Robust Classification of Targets In POL-SAR Using Wavelet Packets

Nirmal Keshava and José M. F. Moura
Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890
keshava@ece.cmu.edu, moura@ece.cmu.edu

***Abstract*— Recent polarimetric SAR (POL-SAR) platforms have indicated the promise of producing imagery of a scene acquired at several frequencies, incidence angles, polarizations, and at multiple time intervals. Such large amounts of data necessitate the construction of robust algorithms that can process a wide range of data with minimal retraining and supervision while minimizing the complexity of the algorithm. For statistical target detection and classification, the challenge is to build a robust classifier that operates satisfactorily even when the statistics of the desired targets migrate from a single prescribed model. In this paper, we present a technique for performing robust target classification when the targets are different terrain types. The algorithm 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 tested using images acquired from the BOREAS field campaigns in Canada.**

I. Introduction

The use of polarimetric synthetic aperture radar (POL-SAR) imagery for detecting and classifying targets has led to a substantial number of applications beyond the realm of military surveillance and reconnaissance. The ability of POL-SAR to capture the salient features of the Earth's surface through the polarization and amplitude of electromagnetic backscatter has made it a leading tool in the effort to monitor environmental processes such as deforestation and erosion on a global scale.

Accompanying the rapidly increasing amount of POL-SAR data is the need to efficiently process measurements at or beyond an acceptable standard of performance while minimizing the necessity for human supervision and system retraining. Future POL-SAR platforms will undoubtedly deliver data at numerous frequencies with increasing spatial resolution. Combined with the desire to capture temporal changes of a region by analyzinging images collected at several time intervals, it becomes easy to comprehend the necessity for efficient and autonomous signal processing algorithms.

In this paper, we introduce an algorithm that processes POL-SAR data in order to detect and classify land cover on the surface of the Earth. Recent efforts in this area have made significant progress using data from current platforms [1],[2],[3]. The emphasis on vegetated terrain as "targets" does not preclude the applicability of our work to targets of other varieties, military or otherwise. The models used to develop the algorithm only employ statistics describing the phenomenology that occurs between the incident radar and the reflecting medium and do not incorporate specific material properties of the targets. Consequently, the generality of the algorithm is retained, and the technique may be adapted for alternative applications by incorporating the statistical backscattering models for the desired targets.

The model derived for a single terrain is dependent on how the terrain is imaged. Temporal, spectral, and angular variations induce changes in the radar-target phenomenology and yield statistics for the backscatter that are subject to variation. Because the signature for each terrain is susceptible to variation, designing a single classifier to operate successfully on a wide spectrum of input data is highly problematic. Only recently have efforts in land cover classification begun to tackle this problem [4].

To address this weakness, 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 greater and more frequent supervision.

To test the performance of the algorithm, images from the BOREAS (BOReal Ecosystem Atmospheric Study) are used. Fully polarimetric data (HH, HV, VV) was collected at three frequencies (P-band, 441 MHz; L-band, 1.25 GHz; C-band, 5.36 GHz) on July 21, 1994. The images were all calibrated and radiometrically corrected

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

0-7803-3731-X/97$10.00©1997

before processing was applied.

II. 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

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

where

$$X(1) = [\ x_{HH}(1)\ \ x_{HV}(1)\ \ x_{VH}(1)\ \ x_{VV}(1)\ ] \tag{3}$$
$$X(2) = [\ x_{HH}(2)\ \ x_{HV}(2)\ \ x_{VH}(2)\ \ x_{VV}(2)\ ] \tag{4}$$

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 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 an M-ary likelihood ratio test that classifies radar data on a pixel-by-pixel basis.

III. 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 [5]:

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

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

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

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.

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

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

and, for $1 \le i \le Q$, 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:

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

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

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

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 $\rho(\Sigma, \hat{\Sigma})$.

IV. The Matching Algorithm

In [6], 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 was extended in [7] to match a wavelet-based process to a family of Q processes when the means are unequal. The detailed procedure discussed in [6],[7] is summarized here. The task is addressed by first assuming the means are all equal in order to find a wavelet-based $\hat{\Sigma}$ to match $\Sigma_1, \ldots, \Sigma_Q$. Based on this intermediate result, an optimal $\hat{m}$ is derived. 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.

A. 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:

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

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

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

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

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

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

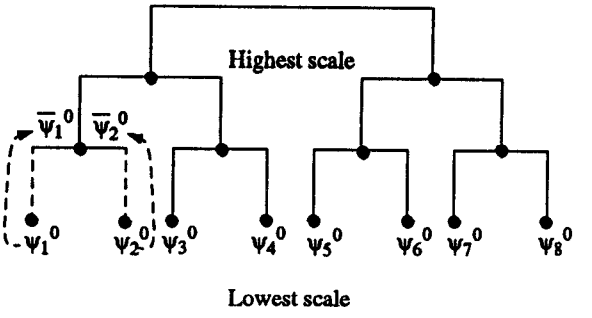

Fig. 1. The wavelet packet tree

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, $i = 1, \ldots, N$. Let $(\cdot)^n$ denote the n-th iteration value and insert, R^{n-1}, a normalizing constant. R^{n-1} assures that the total power in the eigenvalues, $g_1, \ldots, g_N$, upholds the power constraint, $g_1 + g_2 + \ldots + g_N = P$. The expression in (18) can be rewritten in vector form as:

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

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.

B. 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 $\Omega(\Sigma, \hat{\Sigma})$. 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.

$$\hat{U}^0 = [\ \psi_1^0 \quad \psi_2^0 \mid \ \psi_3^0 \quad \psi_4^0 \mid \ \cdots \ \mid \psi_{N-1}^0 \quad \psi_N^0\ ]. \tag{22}$$

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

$$X^i = U_i^H (\hat{U}^0 \hat{S} \hat{U}^{0H}) U_i + S_i. \tag{23}$$

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

$$\Omega^0(\Sigma, \hat{\Sigma}) = \mid \hat{S} \mid^{-\frac{1}{2}} \frac{1}{(\sum_{i=1}^{Q} \mid S_i \mid^{\frac{1}{4}} \mid X^i \mid^{-\frac{1}{2}})^2} \tag{24}$$

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

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

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

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

where

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

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

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

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

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 $\Omega^0(\Sigma, \hat{\Sigma})$.

C. 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:

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

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

D. The Complete Algorithm

The integration of the algorithms presented in Section A. and B. is discussed in [6]. The algorithm presented in Section C. 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.

V. Applications to AIRSAR 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 Canadian boreal forests at L-band and P-band during the AIRSAR missions and have 1024 rows and 1279 columns. Each of the two sets consist of HH, HV, and VV images. All images were co-registered and calibrated, and one pixel corresponds to a 6 m x 12 m rectangle.

Five terrains were identified in the images as having disparate characteristics. Figure 2 identifies these areas. For both frequencies, an 8×1 data vector, $r(x,y)$, was defined for the (x,y)-th pixel point that included data from two polarizations at the specified pixel location as well as from pixels in the surrounding neighborhood. For example, using the images corresponding to L-band, this experiment defined $r(x,y)$ as:

$$r(x,y) = \left[\begin{array}{c} L_{HH}(x,y) \\ L_{HV}(x,y) \\ L_{HH}(x,y-1) \\ L_{HV}(x,y-1) \\ L_{HH}(x,y+1) \\ L_{HV}(x,y+1) \\ L_{VV}(x-1,y) \\ L_{VV}(x+1,y) \end{array} \right]. \quad (32)$$

For each of the five terrains identified in Figure 2, mean and covariance information was collected from 1800 pixels to provide a statistical signature for that terrain type.

$$(L_{HH}, L_{HV}, L_{VV}) \rightarrow (m_{L,T_i}, \Sigma_{L,T_i}) \quad (33)$$

$$(P_{HH}, P_{HV}, P_{VV}) \rightarrow (m_{P,T_i}, \Sigma_{P,T_i}) \quad (34)$$

To generate a representative description for each of the five terrain types, the two mean vectors and covariance matrices describing the same terrain at different frequencies are used to create a representative mean vector and covariance matrix using the algorithm described in Section IV.

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

The Daubechies-8 wavelet was used to spawn a full, orthonormal, dyadic wavelet packet tree. Table I summarizes the values of the Bhattacharyya coefficient when comparing the resulting representative descriptions for all five terrains. The diagonal values of Table I, which is symmetric, are equal to one since the Bhattacharyya coefficient of two identical processes must necessarily be one. By inspecting the values in the first row, T_1 has virtually no similarity to T_2, T_3, and T_4, and only marginal similarity to T_5. Conversely, the high entries in the second,

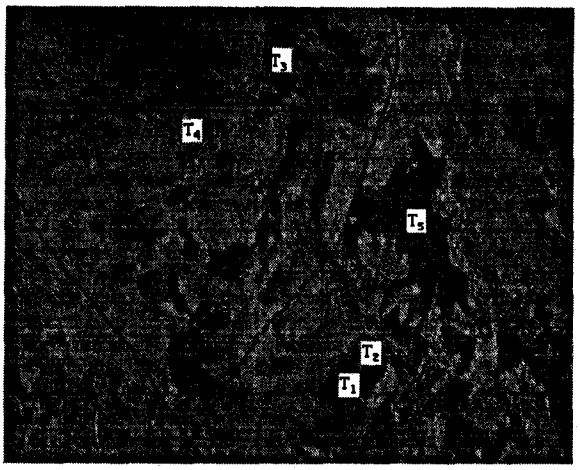

Fig. 2. P-band HH image with T_1, T_2, T_3, T_4, and T_5

	T_1	T_2	T_3	T_4	T_5
T_1	1.0000	0.0319	0.0366	0.0029	0.1528
T_2	0.0319	1.0000	0.8410	0.4576	0.0000
T_3	0.0366	0.8410	1.0000	0.4459	0.0000
T_4	0.0029	0.4576	0.4459	1.0000	0.0000
T_5	0.1528	0.0000	0.0000	0.0000	1.0000

TABLE I
Bhattacharyya Coefficient Values for Representative Processes of Terrains

third, and fourth rows indicate that T_2, T_3, and T_4 possess considerable pairwise similarity.

The quantities described in (35) are used to design an optimal M-ary Bayes classifier. The classifier relies on the result of $M-1$ likelihood ratio tests, which for two N-dimensional, Gaussian processes, $(m_a, \Sigma_a), (m_b, \Sigma_b)$, having prior probabilities, P_a and P_b, is defined by [9]:

$$(x-m_a)^H \Sigma_a^{-1}(x-m_a) - (x-m_b)^H \Sigma_b^{-1}(x-m_b) \underset{H_a}{\overset{H_b}{\gtrless}} 2\log(\frac{P_b}{P_a}) + \log(\mid \Sigma_b \mid) - \log(\mid \Sigma_a \mid). \quad (36)$$

To provide a benchmark of performance, two additional classifiers were applied to each set of images. The first is the optimal P-band Bayes classifier which relies on the terrain statistics obtained from the P-band images; its performance on the P-band images will be considered the paradigm in comparison to other classifiers. The optimal L-band Bayes classifier is constructed similarly and is a paradigm for all other classifiers operating on the L-band images.

Tables II–V are confusion matrices illustrating the performance of the three classifiers on the L-band and P-band images. Diagonal entries represent the proportion of pixels classified correctly, and off-diagonal entries represent

Pixel Class	Classifier assignment				
	T_1	T_2	T_3	T_4	T_5
T_1	0.5820	0.1667	0.2463	0.0038	0.0012
T_2	0	0.1935	0.3509	0.4556	0
T_3	0.0003	0.5168	0.3779	0.1050	0
T_4	0	0	0.2285	0.7715	0
T_5	0.4406	0	0	0	0.5594

TABLE II
Confusion matrix for wavelet classifier on L-band images

Pixel Class	Classifier assignment				
	T_1	T_2	T_3	T_4	T_5
T_1	0.2528	0.5223	0.0852	0.0054	0.1342
T_2	0	0.0192	0.3464	0.6343	0
T_3	0	0.3971	0.4299	0.1730	0
T_4	0	0	0.1995	0.8005	0
T_5	0.1952	0	0	0	0.8048

TABLE III
Confusion matrix for P-band classifier on L-band images

the proportion of pixels in a terrain class that were misclassified. For example, in Table II, the entry in the first row and second column, 0.1667, indicates that 16.67% of all pixels classified as belonging to T_1 by the optimal L-band Bayes classifier were classified by the wavelet classifier as belonging to T_2.

VI. Discussion and Future Work

The results documented in Section V. suggest that classification using the wavelet-based representative processes exceeds the performance when the data and classifier are mismatched. As expected, the performance falls below that of the paradigm, i.e., when the image and the classifier are perfectly matched.

In Tables II and III, a comparison of the diagonal values shows that the wavelet-based classifier outperforms the P-band classifier on the L-band images for T_1 and T_2, provides comparable performance for T_3 and T_4, but falls short for T_5. Tables IV and V demonstrate significantly better overall performance by the wavelet-based classifier. It is worth noting that for T_2, the $(2,2)$ entries

Pixel Class	Classifier assignment				
	T_1	T_2	T_3	T_4	T_5
T_1	0.8104	0	0	0	0.1896
T_2	0.2144	0.4082	0.3604	0.0170	0
T_3	0.0104	0.0827	0.8666	0.0403	0
T_4	0	0.0560	0.0376	0.9064	0
T_5	0.1913	0.0005	0.0005	0.0001	0.8076

TABLE IV
Confusion matrix for wavelet classifier on P-band images

Pixel Class	Classifier assignment				
	T_1	T_2	T_3	T_4	T_5
T_1	0.6374	0	0	0	0.3626
T_2	0.4309	0.0016	0.5675	0	0
T_3	0.0277	0.0571	0.7221	0.1932	0
T_4	0.0010	0.0605	0.2180	0.7205	0
T_5	0.1371	0	0	0	0.8629

TABLE V
Confusion matrix for L-band classifier on P-band images

for Tables II and III demonstrate that the wavelet-based classifier delivered performance one order of magnitude better than the P-band classifier for L-band data; similarly, the $(2,2)$ entries of Tables IV and V show that the wavelet-based classifier outperformed the L-band classifier by over two orders of magnitude. In all four tables, the significant amount of confusion between T_2, T_3, and T_4 can be attributed to the corresponding high values in Table I; the similarity of the processes increases the likelihood of misclassification.

Several factors may be weighed to improve classification. The Daubechies-8 wavelet was used for this experiment, but other wavelets will yield different performance. Selecting an alternative form of classifier corroboration such as using actual ground truth to verify accuracy can boost the results of the wavelet-based classifier. The region pictured in Figure 2 is a portion of a larger region that has been classified with a high degree of accuracy using ground studies and infrared measurements. Utilizing these results will allow the wavelet-based classifier to be compared on equal footing with the optimal P-band and L-band classifiers.

Because the total number of pixels in the test image are not equally distributed among all five classes, there is good reason to believe that the assumption that each terrain type is equally likely can be improved upon to more accurately reflect the real probabilities of each terrain. One alternative to equal priors is to develop a prior model for each pixel based on the classification values in the surrounding neighborhood. Reclassification of the images using priors derived from the first iteration has led to some improvement in performance, at a cost of doubling the processing time. Prior knowledge might also be incorporated into the design of the representative processes by weighting the individual pairwise Bhattacharyya coefficients appropriately.

Finally, the use of wavelet packet bases introduces flexibility into the traditional likelihood ratio test defined in (36) that may lead to significant improvement by exploiting the ability of wavelet packet bases to systematically expose time- or frequency-specific features of an input vector. More rigorous classification can be achieved by determining how specific components of a signal may contribute to classification in a binary hypothesis test. This capability will certainly be an important consideration for future research.

VII. Acknowledgment

The authors would like to thank Dr. Sasan S. Saatchi at the Jet Propulsion Laboratory in Pasadena, California for supplying the AIRSAR data and providing guidance on its use.

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] 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.

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

[6] 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.

[7] N. Keshava and J. M. F. Moura, "Terrain classification in polarimetric SAR using wavelet packets," in *ICASSP*, Apr. 1997.

[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.

[9] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I.* John Wiley and Sons, 1968.