree cannot be pruned independently and still lead to an optimum solution. An alternative is to pick an initial basis and allow its vectors to "migrate" up and down the branches of the tree until it arrives at a new basis which minimizes $\Omega(\Sigma, \hat{\Sigma})$.

For $i = 1, \ldots, N$, let

$$X^i = U_i^H (\hat{U}^0 \hat{S} \hat{U}^{0H}) U_i + S_i. \tag{22}$$

The equation for $\Omega(\Sigma, \hat{\Sigma})$ in (12) can be rewritten as:

$$\Omega^0(\Sigma, \hat{\Sigma}) = \mid \hat{S} \mid^{-\frac{1}{2}} \frac{1}{(\sum_{i=1}^{Q} \mid S_i \mid^{\frac{1}{4}} \mid X^i \mid^{-\frac{1}{2}})^2} \tag{23}$$

where, without sacrificing generality, assume $\hat{U}^0$, the initial choice for the unitary basis, is populated by the vectors at the bottom scale of the wavelet packet tree in Figure 1:

$$\hat{U}^0 = \left[\begin{array}{cc|cc|ccc|cc} \psi_1^0 & \psi_2^0 & \psi_3^0 & \psi_4^0 & \cdots & \psi_{N-1}^0 & \psi_N^0 \end{array} \right]. \tag{24}$$

Consider the migration of $\hat{U}^0$ to $\hat{U}^1$ as pictured in Figure 1 so that $\hat{U}^1$ is defined as:

$$\hat{U}^1 = \left[\begin{array}{cc|cc|ccc|cc} \overline{\psi_1}^0 & \overline{\psi_2}^0 & \psi_3^0 & \psi_4^0 & \cdots & \psi_{N-1}^0 & \psi_N^0 \end{array} \right]. \tag{25}$$

The change in $\Omega^0(\Sigma, \hat{\Sigma})$, $\Delta\Omega^0(\Sigma, \hat{\Sigma})$, can be shown to be approximately:

$$\Delta\Omega^0(\Sigma, \hat{\Sigma}) = \sum_{i=1}^{N} \epsilon_i \frac{\partial \Omega^0(\Sigma, \hat{\Sigma})}{\partial \epsilon_i} \tag{26}$$

where

$$\epsilon_i = \sum_{k,l=1}^{N} \delta_i(k,l) \nabla_i(k,l) \tag{27}$$

$$\delta_i = U_i^H [\sum_{i=1}^{i=2} g_i (\overline{\psi_i}^0 \overline{\psi_i}^{0H} - \psi_i^0 \psi_i^{0H})] U_i \tag{28}$$

$$\nabla_i = \mathrm{Adj}(U_i^H \hat{U}^0 \hat{S} \hat{U}^{0H} U_i + S_i) \tag{29}$$

$$\frac{\partial \Omega^0(\Sigma, \hat{\Sigma})}{\partial \epsilon_i} = \frac{\mid \hat{S} \mid^{-\frac{1}{2}} (\mid S_i \mid^{\frac{1}{4}} [\mid X^i \mid + \epsilon_i])^{-\frac{3}{2}}}{(\sum_{j=1}^{Q} (\mid S_j \mid^{\frac{1}{4}} [\mid X^j \mid + \epsilon_j])^{-\frac{1}{2}})^3}. \tag{30}$$

If $\Delta\Omega^0(\Sigma, \hat{\Sigma}) > 0$, then the migration increases $\Omega^0(\Sigma, \hat{\Sigma})$ and necessarily decreases the similarity between the processes. If $\Delta\Omega^0(\Sigma, \hat{\Sigma}) < 0$ then the new basis will result in a wavelet-based process that is closer to the family of processes in Σ, and the migration is justified. This procedure is repeated for each group of vectors in the basis. Migration of vectors to another scale only occurs if its impact is to decrease $\Omega^0(\Sigma, \hat{\Sigma})$.

4.4. Optimal Mean Vector

Having determined $\hat{\Sigma}$, $\hat{m}$ can be found to maximize (8). It can be shown that $\hat{m}$ may be determined from a fixed-point algorithm having the form of:

$$\hat{m}^n = [\sum_{i=1}^{Q} \rho(m_i, \Sigma_i, \hat{m}^{n-1}, \hat{\Sigma}) (\frac{\Sigma_i + \hat{\Sigma}}{2})^{-1}]^{-1} \cdot [\sum_{i=1}^{Q} \rho(m_i, \Sigma_i, \hat{m}^{n-1}, \hat{\Sigma}) (\frac{\Sigma_i + \hat{\Sigma}}{2})^{-1} m_i]. \tag{31}$$

It has been found that a good starting point for $\hat{m}$ is the average value of $m_1, \ldots, m_Q$.

4.5. The Complete Algorithm

The integration of the algorithms presented in Section 4.2. and 4.3. is discussed in [7]. The algorithm presented in Section 4.4. permits processes with unequal means to be matched. Together, the complete algorithm constructs a wavelet-based process that is matched in the Bhattacharyya sense to a family of Q means and covariances.

5. APPLICATIONS TO SIR-C DATA

To test the ability of the algorithm to robustly classify terrain across images, two sets of images were analyzed with the objective of designing a single classifier that satisfactorily categorizes terrains in both sets. The images were collected from the Alaskan boreal forests during the 1994 space shuttle missions. The first set consists of HH, HV, and VV images obtained at L-band. The second set consists of HH, HV, and VV images at P-band that have been co-registered with the first set. An area of the images that clearly reveals the perimeters of two abutting terrains, T_1 and T_2, was isolated and is shown in Figure 2.

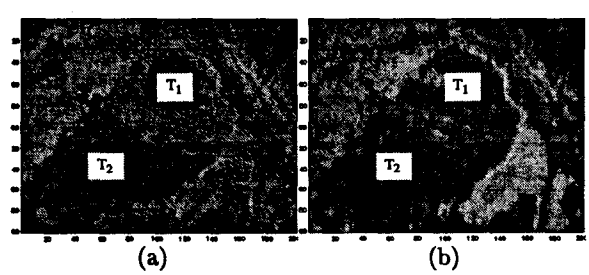

Figure 2. (a) L-band HH image of T_1 and T_2 (b) P-band HH image of T_1 and T_2.

A length-8 vector, r, was defined around the neighborhood of a central pixel to capture both spatial and polarimetric information at each frequency. If, for example,

$L_{HH}(x,y)$ represents the value of the (x,y) pixel in the HH image captured at L-band, then

$$r = \begin{bmatrix} L_{HH}(x,y) \\ L_{HV}(x,y) \\ L_{HH}(x,y-1) \\ L_{HV}(x,y-1) \\ L_{HH}(x,y+1) \\ L_{HV}(x,y+1) \\ L_{VV}(x-1,y) \\ L_{VV}(x+1,y) \end{bmatrix}. \quad (32)$$

Mean and covariance information for r was compiled for T_1 and T_2 at each frequency.

$$(L_{HH}, L_{HV}, L_{VV}) \rightarrow (m_{L,T_1}, \Sigma_{L,T_1}, m_{L,T_2}, \Sigma_{L,T_2}) \quad (33)$$
$$(P_{HH}, P_{HV}, P_{VV}) \rightarrow (m_{P,T_1}, \Sigma_{P,T_1}, m_{P,T_2}, \Sigma_{P,T_2}) \quad (34)$$

Using the quantities in (33) and (34), optimal Bayes classifiers were implemented for each frequency with the assumption of equal priors for T_1 and T_2. The classifiers were applied, pixel-by-pixel, to both sets of images to demonstrate their performance when matched, and then mismatched, to the input data set. Figure 3 illustrates the results.

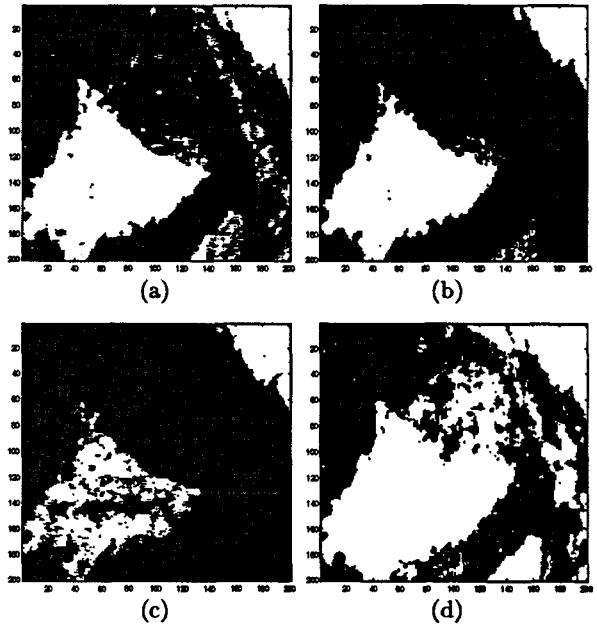

Figure 3. Classification results: (a) L-band data with L-band classifier (b) P-band data with P-band classifier (c) L-band data with P-band classifier (d) P-band data with L-band classifier.

Using the mean and covariance information in (33) and (34), representative statistics were generated for both T_1 and T_2 using the algorithm described in Section 4.

$$(m_{L,T_1}, \Sigma_{L,T_1}, m_{P,T_1}, \Sigma_{P,T_1}) \rightarrow (\hat{m}_{T_1}, \hat{\Sigma}_{T_1}) \quad (35)$$
$$(m_{L,T_2}, \Sigma_{L,T_2}, m_{P,T_2}, \Sigma_{P,T_2}) \rightarrow (\hat{m}_{T_2}, \hat{\Sigma}_{T_2}) \quad (36)$$

The quantities in (35) and (36) were used to design an optimal Bayes classifier which was applied to both sets of images. Figure 4 illustrates the performance of the wavelet-based classifier. As expected, the classifier yields performance superior to the mismatched scenarios documented in Figure 3(c) and 3(d), but not on par with the matched results in Figure 3(a) and 3(b). Using the matched classifier

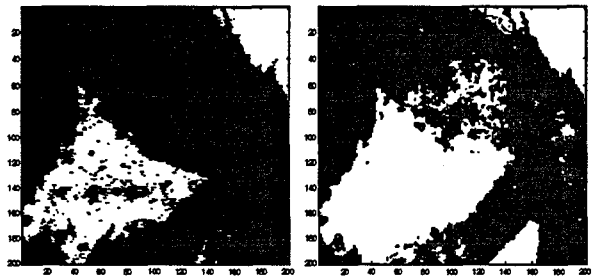

Figure 4. Classification results using optimal classifier constructed from representative statistics: (a) L-band data set (b) P-band data set.

outputs as paradigms, the wavelet-based classifier yielded 4127 misclassifications out of 40000 pixels compared to 6166 for the mismatched classifier on the L-band data, and for the P-band data, 4838 misclassifications compared to 8952. The performance of the wavelet-based classifier can be improved by locally adapting the prior values for T_1 and T_2 based on neighboring classification results, and assembling mean and covariance statistics for each region from disparate locations within the perimeter.

6. CONCLUSION

The results of the experiment in Section 5. demonstrate the feasibility of constructing robust classifiers that operate satisfactorily across families of images with little supervision and retraining. Future work will concentrate on classification algorithms that operate on several classes of terrain, integrating ground truth into the analysis. Finally, the wavelet packet representation of the terrain covariance matrices will be investigated to build a classifier that takes advantage of its unique scale-space localization properties.

REFERENCES

[1] D. R. Sheen and L. P. Johnston, "Statistical and spatial properties of forest clutter measured with polarimetric synthetic aperture radar (SAR)," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 30, pp. 578–588, May 1992.

[2] E. J. M. Rignot, C. L. Williams, J. Way, and L. A. Viereck, "Mapping of forest types in Alaskan boreal forests using SAR imagery," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 32, pp. 1051–1058, Sept. 1994.

[3] H. Anys and D.-C. He, "Evaluation of textural and multipolarization radar features for crop classification," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 33, pp. 1170–1181, Sept. 1995.

[4] R. L. Jordan, B. L. Huneycutt, and M. Werner, "The SIR-C/X-SAR synthetic aperture radar system," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 33, pp. 829–839, July 1995.

[5] L. E. Pierce, K. Bergen, M. C. Dobson, and F. T. Ulaby, "Land-cover classification using SIR-C/X-SAR data," in *IGARSS*, pp. II–918–920, July 1995.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Philadelphia, Pennsylvania: Academic Press, Inc., 1990.

[7] N. Keshava and J. M. F. Moura, "Wavelets and random processes: Optimal matching in the Bhattacharyya sense," in *Proc. of the 30th Asilomar Conf. on Sig., Sys., and Comp.*, Nov. 1996.

[8] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Info. Theory*, vol. 38, pp. 713–718, Mar. 1992.