Matching Wavelet Packets to Gaussian Random Processes

Nirmal Keshava and José M. F. Moura, *Fellow, IEEE*

Abstract— In this paper, we consider the problem of approximating a set of arbitrary, discrete-time, Gaussian random processes by a single, representative wavelet-based, Gaussian process. We measure the similarity between the original processes and the wavelet-based process with the Bhattacharyya coefficient. By manipulating the Bhattacharyya coefficient, we reduce the task of defining the representative process to finding an optimal unitary matrix of wavelet-based eigenvectors, an associated diagonal matrix of eigenvalues, and a mean vector. The matching algorithm we derive maximizes the nonadditive Bhattacharyya coefficient in three steps: a migration algorithm that determines the best basis by searching through a wavelet packet tree for the optimal unitary matrix of wavelet-based eigenvectors; and two separate fixed-point algorithms that derive an appropriate set of eigenvalues and a mean vector. We illustrate the method with two different classes of processes: first-order Markov and bandlimited. The technique is also applied to the problem of robust terrain classification in polarimetric SAR images.

Index Terms—Basis functions, best basis search, Bhattacharyya coefficient, binary detection, classification, nonadditive cost function, random processes, wavelet packet.

I. INTRODUCTION

A S AN alternative to the Fourier transform, the wavelet transform provides a multiresolution decomposition that can methodically reveal successive levels of detail in a signal. The wavelet transform, however, is only one of many possible transforms that are derived from the same scaling/wavelet filter pair. In [1], a family of wavelet packet bases is introduced as a generalization of the wavelet transform basis, and an algorithm for determining the best basis to represent a deterministic signal based on entropy is presented. This generalization of the traditional wavelet filter bank structure permits the representation of a signal by one of many bases, each of which is constructed by a unique ensemble of scalings and translations of the same wavelet/scaling filter pair.

The idea of developing flexible, wavelet-based representations to represent deterministic signals is extended in [2] and [4], which employ information theoretic cost criteria such as entropy and rate-distortion and are particularly well suited to signal compression; the search proposed in these references is simplified by the additivity property possessed by these cost functions. For the analysis of stochastic processes, no comprehensive framework exists for optimally representing random signals using wavelets, although efforts have been made [5]–[8] to incorporate the statistics of stochastic processes into the mathematical framework of wavelets.

The motivation for using wavelet packet bases, however, stems from three useful properties of wavelet packets. Wavelets can isolate nonstationary behavior in both time and frequency and are suited for signal processing applications where transient and nonstationary phenomena occur. Hence, the desire to represent random processes using wavelets, as opposed to another arbitrary basis construction, is well justified. Second, wavelet packets permit a whole ensemble of localized representations beyond the traditional wavelet transform. Finally, the key notion of regularity in wavelets provides a launching point for further investigations in areas such as radar and medical imaging, where the notion of smoothness can be adapted to application-specific characteristics.

In this paper, we adapt wavelet packet bases to the task of approximating a set (one or more) of Gaussian random processes by a single wavelet-based representation. In considering the general problem of approximating a *family* of Gaussian processes, we note the applicability of our technique to the problem of robust classification, where the inherent variability in target signatures requires representative statistical descriptions to be derived from a set of known signatures. In order to gauge the quality of the approximation, we use the Bhattacharyya coefficient [9], [10] as a cost criteria. Unlike its close relative (the divergence), the Bhattacharyya coefficient provides an upper bound on the probability of error when detecting the presence of either one of two processes; this upper bound is a useful metric for measuring the stochastic similarity between a process and its wavelet-based approximation. As will be shown later, analytical expressions for the Bhattacharyya coefficient can be optimized to achieve better approximations. In the same way that information theoretic measures in [1] and [4] were relevant to signal compression and quantization, the Bhattacharyya coefficient exhibits unique properties that are organic to the classical tasks of detection and classification.

We restrict ourselves to N-dimensional, Gaussian random processes. By rewriting the expression for the Bhattacharyya coefficient, we can reduce the design of the representative

N. Keshava was with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA. He is now with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173 USA.

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

process to the selection of the three defining quantities of the wavelet-based process:

1) its mean vector;
2) its eigenvalues;
3) the associated $N \times N$ unitary matrix of eigenvectors.

Two separate fixed-point algorithms determine the set of eigenvalues for the wavelet-based covariance matrix and the accompanying mean vector. The unitary matrix of eigenvectors for the wavelet-based covariance matrix may be any of the admissible wavelet packet bases in the tree spawned by a wavelet/scaling filter pair. Given a value of $N \geq 2$ that is a power of 2, the number of possible bases $\ell(N)$ in a dyadic, orthonormal, wavelet packet tree is

$$\ell(N) = \left(\ell\left(\frac{N}{2}\right) + 1 \right)^2. \tag{1}$$

Given that $\ell(2) = 2$, this number grows very rapidly so that the task of exhaustively testing each basis quickly becomes unwieldy as N increases past 8. To avoid this exhaustive search, we develop a best-basis search constructed around the Bhattacharyya coefficient.

The paper first formulates the relevant mathematical concepts and derives the key expression to optimize in Section II. In Section III, we present the complete matching algorithm. Section III-A introduces the fixed-point algorithm for determining the optimal eigenvalues of the representative process. Section III-B discusses the basis migration algorithm for finding the best basis in a wavelet packet tree. Section III-C presents the algorithm for determining the optimal mean vector. Using standard Markov and bandlimited processes, quantitative results are documented in Section IV-A. To demonstrate the utility of the matching algorithm for multiple processes, Section IV-B presents the results of an algorithm for binary detection that utilizes a wavelet-based representative process and exploits the definition of the Bhattacharyya coefficient as well as the vectors of the wavelet packet basis. Section V summarizes the results of the entire paper.

II. PROBLEM FORMULATION

We consider Q Gaussian random processes $\{x_i[1], \ldots, x_i[N]\}, i = 1, \ldots, Q$. We collect the N samples of each process in the vector x_i. We assume that the Q Gaussian random processes are equally probable, real-valued, and parameterized by their means m_i and covariances Σ_i. Let $(m, \Sigma) = \{(m_1, \Sigma_1), (m_2, \Sigma_2), \ldots, (m_Q, \Sigma_Q)\}$ be the set of Q means and covariances.

Our task is to design a single wavelet-based process that represents the Q processes $x_i, i = 1, \ldots, Q$. Because the processes are Gaussian, the problem reduces to finding the pair $(\hat{m}, \hat{\Sigma})$ corresponding to the wavelet-based process that best approximates the arbitrary means and covariances in (m, Σ). In the context of the Bhattacharyya coefficient [9], [10], the problem is to find $(\hat{m}, \hat{\Sigma})$ such that they maximize

$$\rho(m, \Sigma; \hat{m}, \hat{\Sigma}) = \frac{1}{Q} \sum_{i=1}^{Q} \rho(m_i, \Sigma_i; \hat{m}, \hat{\Sigma}) \tag{2}$$
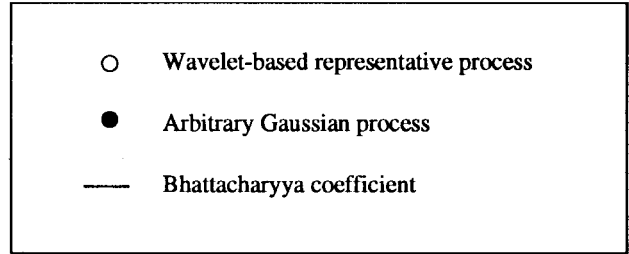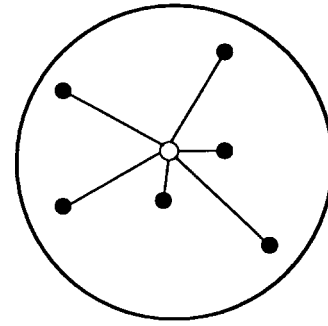

Fig. 1. Model for representative process.

where $\rho(m_i, \Sigma_i; \hat{m}, \hat{\Sigma})$ is the Bhattacharyya coefficient between (m_i, Σ_i) and $(\hat{m}, \hat{\Sigma})$. Fig. 1 illustrates the mathematical concept in (2).

A. Bhattacharyya Coefficient

We seek to simplify the expression in (2). First, we introduce the definition of the Bhattacharyya coefficient. For two random processes x_1 and x_2, whose probability density functions are denoted by $p_1(x)$ and $p_2(x)$, the Bhattacharyya coefficient is defined by the N-fold integral

$$\rho(x_1, x_2) = \int_{-\infty}^{\infty} \sqrt{p_1(x)p_2(x)} \, dx. \tag{3}$$

The value of ρ is on the closed interval $[0, 1]$.

If the processes are equally probable, real-valued, N-dimensional, and Gaussian, each having mean vectors m_1 and m_2 and covariance matrices Σ_1 and Σ_2, respectively, the Bhattacharyya coefficient $\rho(m_i, \Sigma_i; \hat{m}, \hat{\Sigma})$ is [11]

$$\rho(m_1, \Sigma_1; m_2, \Sigma_2) = e^{-\mu} \tag{4}$$

where μ varies from 0 to ∞ and is defined by

$$\mu = \frac{1}{8}(m_2 - m_1)^H \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} (m_2 - m_1)$$
$$+ \frac{1}{2} \ln \frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}. \tag{5}$$

Cost Function for Zero-Mean Processes: To proceed with the maximization of (2), we assume that the means of all Q processes are equal. This reduces the problem posed in (2) to only finding the optimal $\hat{\Sigma}$ to represent $\{\Sigma_1, \ldots, \Sigma_Q\}$. Based on this value for $\hat{\Sigma}$, an optimal $\hat{m}$ is subsequently derived from an iterative fixed-point algorithm. For equal means, we

rewrite (2) as

$$\rho(\Sigma; \hat{\Sigma}) = \frac{1}{Q} \sum_{i=1}^{Q} \rho(\Sigma_i; \hat{\Sigma}). \qquad (6)$$

We expand $\hat{\Sigma}$ and Σ_i into their eigendecompositions

$$\hat{\Sigma} = \hat{U} \cdot \hat{S} \cdot \hat{U}^H \qquad (7)$$

$$\Sigma_i = U_i \cdot S_i \cdot U_i^H, \quad 1 \le i \le Q \qquad (8)$$

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

Using the analytical definitions in (4) and (5) and substituting the eigendecompositions from (7) and (8) into (6) yields

$$\rho(\Sigma; \hat{\Sigma}) = 2^{-\frac{N}{2}} \sum_{i=1}^{Q} |\hat{S}|^{\frac{1}{4}} |S_i|^{\frac{1}{4}} |\hat{S} + \hat{U}^H \Sigma_i \hat{U}|^{-\frac{1}{2}}. \qquad (9)$$

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

$$\Omega(\Sigma; \hat{\Sigma}) = |\hat{S}|^{-\frac{1}{2}} \frac{1}{\left(\sum_{i=1}^{Q} |S_i|^{\frac{1}{4}} |\hat{S} + \hat{U}^H \Sigma_i \hat{U}|^{-\frac{1}{2}}\right)^2}. \qquad (10)$$

The maximization of $\rho(\Sigma; \hat{\Sigma})$ is now rephrased as the equivalent task of minimizing $\Omega(\Sigma, \hat{\Sigma})$ with respect to the same quantities as before: $\hat{U}$ and $\hat{S}$. We constrain $\hat{U}$ to be a unitary matrix whose columns are an admissible basis from a wavelet packet tree, such as the one depicted in Fig. 2 for $\mathbb{R}^8$. The tree in Fig. 2 is spawned by a scaling/wavelet filter pair $\{c_k, d_k\}$, which are quadrature mirror filter banks that satisfy constraints guaranteeing the orthonormal, translational, and dilational properties of the scaling function $\phi(x)$ and the mother wavelet $\psi(x)$ [12].

III. THE COMPLETE MATCHING ALGORITHM

We assume zero-mean processes and minimize $\Omega(\Sigma; \hat{\Sigma})$ as given by (10) with respect to $\hat{U}$ and $\hat{S}$ in two steps: 1) an eigenvalue fixed-point algorithm that minimizes $\Omega(\Sigma; \hat{\Sigma})$ with respect to $\hat{S}$ when $\hat{U}$ is fixed and 2) a basis migration algorithm that uses $\hat{U}$ to minimize $\Omega(\Sigma; \hat{\Sigma})$ when $\hat{S}$ is fixed.

When the processes are not zero-mean, and once $\hat{\Sigma}$ has been determined, $\hat{m}$ is found by a subsequent algorithm that is discussed in Section III-C. The pictorial description of the complete matching algorithm is shown in Fig. 3. The algorithm starts by using initial guesses for the best eigenvectors $\hat{U}^0$ and eigenvalues $\hat{S}^0$ of the representative process. Subsequent iterations of the basis migration algorithm and the fixed-point algorithm update the eigenvectors and eigenvalues, based on gradients of the existing system, until the difference between successive approximations of Σ fall below a desired tolerance, as measured by the Bhattacharyya coefficient. Once $\hat{\Sigma}$ has been determined, $\hat{m}$ is then derived through the mean vector algorithm.
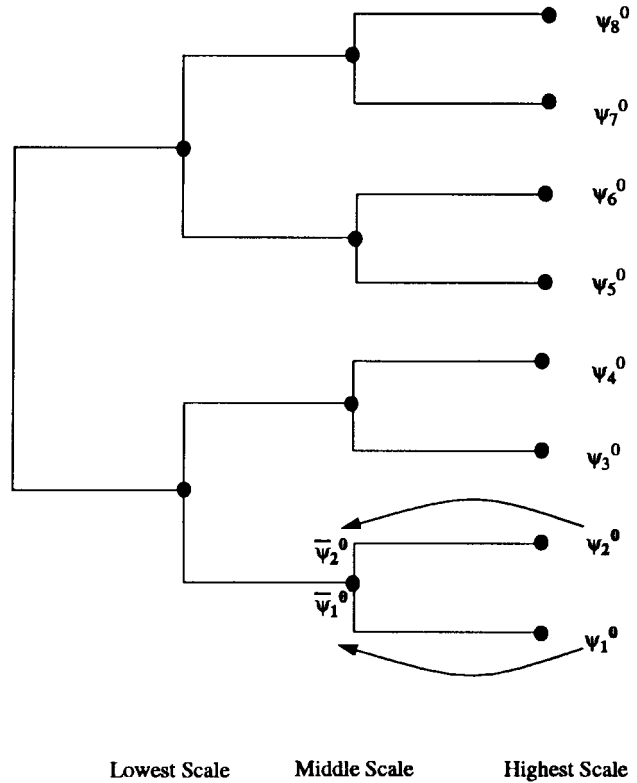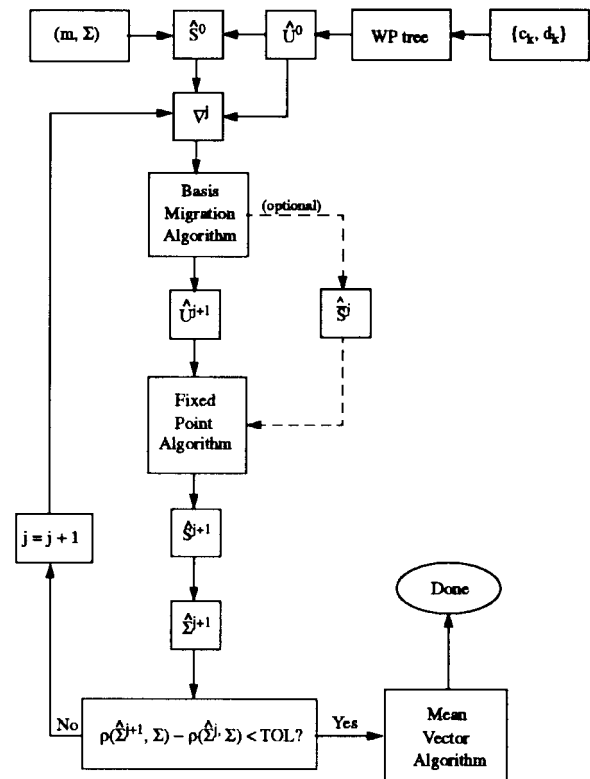

Fig. 2. Wavelet packet tree for $\mathbb{R}^8$.

Fig. 3. Complete matching algorithm.

A. Optimal Eigenvalues: Fixed-Point Algorithm

In this section, we demonstrate an algorithm for determining the set of eigenvalues for a wavelet packet basis that minimizes $\Omega(\Sigma; \hat{\Sigma})$. To do this, we seek to derive a fixed-point expression

from $\Omega(\Sigma; \hat{\Sigma})$ that can be iterated to arrive at $\hat{S}$. Continuity of the fixed-point expression, its partial derivatives, and the equivalency of its range and domain substantiate the existence of at least one fixed point. The proof appears in [13].

Consider the problem of minimizing $\Omega(\Sigma; \hat{\Sigma})$ when a unitary basis matrix $\hat{U}$ has been fixed. What remains is to find the diagonal matrix of eigenvalues $\hat{S} = \text{diag}[g_1, \ldots, g_N]$ for the wavelet-based process $\hat{\Sigma}$ that minimizes $\Omega(\Sigma; \hat{\Sigma})$ subject to $g_1 + g_2 + \cdots + g_N = P$, where P is a power constraint imposed on the eigenvalues. One useful constraint is to make P the average trace of the matrices $\Sigma_1, \ldots, \Sigma_Q$.

In (10), let $G = |\hat{S}|^{-\frac{1}{2}}$, and let V be

$$V = \frac{1}{\left(\sum_{j=1}^{Q} |S_j|^{\frac{1}{4}} |V_j|^{-\frac{1}{2}} \right)^2} \tag{11}$$

where

$$V_j = \hat{S} + \hat{U}^H \Sigma_j \hat{U}. \tag{12}$$

Then, taking partial derivatives of $\Omega(\Sigma; \hat{\Sigma})$ with respect to g_i and setting them equal to zero

$$\frac{\partial \Omega}{\partial g_i} = \frac{\partial G}{\partial g_i} V + G \frac{\partial V}{\partial g_i} \tag{13}$$

$$= -\frac{1}{2} \frac{1}{g_i} GV + G \frac{\sum_{j=1}^{Q} |S_j|^{\frac{1}{4}} |V_j|_{ii} |V_j|^{-\frac{3}{2}}}{\left(\sum_{k=1}^{Q} |S_k|^{\frac{1}{4}} |V_k|^{-\frac{1}{2}} \right)^3} \tag{14}$$

$$= 0 \tag{15}$$

where $|V_j|_{ii}$ represents the i-th principal minor of $|V_j|$. Rearranging (13)–(15) yields

$$g_i = \frac{z}{z_i} \tag{16}$$

where

$$z = \sum_{k=1}^{Q} |S_k|^{\frac{1}{4}} |V_k|^{-\frac{1}{2}} \tag{17}$$

$$z_i = 2 \sum_{j=1}^{Q} |S_j|^{\frac{1}{4}} \frac{|V_j|_{ii}}{|V_j|^{\frac{3}{2}}}. \tag{18}$$

Both z and z_i, $i = 1, \ldots, N$ are functions of $g_1, \ldots, g_N$.

The expression for g_i in (16) does not incorporate the power constraint P:

$$\sum_{i=1}^{N} g_i = P. \tag{19}$$

To include the power constraint, define a normalizing constant k, and rewrite (16) with it:

$$k = \frac{P}{\sum_{i=1}^{N} \frac{z}{z_i}} \tag{20}$$

$$g_i = k \frac{z}{z_i} \tag{21}$$

$$= \frac{P}{z_i \sum_{m=1}^{N} \frac{1}{z_m}}. \tag{22}$$

In (22), z_i is a function of $g_1, \ldots, g_N$, and by the definition of z_i in (18), z_i is guaranteed to be positive for all values of $i = 1, \ldots, N$.

The normalization of (22) by k assures that the right-hand side is bounded on the interval $(0, P)$. The partial derivatives of the right-hand side of (22) with respect to g_k, $k = 1, \ldots, N$ can also be shown to be continuous and bounded. Under these restrictions, the expression in (22) results in a fixed-point algorithm for g_i [14], and the optimal g_i, $i = 1, \ldots, N$ may be found by an iterative algorithm based on (22).

Using a superscript n to denote iteration order, the iterative expression that finds the optimal values for $g_1, \ldots, g_N$ is expressed in matrix form as

$$\begin{bmatrix} g_1^n \\ \vdots \\ g_N^n \end{bmatrix} = \frac{P}{\sum_{m=1}^{N} \frac{1}{z_m^{n-1}}} \begin{bmatrix} \frac{1}{z_1^{n-1}} \\ \vdots \\ \frac{1}{z_N^{n-1}} \end{bmatrix}. \tag{23}$$

We have found that an appropriate choice for the starting values $g_1^0, \ldots, g_N^0$ is provided by the diagonal values of the matrix A that is obtained from Σ_{avg}

$$A = \frac{P}{\text{trace}(\Sigma_{\text{avg}})} \hat{U}^{0H} \Sigma_{\text{avg}} \hat{U}^0 \tag{24}$$

$$\Sigma_{\text{avg}} = \frac{1}{Q} \sum_{i=1}^{Q} \Sigma_i. \tag{25}$$

B. Optimal Eigenvectors: Basis Migration Algorithm

We now consider the optimization complementary to the one in Section III-A: the minimization of $\Omega(\Sigma, \hat{\Sigma})$ when $\hat{S}$ is fixed by finding the best unitary matrix, $\hat{U}$, from a wavelet packet tree like the one depicted in Fig. 2 for $\mathbb{R}^8$. The Bhattacharyya coefficient is not an additive cost function, i.e., the branches of the wavelet packet tree cannot be pruned independently and still lead to an optimum solution. Best-basis search algorithms have almost exclusively utilized additive cost functions as a criteria to find the best basis for a desired application [1], [4], [15]. Additivity allows searches to be systematically disassembled into smaller units, and the "pruning" that results quickly eliminates suboptimal branches and, hence, bases, from further consideration as members of the best basis. Using (1) and noting that when $N = 8, \ell(N) = 25$, $N = 16, \ell(N) = 676$, and when $N = 32, \ell(N) = 458329$, the efficiency that results from additivity is valuable as N increases.

However, a large number of desirable distance and cost measures are not additive. Consequently, for an application in which the optimal wavelet packet basis must be found, the search is neither analytically nor computationally straightforward. Nonadditive cost measures do not permit the wavelet packet tree to be pruned irrevocably, i.e., no vector or basis may be uncategorically eliminated from consideration. The implication for a best basis search algorithm using a nonadditive cost measure is that the surface it must traverse from any starting point may have local extrema.

The alternative we use to minimize $\Omega(\Sigma; \hat{\Sigma})$ when $\hat{S}$ is fixed 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 that minimizes the expression for $\Omega(\Sigma; \hat{\Sigma})$ in (10). While the nonadditivity of our cost function does not permit us to analytically guarantee that we always arrive at

the optimal basis, our method relies on gradient-based values to steer the algorithm toward bases that minimize $\Omega(\Sigma; \hat{\Sigma})$. Our observations show that a careful choice of the starting basis results in a final basis that, if not the best basis, is very close to being optimal.

We introduce a superscript for variables to indicate their values at successive iterations of the basis migration algorithm. Without sacrificing generality, assume that $\hat{U}^0$, which is the initial choice for the unitary basis, is populated by vectors at the highest scale of the wavelet packet tree in Fig. 2.

$$\hat{U}^0 = \begin{bmatrix} \psi_1^0 & \psi_2^0 \end{bmatrix} \quad \psi_3^0 \quad \psi_4^0 \end{bmatrix} \quad \cdots \quad \begin{bmatrix} \psi_{N-1}^0 & \psi_N^0 \end{bmatrix}. \quad (26)$$

The vertical bars ($|$) in (26) demarcate vector pairs that are descendants of a common branch in the wavelet packet tree. Recalling the eigendecompositions in (7) and (8), for $i = 1, \ldots, Q$, let

$$X_i^0 = U_i^H (\hat{U}^0 \hat{S} \hat{U}^{0H}) U_i + S_i. \quad (27)$$

X_i^0 is an $N \times N$ matrix that designates the ith *system* comprised by Σ_i, $\hat{U}^0$, and $\hat{S}$. The equation for $\Omega(\Sigma; \hat{\Sigma})$ in (10) can be rewritten as

$$\Omega^0(\Sigma; \hat{\Sigma}) = |\hat{S}|^{-\frac{1}{2}} \frac{1}{\left(\sum_{i=1}^{Q} |S_i|^{\frac{1}{4}} |X_i^0|^{-\frac{1}{2}} \right)^2}. \quad (28)$$

Consider the migration of $\hat{U}^0$ to $\hat{U}^1$, as pictured in Fig. 2, so that $\hat{U}^1$ is defined as

$$\hat{U}^1 = \begin{bmatrix} \bar{\psi}_1^0 & \bar{\psi}_2^0 \end{bmatrix} \quad \psi_3^0 \quad \psi_4^0 \end{bmatrix} \quad \cdots \quad \begin{bmatrix} \psi_{N-1}^0 & \psi_N^0 \end{bmatrix}. \quad (29)$$

After lengthy algebraic manipulations [13], the change in $\Omega^0(\Sigma; \hat{\Sigma})$ incurred by exchanging $[\psi_1^0 \psi_2^0]$ for $[\bar{\psi}_1^0 \bar{\psi}_2^0]$, $\Delta \Omega_{\overline{12}}^0(\Sigma; \hat{\Sigma})$ is approximately

$$\Delta \Omega_{\overline{12}}^0(\Sigma; \hat{\Sigma}) \cong \sum_{i=1}^{Q} \epsilon_i^0 \frac{\partial \Omega^0(\Sigma; \hat{\Sigma})}{\partial \epsilon_i^0} \quad (30)$$

$$\epsilon_i^0 = \sum_{k,l=1}^{N} \left[U_i^H \delta_{\overline{12}} U_i \right]_{kl} \nabla_i^0(k, l) \quad (31)$$

$$\delta_{\overline{12}} = \sum_{j=1}^{2} \left[\overline{g_j} \bar{\psi}_j^0 \bar{\psi}_j^{0H} - g_j \psi_j^0 \psi_j^{0H} \right] \quad (32)$$

$$\nabla_i^0 = \text{Adj}(X_i^0) \quad (33)$$

$$\frac{\partial \Omega^0(\Sigma; \hat{\Sigma})}{\partial \epsilon_i} = \frac{|\hat{S}|^{-\frac{1}{2}} \left(|S_i|^{\frac{1}{4}} \left[|X_i^0| + \epsilon_i^0 \right] \right)^{-\frac{3}{2}}}{\left(\sum_{j=1}^{Q} \left(|S_j|^{\frac{1}{4}} \left[|X_j^0| + \epsilon_j^0 \right] \right)^{-\frac{1}{2}} \right)^3}. \quad (34)$$

In (33), $\text{Adj}(\cdot)$ is the adjoint of the matrix, which, for a nonsingular matrix M, is related to its inverse by

$$\text{Adj}(M) = |M| M^{-1}. \quad (35)$$

The (j, i)th element of $\text{Adj}(M)$ is $(-1)^{i+j}$ times the (i, j)th minor of M. Let m_{ij} be the (i, j)th element of M. If $|M|_{ij}$ is the (i, j)th minor of M formed by taking the determinant of M after its ith row and jth column have been removed,

the (j, i)th element of $\text{Adj}(M)$ and the derivative of $|M|$ with respect to m_{ij} are

$$[\text{Adj}]_{ji} = (-1)^{i+j} |M|_{ij} \quad (36)$$

$$\frac{\partial |M|}{\partial m_{ij}} = (-1)^{i+j} |M|_{ij}. \quad (37)$$

∇_i^0 acts as an element-by-element first derivative of the current ith system given by X_i^0. $\delta_{\overline{12}}$ is $N \times N$ and represents the change incurred by X_i^0 due to the exchange of eigenvectors $[\psi_1^0 \psi_2^0]$ for $[\bar{\psi}_1^0 \bar{\psi}_2^0]$ in the covariance matrix of the representative process. When ∇_i^0 is multiplied element-by-element by $\delta_{\overline{12}}$ and summed, the result yields ϵ_i^0, which is the effective impact of the migration on the ith system. The final summation in (30) combines the impacts from each of the Q systems with appropriate weights for each system given by (34).

The exchange of $[\psi_1^0 \psi_2^0]$ with $[\bar{\psi}_1^0 \bar{\psi}_2^0]$ can be accompanied by assigning each of the eigenvalues g_1 and g_2 to each of the new vectors $[\bar{\psi}_1^0 \bar{\psi}_2^0]$. However, the new pair of vectors will be oriented differently than the pair they replace; hence, the allocation of the total power $g_1 + g_2$ in the subspace spanned by $[\bar{\psi}_1^0 \bar{\psi}_2^0]$ should be different in order to decrease $\Omega^0(\Sigma; \hat{\Sigma})$. In (32), $\overline{g}_1$ and $\overline{g}_2$ are a new distribution of the power in g_1 and g_2 when moving from one scale to another. This *local* reallocation of power among vectors on a common branch is performed to maximize the similarity between Σ and $\hat{\Sigma}$ and, ultimately, enhances the speed of the complete algorithm. A significant point about this reallocation is that the total power *within* a branch is reorganized, yet completely conserved. The eigenvalue fixed-point algorithm described in Section III-A is responsible for *global*, optimal redistribution of power in the eigenvalues, i.e., *between all vectors* while maintaining the overall power constraint. The values for $\overline{g}_i$, $i = 1, 2$ are

$$\overline{g}_i = \frac{\bar{\psi}_i^{0H} \Sigma_{\text{avg}} \bar{\psi}_i^0}{\text{trace}\left(\left[\bar{\psi}_1^0 \bar{\psi}_2^0 \right]^H \Sigma_{\text{avg}} \left[\bar{\psi}_1^0 \bar{\psi}_2^0 \right] \right)} (g_1 + g_2). \quad (38)$$

For $\hat{U}^0$ in (26), $\Delta \Omega_{\overline{34}}^0(\Sigma; \hat{\Sigma})$, $\Delta \Omega_{\overline{56}}^0(\Sigma; \hat{\Sigma})$, and $\Delta \Omega_{\overline{78}}^0(\Sigma; \hat{\Sigma})$ are determined by selecting the appropriate pairs of vectors from the wavelet packet tree for $\delta_{\overline{34}}$, $\delta_{\overline{56}}$, and $\delta_{\overline{78}}$. Migration of vectors to another scale only occurs if its impact is to decrease $\Omega^0(\Sigma; \hat{\Sigma})$. If $\Delta_{\overline{ij}} \Omega^0(\Sigma; \hat{\Sigma}) > 0$, then the migration increases $\Omega^0(\Sigma; \hat{\Sigma})$. If $\Delta_{\overline{ij}} \Omega^0(\Sigma; \hat{\Sigma}) < 0$, then the new basis will result in a wavelet-based process that is closer to the family of covariances in $\{\Sigma_1, \ldots, \Sigma_Q\}$, and the migration is justified. The migration to be executed is the one whose value of $\Delta_{\overline{ij}} \Omega^0(\Sigma; \hat{\Sigma})$ is the smallest, and $\hat{U}^1$ is equal to $\hat{U}^0$ modified by the appropriate swap of vectors. However, if no value of $\Delta_{\overline{ij}} \Omega^0(\Sigma; \hat{\Sigma})$ is less than zero, then none of the possible migrations decreases $\Omega^0(\Sigma; \hat{\Sigma})$, and $\hat{U}^1 = \hat{U}^0$. With the new basis $\hat{U}^1$, a new migration is initiated by calculating ∇^1, using $\hat{U}^1$ and $\overline{S}$.

Inherent in this migration rule, however, is the fact that the expression for $\Delta_{\overline{ij}} \Omega(\Sigma; \hat{\Sigma})$ is only a first-order approximation to the change incurred by $\Omega(\Sigma; \hat{\Sigma})$ when the ith and jth basis vectors are replaced. If the approximation fails to hold, the migration algorithm may identify false migrations, and
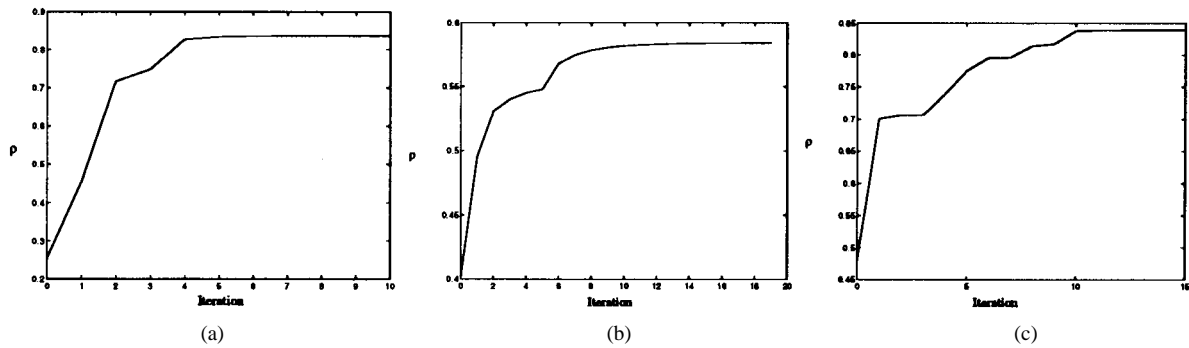
Fig. 4.   Iterative results for (a) 8×8 Markov process with $\kappa = 0.95$. (b) 8×8 bandlimited process with $\omega = 0.35$. (c) 16×16 Markov process with $\kappa = 0.75$.

under these circumstances, finding the best basis and wavelet-based random process becomes difficult. Experiments indicate that the approximation leading to $\Delta_{\overline{ij}}\Omega(\Sigma; \hat{\Sigma})$ remains useful and intact except under certain situations. These issues are discussed in detail in Sections IV and V.

C. Optimal Mean Vector Algorithm

In this section, we use the results of the algorithms derived in Sections III-A and III-B to obtain a mean vector $\hat{m}$ for the representative process. The same approach used in Section III-A to determine $\hat{S}$ is employed here to find $\hat{m}$: Derive a fixed-point expression that can be iterated to yield a value of $\hat{m}$ that maximizes $\rho(m, \Sigma; \hat{m}, \hat{\Sigma})$. Continuity of the fixed-point expression, its partial derivatives, and the equivalency of its range and domain substantiate the existence of at least one fixed point. The proof appears in [13].

Having determined $\hat{\Sigma}$, the representative mean vector $\hat{m}$ is found by maximizing (2). Starting with (2), and using (4) and (5), let $\hat{m}(j)$ denote the jth element of $\hat{m}$. Then, taking partial derivatives of (2) with respect to $\hat{m}_j$

$$\frac{\partial \rho(m, \Sigma; \hat{m}, \hat{\Sigma})}{\partial \hat{m}(j)} = -2 \sum_{i=1}^{Q} \rho(m_i, \Sigma_i; \hat{m}, \hat{\Sigma})$$
$$\times u_j^H \left(\frac{\Sigma_i + \hat{\Sigma}}{2} \right)^{-1} (m_i - \hat{m}) \quad (39)$$

where u_j denotes a length-N vector with all zeros, except for a value of 1 at the jth element. The expression in (39) is a scalar equation and is valid for all $j, 1 \leq j \leq N$; it can be rewritten in matrix form for all values of j by replacing u_j by the identity matrix. Setting this quantity equal to zero and solving for $\hat{m}$ yields a fixed-point expression whose iterations are denoted by a superscript

$$\hat{m}^n = \left[\sum_{i=1}^{Q} \rho(m_i, \Sigma_i; \hat{m}^{n-1}, \hat{\Sigma}) \left(\frac{\Sigma_i + \hat{\Sigma}}{2} \right)^{-1} \right]^{-1}$$
$$\times \left[\sum_{i=1}^{Q} \rho(m_i, \Sigma_i; \hat{m}^{n-1}, \hat{\Sigma}) \left(\frac{\Sigma_i + \hat{\Sigma}}{2} \right)^{-1} m_i \right]. \quad (40)$$

A good starting point for $\hat{m}$ is the average value of $m_1, \ldots, m_Q$.

IV. RESULTS

We present our experimental results in two parts. In Section IV-A, we consider the problem of matching a single Gaussian process to a wavelet-based process, and we apply the complete matching algorithm to common random processes. The key analytical expressions are specialized from the formulations in Section III and appear in [13] and [16]. To illustrate how the algorithm rapidly eliminates bases from consideration, intermediate results are presented at each iteration to indicate the incremental basis migrations and the convergence of the Bhattacharyya coefficient to its final value. A discussion about using wavelets of different regularities concludes this part. In Section IV-B, we address the general case of matching multiple Gaussian processes to a wavelet-based process in the context of robust terrain classification in POL-SAR images.

A. Single-Process Matching

We consider the problem of matching a wavelet-based random process to one arbitrary, N-dimensional Gaussian random process. For the following examples, the Daubechies-4 mother wavelet is used to spawn a full dyadic wavelet packet tree. In all cases, the initial basis $\hat{U}^0$ was chosen to be the set of vectors residing on the lowest scale, which is denoted as scale 1.

For the first example, define the entries of an 8×8 first-order Markov process Σ as

$$\Sigma_{ij} = \kappa^{|i-j|} \quad 1 \leq i, \quad j \leq 8. \quad (41)$$

Fig. 4(a) and Table I summarize the results of the matching algorithm when $\kappa = 0.95$. Iterations one through four show rapid increases in the Bhattacharyya coefficient due to the exchange of basis vectors that occurs. The final five iterations not shown in Table I do not alter the basis, and the modest increase in the Bhattacharyya coefficient results from a refinement of the eigenvalues by the fixed-point algorithm.

The iterative results were compared with an exhaustive search of all 25 bases in the wavelet packet tree. This exhaustive search confirmed that the final basis in iteration nine achieved the best match between Σ and $\hat{\Sigma}$.

TABLE I
BASIS MIGRATIONS FOR 8×8 MARKOV PROCESS: $\kappa = 0.95$

Iteration	Scales of basis vectors	$\rho(\Sigma; \hat{\Sigma}^i)$
0	{1,1,1,1,1,1,1,1}	0.2547
1	{1,1,1,1,2,2,2,2}	0.4548
2	{1,1,1,1,2,2,3,3}	0.7467
4	{1,1,1,1,3,3,3,3}	0.8338
9	{1,1,1,1,3,3,3,3}	0.8362

TABLE II
BASIS MIGRATIONS FOR 8×8 BAND-LIMITED PROCESS: $\omega = 0.35$

Iteration	Scales of basis vectors	$\rho(\Sigma; \hat{\Sigma}^i)$
0	{1,1,1,1,1,1,1,1}	0.4015
1	{2,2,2,2,1,1,1,1}	0.5052
2	{3,3,2,2,1,1,1,1}	0.5401
19	{3,3,2,2,1,1,1,1}	0.5842

For the second example, consider an 8×8 bandlimited process Σ, whose entries are given by

$$\Sigma_{ij} = \frac{\sin(2\pi\omega(i-j))}{\pi(i-j)} \quad 1 \leq i, \quad j \leq 8. \qquad (42)$$

For the case when $\omega = 0.35$, Fig. 4(b) and Table II summarize the results of the complete matching algorithm. Iterations one and two show rapid increases in the Bhattacharyya coefficient as the best basis vectors are found, and the remaining iterations increase the similarity only marginally as the wavelet packet basis stays constant.

The iterative results were compared with an exhaustive search of all 25 bases in the wavelet packet tree. Again, this search confirmed that the matching algorithm led to the best match between Σ and $\hat{\Sigma}$, yielding the highest value of the Bhattacharyya coefficient.

A final example using a Markov process illustrates the algorithm for $N = 16$ and demonstrates how the matching algorithm may not always find the best basis but may find a basis that is close to being optimal. When $\kappa = 0.75$, Fig. 4(c) and Table III summarize the results of the matching algorithm. When compared with the exhaustive tests conducted for each of the admissible bases, the result generated in the final row of Table III ranks 14th out of 676. In the optimal basis, the pair of scale-3 vectors in the 13th iteration of Table III migrate to scale 4; the corresponding value of $\rho(\Sigma; \hat{\Sigma})$ is 0.8449. The matching algorithm was unable to arrive at the optimal basis because it did not swap these subspaces. The difference in matching performance, however, is considerably small.

While exhaustive verification of the matching algorithm becomes prohibitive for N larger than 16, there is no algorithmic dependence on N that hinders its operation as N increases. Matches for processes with $N = 32$ were also successfully

TABLE III
BASIS MIGRATIONS FOR 16×16 MARKOV PROCESS: $\kappa = 0.75$

Iteration	Scales of basis vectors	$\rho(\Sigma; \hat{\Sigma}^i)$
0	{1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1}	0.4803
1	{1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2}	0.7007
2	{2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2}	0.7065
4	{2,2,2,2,3,3,3,3,2,2,2,2,2,2,2,2}	0.7425
5	{2,2,2,2,4,4,3,3,2,2,2,2,2,2,2,2}	0.7768
6	{3,3,3,3,4,4,3,3,2,2,2,2,2,2,2,2}	0.7956
8	{3,3,3,3,4,4,4,4,2,2,2,2,2,2,2,2}	0.8167
10	{4,4,3,3,4,4,4,4,2,2,2,2,2,2,2,2}	0.8390
13	{4,4,3,3,4,4,4,4,2,2,2,2,2,2,2,2}	0.8394

conducted, although not exhaustively corroborated. The issue of computational complexity is addressed at the end of Section V.

1) Different Wavelets We also considered the impact of changing the scaling/wavelet filter pair that spawns the wavelet packet tree. Four different wavelets were used in the complete matching algorithm to approximate first-order Markov processes with $0 \leq \kappa \leq 1$, where κ is discretized in increments of 0.01. The results appear in Fig. 5(a). Similarly, Fig. 5(b) depicts the results when the same four wavelets were used to approximate band-limited processes having $0 \leq \omega \leq 0.50$ in increments of 0.01.

Fig. 5(a) and (b) illustrates that wavelets of higher regularity consistently yielded better matching results for all values of κ and ω. For both cases of processes having lower bandwidths (higher values of κ and lower values of ω), the matching algorithm is unable to yield approximations commensurate with performance achieved for processes having a higher bandwidth.

B. Multiple-Process Matching: POL-SAR

In this section, the complete matching algorithm is applied to the problem of robust classification of land cover types in POL-SAR. This issue was first addressed in [17] and [18]. The scenario occurs when the statistics describing the target terrains change. Variations in frequency, polarization, and observation angle induce changes in the radar-target phenomenology and yield multiple signatures for the backscatter from each terrain type, making the design of a single classifier that operates successfully on a wide spectrum of input data highly problematic.

To illustrate our technique, we develop a *representative* terrain model for each land cover type that serves as an aggregate description for the scattering behavior of a terrain under different circumstances. As would be expected, the robustness of a classifier to changes in terrain characteristics depends on both the tightness of the original cluster of signatures and the quality of the ensuing approximation. Since
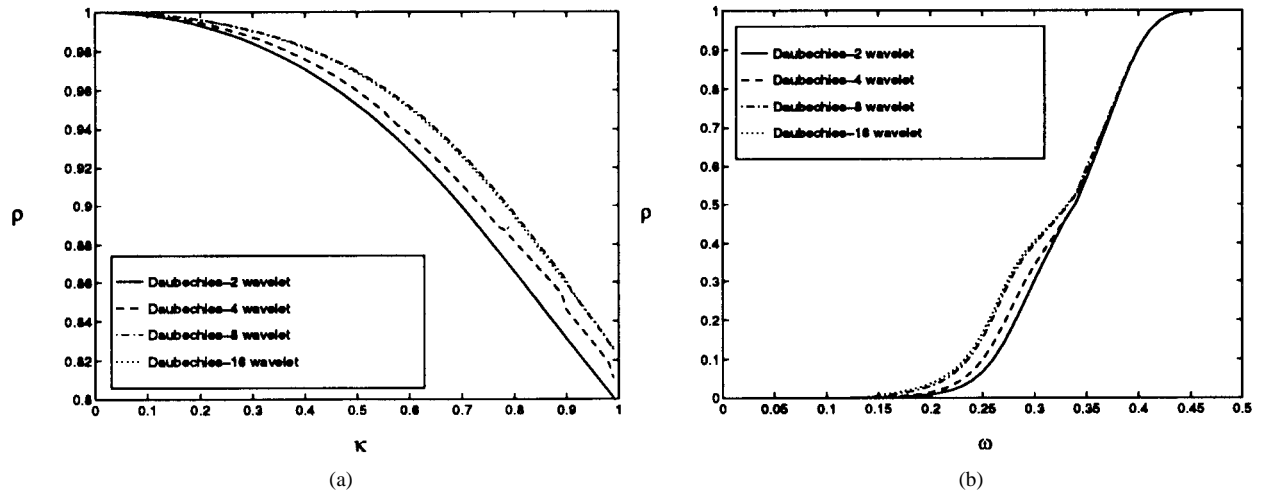
Fig. 5. (a) Bhattacharyya coefficient between first-order Markov processes and wavelet-based approximations as a function of κ. (b) Bhattacharyya coefficient between bandlimited processes and wavelet-based approximations as a function of ω.
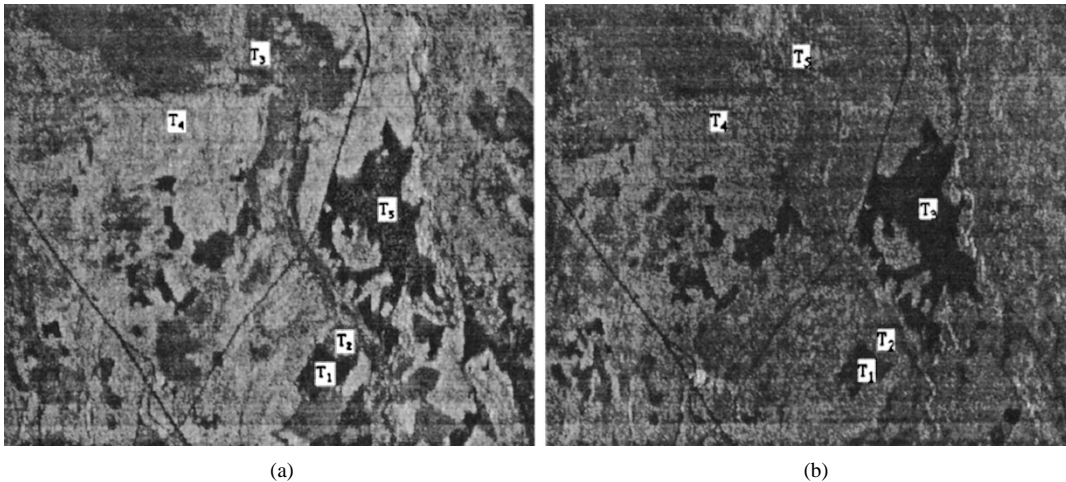

Fig. 6. (a) P-band, *HH* image. (b) P-band *VV* image.

the former is dictated by empirically collected results, we address the latter constraint using wavelet-based representative processes. To test the performance of the algorithm, we use images from the boreal ecosystem atmospheric study (BOREAS) [19]. 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 collected from Canadian boreal forests during the AIRSAR missions and have 1024 rows and 1279 columns. One pixel corresponds to approximately a 6 m × 12 m rectangle.

The experiment we propose utilizes *HH* and *VV* images acquired at P-band, which are shown in Fig. 6(a) and (b). Five different land cover types are identified: $T_1, T_2, T_3, T_4,$ and T_5. We seek to construct a representative description of each terrain type from wavelets, $T_i, i = 1, \ldots, 5$ from its *HH* and *VV* statistics. We recognize that classifiers generally do not require robustness to changes in polarization. Changes in phenomenology due to variations in incidence angle and frequency are far more critical to classifier design. However, limitations in available data constrain us to demonstrate our technique using data acquired at multiple polarizations. We emphasize, nevertheless, that our algorithmic concept for gain-

ing robustness to changes in polarization would be identical for variations in any other SAR modality.

1) Constructing Terrain Signatures: We create signatures for a land cover type in a POL-SAR image by first constructing feature vectors from pixels residing in the desired terrain. The features may be any quantity derived from the central pixel and/or its surrounding pixels. Feature vectors associated with a pixel can be as simple as a lexicographic collection of the raw pixel values surrounding the central pixel, or they may be texture values of several orders derived from a neighborhood of pixels. For our analysis, we demonstrate the classification algorithm with feature vectors of length 8 for each terrain containing a collection of eight raw pixel values in the immediate neighborhood of the pixel of interest.

We employ the common assumption that the feature vectors for a terrain obey the properties of Gaussianity; hence, to construct a probability density function for the feature vector, we need only determine the mean and covariance of the feature vector. Using pixels from each terrain group, we derive means and covariances and calculate the corresponding Bhattacharyya coefficients between classes for each polarization. Tables IV and V summarize this information.

TABLE IV
BHATTACHARYYA COEFFICIENT VALUES FOR *HH* IMAGE AT P-BAND

	T_1	T_2	T_3	T_4	T_5
T_1	1.0000	0.0584	0.1728	0.0112	0.2344
T_2	0.0584	1.0000	0.9042	0.5850	0.0000
T_3	0.1728	0.9042	1.0000	0.4602	0.0001
T_4	0.0112	0.5850	0.4602	1.0000	0.0000
T_5	0.2344	0.0000	0.0001	0.0000	1.0000

TABLE V
BHATTACHARYYA COEFFICIENT VALUES FOR *VV* IMAGE AT P-BAND

	T_1	T_2	T_3	T_4	T_5
T_1	1.0000	0.0727	0.1293	0.0057	0.2820
T_2	0.0727	1.0000	0.9576	0.6010	0.0000
T_3	0.1293	0.9576	1.0000	0.4752	0.0002
T_4	0.0057	0.6010	0.4752	1.0000	0.0000
T_5	0.2820	0.0000	0.0002	0.0000	1.0000

TABLE VI
BHATTACHARYYA COEFFICIENT VALUES FOR
REPRESENTATIVEPROCESSES CONSTRUCTED FROM P-BAND IMAGES

	T_1	T_2	T_3	T_4	T_5
T_1	1.0000	0.0690	0.0612	0.0010	0.7223
T_2	0.0690	1.0000	0.8457	0.2643	0.0118
T_3	0.0612	0.8457	1.0000	0.6859	0.0140
T_4	0.0010	0.2643	0.6859	1.0000	0.0001
T_5	0.7223	0.0118	0.0140	0.0001	1.0000

2) The Representative Process Using the statistics derived for the five terrains, we use the complete matching algorithm to generate a representative process for each terrain class $i = 1, \ldots, 5$ from its *HH* and *VV* descriptions.

$$(m_{iPHH}, \Sigma_{iPHH}; m_{iPVV}, \Sigma_{iPVV}) \rightarrow (\hat{m}_{iPW}, \hat{\Sigma}_{iPW}). \quad (43)$$

The representative process for each terrain $(\hat{m}_{iPW}, \hat{\Sigma}_{iPW})$ is constructed from the original two processes in the same way depicted in Fig. 1.

The Daubechies-4 wavelet was used to spawn a full, orthonormal, dyadic wavelet packet tree. Representative processes were generated using the *HH* and *VV* statistics for each terrain. Table VI summarizes the pairwise Bhattacharyya coefficients of the wavelet-based, representative processes.

To gauge the robustness of the representative processes, we insert the wavelet-based signatures for T_1, T_2, T_3, T_4, and T_5 into a five-class optimal Bayesian classifier. The five training regions in both the *HH* and *VV* image for each frequency are submitted to the classifier, and the rates of correct classification from both images are summed to ascertain the robustness of the classifier when two disparate images serve as the input. The classifier relies on the result of $M - 1$ likelihood ratio tests [20] that assume each terrain is equally probable.

To provide a benchmark of performance, two additional classifiers were applied to the set of images corresponding to each frequency. The first is the optimal *HH* Bayes classifier, which relies on the terrain statistics obtained solely from the *HH* image; its performance on the *HH* images will necessar-

ily yield the highest accuracy for the *HH* training regions; however, its performance with respect to the *VV* training regions will be suboptimal because of the inherent mismatch between the classifier and the statistics of the input image. The second classifier is the optimal *VV* classifier constructed from the terrain statistics obtained from the *VV* images. The performance of both the *HH*- and *VV*-based classifiers is also gauged by averaging their rates of correct classification on both the *HH* and *VV* images.

Table VII contains the results for the P-band classifiers and images. The entries represent the fraction of pixels from a terrain class in either the *HH* or *VV* image that have been classified correctly by the associated classifier. The robust average of the wavelet-based classifier exceeds the performance of the *HH*- and *VV*-based classifiers. As we expect, for an M-ary test, the performance on individual terrain classes exhibits a wide variation compared with the average and robust average values; the optimal Bayesian classifier is expected to yield the best aggregate probability of error over all terrain classes, but there is no statistical guarantee of performance for individual classes.

V. DISCUSSION

The results documented in Tables I–III show that the complete matching algorithm is able to approximate random processes by wavelet-based representations. The algorithm succeeds in replacing an exhaustive search of a wavelet packet tree for a set of eigenvectors with a fast, iterative, computationally simple procedure. In one case, we observed that although the algorithm was unable to arrive at the best basis, the final solution was nearly as good as the one obtained by the best basis.

Although the complete matching algorithm consistently found the best, or near-best, basis, the quality of the approximation sometimes still suffers. Our experiments indicated that processes having high bandwidths were better approximated by wavelet-based eigenstructures than processes with low bandwidths. The conclusion is reinforced by the curves in Figs. 5(a) and (b), which suggest that while wavelet packet bases may be unsuitable for approximating some low-bandwidth processes, a wavelet with higher regularity can improve the final result.

TABLE VII
PERFORMANCE OF P-BAND HH, VV, AND WAVELET-BASED CLASSIFIERS

Pixel Class	HH Classifier		VV Classifier		Wavelet-based Classifier	
	HH	VV	HH	VV	HH	VV
T_1	0.7905	0.8615	0.3610	0.8925	0.5475	0.9050
T_2	0.8095	0.6085	0.8285	0.8620	0.8705	0.7760
T_3	0.4365	0.7910	0.1530	0.2530	0.1860	0.4240
T_4	0.9875	0.0470	0.0215	0.7180	1.0000	0.4305
T_5	0.7275	0.4720	0.9180	0.8930	0.8635	0.7460
Avg.	0.7503	0.5560	0.4564	0.7237	0.6935	0.6563
Robust Avg.	0.6532		0.5900		0.6749	

When matching single processes, we recognized that the basis migration algorithm occupies the dominant role in the matching algorithm. This is understandable since migrations are also accompanied by a local redistribution of eigenvalues in the associated basis vectors. Intermediate updating of the eigenvalues (assuming the powers of the eigenvalues are locally redistributed in each migration) is unnecessary until the overall algorithm converges to a final basis. Then, the eigenvalue fixed-point algorithm can be iterated to obtain the associated optimal eigenvalues.

As noted earlier, the matching algorithm sometimes has difficulty identifying the migrations necessary to arrive at the best basis. This phenomenon was observed in Table III. In some instances, multiple iterations were necessary between basis migrations because the algorithm had trouble finding the migration that increased the Bhattacharyya coefficient. This can be attributed to the inaccuracy of the first-derivative approximation stated in (33) near regions around the extrema, and, as stated above, the problem is compounded in processes possessing low bandwidths. Consequently, the algorithm may halt before reaching the best basis, but the final result is usually very close to that which would be achieved by the best basis. Mechanisms installed in the algorithm prevent migrations in the backward direction toward previous bases and calculate the Bhattacharyya coefficient between iterations to assure that the migration is always in the direction of increasing the similarity between the two processes. Generally, this information is only required for the final migrations. Although this entails extra calculation, it ensures that the migration stays "on track." In comparison with the effort required to exhaustively investigate each basis in the wavelet packet tree, this feature remains insignificant and justifiable.

We address the issue of computational cost by first considering the cost incurred by an exhaustive search for the simplest case of single-process matching that appeared in Section IV-A. The exhaustive search for the best set of eigenvectors and eigenvalues can be considered separately and then multiplied to determine the number of total Bhattacharyya coefficient calculations that are necessary. For $N = 16$, the total number of wavelet packet bases is 676. The total number of sets of eigenvalues that require testing for each basis depends on the granularity of the discretized interval being searched and the power constraint. For the example documented in Table III, each basis requires an $(N - 1)$-dimensional search, where each dimension investigates the range from 0 to P, which is the total power in the process. This example requires an enormous search, and the order of both the eigenvector and eigenvalue searches increases dramatically as N increases. In contrast, the complete matching algorithm requires tests of only 13 iterations, where the most significant operation is the calculation of principal minors, which, in this case, are determinants of 15×15 matrices.

Our experiments in Section IV-A were conducted when the starting unitary basis $\hat{U}^0$ was chosen to be the basis whose vectors came exclusively from the lowest scale in the wavelet packet tree. Since any migration of vectors from the lowest scale to the next highest scale (or vice-versa) involves $N/2$ vectors, it is best to complete such a sizable migration early in the iterative process—particularly in the initial steps—when the impact of possible migrations is relatively clear and unambiguous, and the first-derivative approximations are reliable. Starting bases comprised of vectors residing on higher scales were also tested, but the basis from the lowest scale almost always provided performance at least equal to all other selections.

VI. CONCLUSION

In this paper, we developed an algorithm that constructs a single, wavelet-based process that matches a set of arbitrary, discrete-time, Gaussian, N-dimensional random processes. Because of its applicability to detection, our technique employed the Bhattacharyya coefficient to measure the quality of the approximation and used three separate algorithms to obtain the final result.

Pivotal to the complete matching algorithm is a best-basis search algorithm that finds the optimal set of eigenvectors for the approximation from a wavelet packet tree. In spite of the complications incurred by the nonadditivity of the Bhattacharyya coefficient, the algorithm is able to locate the best or near-best basis consistently through gradient-based techniques that are computationally simple and inexpensive.

Our application of the matching algorithm to the problem of robust terrain classification in POL-SAR imagery found that the representative, wavelet-based terrain models were able to deliver better overall performance when the terrain signatures varied. Table VII demonstrates this result for P-band images of Canadian boreal forests acquired at two different polarizations.

The experimental results in Section IV-A demonstrate that for first-order Markov processes and bandlimited processes, the matching algorithm consistently finds the best or near-best basis that yields the best approximation in the Bhattacharyya sense. While the curves in Fig. 5(a) and (b) demonstrate that low-bandwidth processes are not as well approximated by any wavelet packet basis as processes with higher bandwidths, the experimental observation that wavelets of higher regularity yield increasingly better approximations is a very promising result. Just as higher regularity can be proven mathematically to lead to better approximations of deterministic signals, our results at least provide empirical evidence that the same is true for certain classes of random Gaussian signals.

REFERENCES

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

[2] C. Herley, J. Kovačević, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 3341–3359, Dec. 1993.

[3] C. Herley and M. Vetterli, "Orthogonal time-varying filter banks and wavelet packets," *IEEE Trans. Signal Processing*, vol. 42, pp. 2650–2663, Oct. 1994.

[4] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[5] G. W. Wornell, "A Karhunen-Loéve-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859–861, July 1990.

[6] M. Bassevile, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, Mar. 1992.

[7] R. W. Dijkerman and R. R. Mazumdar, "Wavelet representations of stochastic processes and multiresolution stochastic models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 42, pp. 1640–1652, July 1994.

[8] J. Zhang and G. Walter, "A wavelet-based KL-like expansion for wide-sense stationary random processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 42, pp. 1737–1745, July 1994.

[9] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[10] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans.Commun. Technol.*, vol. 15, pp. 52–60, Feb. 1967.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Philadelphia, PA: Academic, 1990.

[12] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Applied Math.*, vol. 41, no. 7, pp. 909–996, Jan. 1988.

[13] N. Keshava, "Robust detection of stochastic targets using wavelet packets," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 1997.

[14] R. L. Burden and J. D. Faires, *Numerical Analysis*. Boston, MA: PWS, 1985.

[15] C. Taswell, "Near-best basis selection algorithms with nonadditive information cost functions," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Oct. 1994, pp. 13–16.

[16] N. Keshava and J. M. F. Moura, "Wavelets and random processes: Optimal matching in the Bhattacharyya sense," in *Proc. 30th Asilomar Conf. Signals, Syst., Comput.*, Nov. 1996.

[17] ——, "Terrain classification in polarimetric SAR using wavelet packets," in *Proc. ICASSP*, Apr. 1997, pp. 551–554.

[18] ——, "Robust classification of targets in POL-SAR using wavelet packets," in *Proc. 1997 IEEE Nat. Radar Conf.*, May 1997, pp. 105–110.

[19] 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. Remote Sensing*, vol. 32, pp. 1051–1058, Sept. 1994.

[20] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I.* New York: Wiley, 1968.

Nirmal Keshava was born in Fargo, ND. He received the B.S. and M.S. in electrical engineering from the University of California, Los Angeles, and Purdue University, West Lafayette, in 1989 and 1990, respectively. In 1997, he received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA.

From 1989 to 1992, he was a Member of the Technical Staff at Bellcore, Red Bank, NJ, working on protocols for network maintenance. Currently, he develops algorithms for radar imaging at Lincoln Laboratory, Massachusetts Institute of Technology, Lexington.
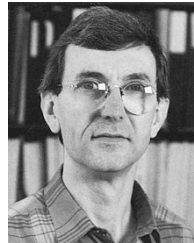
José M. F. Moura (S'71–M'75–SM'90–F'94) received the engenheiro electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and the D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

He is presently a Professor of Electrical and Computer Engineering at Carnegie Mellon University (CMU), Pittsburgh, PA, which he joined in 1986. Prior to this, he was on the faculty of IST, where he was an Assistant Professor (1975), Professor Agregado (1978), and Professor Catedrático (1979). He has held visiting appointments at several Institutions, including MIT (Genrad Associate Professor of Electrical Engineering and Computer Science from 1984 to 1986) and the University of Southern California (Research Scholar, Department of Aerospace Engineering, Summers 1978 to 1981). His research interests include statistical signal processing and telecommunications, wavelets and time-frequency transforms, image processing, video representation, and video editing and manipulation. He has over 200 published technical contributions, is the co-editor of two books, holds three patents in the areas of image and video processing and digital communications with the U.S. Office of Patents and Trade, and has given numerous invited seminars at U.S. and European Universities and Laboratories.

Dr. Moura has been the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a member of the IEEE Signal Processing Society Publications Board since October 1995. Since January 1999, he has been a member-at-large of the Board of governors of the IEEE Signal Processing Society as well as a member of the Editorial Board of the PROCEEDINGS OF THE IEEE. He is a Member of the *Multimedia Signal Processing Technical Committee* and a Member of the *Statistical Array and Multichannel Processing Technical Committee* of the Signal Processing Society. He was a Member of the *IEEE Press Board* from 1991 to 1995, a Technical Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1993 to 95, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1988 to 1992. He was a Program Committee Member for the IEEE International Conference on Image Processing (ICIP'95) and for the IEEE International Symposium on Information Theory (ISIT'93). He has organized and codirected two international scientific meetings on signal processing theory and applications and has been a member of the organizing committee for several other signal processing international technical meetings. He is a Corresponding Member of the *Academy of Sciences of Portugal* (Section of Sciences). He is affiliated with several IEEE societies, Sigma Xi, AMS, IMS, and SIAM.