

Improved Manifold Coordinate Representations of Large-Scale Hyperspectral Scenes

Charles M. Bachmann, *Senior Member, IEEE*, Thomas L. Ainsworth, *Senior Member, IEEE*, and Robert A. Fusina, *Member, IEEE*

Abstract—In recent publications, we have presented a data-driven approach to representing the nonlinear structure of hyperspectral imagery using manifold coordinates. The approach relies on graph methods to derive geodesic distances on the high-dimensional hyperspectral data manifold. From these distances, a set of intrinsic manifold coordinates that parameterizes the data manifold is derived. Scaling the solution relied on divide-conquer-and-merge strategies for the manifold coordinates because of the computational and memory scaling of the geodesic coordinate calculations. In this paper, we improve the scaling performance of isometric mapping (ISOMAP) and achieve full-scene global manifold coordinates while removing artifacts generated by the original methods. The CPU time of the enhanced ISOMAP approach scales as $O(N \log^2(N))$, where N is the number of samples, while the memory requirement is bounded by $O(N \log(N))$. Full hyperspectral scenes of $O(10^6)$ samples or greater are obtained via a reconstruction algorithm, which allows insertion of large numbers of samples into a representative “backbone” manifold obtained for a smaller but representative set of $O(10^5)$ samples. We provide a classification example using a coastal hyperspectral scene to illustrate the approach.

Index Terms—Automatic classification, hyperspectral imagery, isometric mapping (ISOMAP), Jeffries-Matsushita distance, manifold coordinates, manifold geodesics, manifold learning, multi-dimensional scaling, nonlinear dimensionality reduction, tree searching, trees (graphs), Vantage Point Forest, vantage point tree, Virginia Coast Reserve.

I. INTRODUCTION AND BACKGROUND

A. Nonlinearity in Hyperspectral Imagery (HSI)

Nonlinearity in HSI is a significant source of estimation errors in derived products. Sources of nonlinearity include 1) nonlinear variations in reflectance produced by variations in sun-canopy-sensor geometry in the landscape [32], [49], 2) multipath scatter among subpixel constituents [44], [46], violating the traditional linear mixing assumptions, and 3) the variable presence of water, an attenuating medium [42] in the scene. Some of the errors that we observed in land-cover classification products [6], [7] became the motivation for finding new methods of modeling nonlinear structure in hyperspectral data [8]. In the next two subsections, we give a brief overview of the approach that we presented in [8] as a preamble to introducing improvements.

Manuscript received September 30, 2005; revised July 19, 2006. This work was supported by the Office of Naval Research under Program 0601153N, Task Area BE032-08-41.

The authors are with the Remote Sensing Division, Naval Research Laboratory, Washington, DC 20375 USA (e-mail: bachmann@nrl.navy.mil).

Digital Object Identifier 10.1109/TGRS.2006.881801

The introduction of manifold coordinate estimation algorithms began in applications outside of the field of remote sensing [47], [53], and since that time, a considerable number of papers have begun to appear in these other application areas that include, for example, face recognition, optical character recognition, phoneme recognition, and computational chemistry [1], [15], [34]. Before proceeding, we note that the use of methods for the estimation of manifold coordinates is relatively new both in remote sensing [2]–[4] and, in particular, in hyperspectral remote sensing [8]–[12], [21], [29], [35].

B. Manifold Coordinate Representations

In [8], we described a new method for modeling the nonlinear effects in HSI and demonstrated that it provided a better means of discriminating land-cover types with a high degree of spectral similarity. Using examples from AVIRIS and PROBE2 imagery, we also showed that our new approach provides better compression of HSI data in both terrestrial and aquatic imagery. This method involves a data-driven estimation of a set of coordinates that parameterizes the high-dimensional hyperspectral data, directly modeling its nonlinear structure with a set of intrinsic coordinates. The method proceeds by calculating the local spectral neighborhood distances where linearity is assumed to hold about each sample and then determining the shortest nonlinear path (geodesic) distances to all other spectral samples outside the spectral neighborhood. These geodesic distances are then used to derive the manifold coordinate system that parameterizes the high-dimensional hyperspectral data using standard multidimensional scaling methods. In [8], we also described methods for achieving computationally scalable implementations of this approach. Fig. 1 provides a conceptual representation of manifold coordinate estimation. Note that the manifold coordinate system parameterizes the higher dimensional (124 channels in this example) HSI data so that the linear distance in the manifold coordinates corresponds to a nonlinear distance over the surface of the original higher dimensional data. In Section III, examples of this processing applied to PROBE2 HSI from our Virginia Coast Reserve barrier islands study site [5]–[8], [45], [54] are provided.

In [8], we proposed several strategies for computing the global manifold coordinates. All of these strategies involved partitioning the scene into subsets, solving for the optimal manifold coordinates on these subsets, and either aligning or merging subset coordinates to obtain a global manifold coordinate system. Manifold coordinates on subsets were determined by using the optimal isometric mapping (ISOMAP) [53]

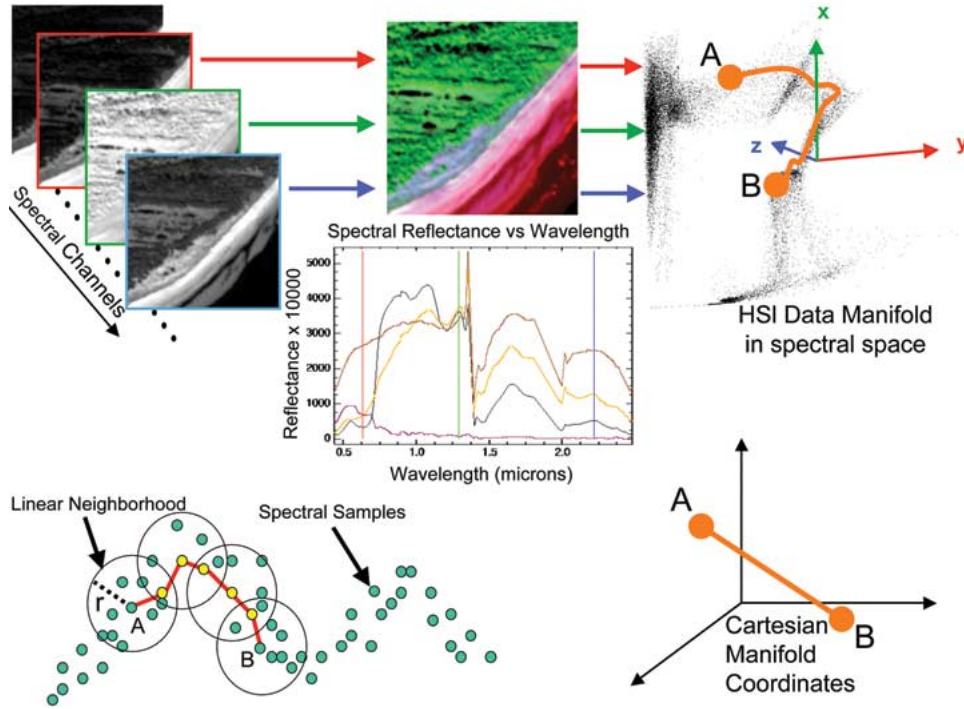


Fig. 1. Conceptual view of the manifold coordinate system. (Left) PROBE2 image subset showing the source data of 124 spectral channels. (Center) RGB triplet from wavelengths at 0.63, 1.29, and 2.22 μm for the subset over Smith Island, VA, October 18, 2001; shown: uplands, brackish and fresh water marshes, dunes, beach, and surf zone. (Right, top) Corresponding scatterplot of the reflectance of these arbitrarily chosen channels reveals a highly nonlinear HSI data manifold. (Right, bottom) Manifold coordinate system parameterizes the HSI spectral data (note that the manifold coordinates are a parameterization of the full spectral data, not just the three arbitrary channels displayed). (Bottom, left) Geodesic distances estimated via the shortest path between neighbors common to overlapping neighborhoods.

TABLE I
ISOMAP PROCESSING STEPS AND SCALING WITH NUMBER OF SAMPLES N

	Process	Description	Computational Scaling	Memory Scaling	Comments
1	Compute metric distances	precursor to neighborhood calculation	$O(N^2)$	$O(N^2)$	all sample pairs
2	Find Neighbors, each sample	sparse initial graph	$O(N^2 \log(N))$	$O(N^2)$	exhaustive search
3	Estimate geodesic shortest paths		$O(N^2 \log(N))$	$O(N^2)$	all pairs, Dijkstra shortest paths algorithm
4	Calculate variance landmarks geodesic distances		$O(N)$	$O(N)$	preparatory to eigendecomposition
5	Eigendecomposition, manifold coordinates	Iterative eigensolver	$O(N^2)$	$O(N^2)$	potential convergence problems

algorithm. In its original form (Table I), the ISOMAP portion of the computations involves the following steps: 1) For a specific metric such as Euclidean distance, spectral angle, or some other appropriate choice, determine the local neighborhoods (initial sparse neighborhood graph d_G) of the input data space where linearity holds, maintaining a list for each sample of its neighbors and metric distances. 2) At each sample, for all distances outside the neighborhood, estimate the geodesic distances using Floyd’s algorithm [37], a shortest path algorithm that exhaustively relaxes all edges of the graph d_G according to

$$d_G(i, j) = \min_k (d_G(i, j), d_G(i, k) + d_G(k, j)). \quad (1)$$

Floyd’s algorithm scales as $O(N^3)$; however, the scaling was later improved [14] by replacing Floyd’s algorithm with Dijkstra’s all shortest paths algorithm [24], [50], which scales

as $O(N^2 \log(N))$. Dijkstra’s algorithm uses the same “relaxation” rule for edges, but the choice of which edges are relaxed at each iteration is controlled by a minimum priority queue [22], [50]. Specifically, at each iteration, the closest edge not already attached to the graph d_G is extracted from the minimum priority queue, and all edges leading from the newly attached edge are relaxed according to (1) to compute the shortest nonlinear path (geodesic) distance to all other samples (note that this is a graph calculation and that the metric, therefore, is only evaluated in step 1) inside the neighborhood). 3) With the full $N \times N$ (where N is the number of spectral samples) geodesic distance matrix calculated in steps 1) and 2), compute the second-order variation in the geodesic distances

$$\tau = -\frac{1}{2} H^T S H \quad (2)$$

where

$$S_{ij} = ((d_G)_{ij})^2 \quad (3)$$

$$H_{ij} = \delta_{ij} - \frac{1}{N} \quad (4)$$

is a centering matrix. 4) Extract s ($s \ll N$) manifold coordinates from the most significant eigenvectors and eigenvalues of the $N \times N$ matrix τ with the i th manifold coordinate vector \vec{m}_i given by

$$\vec{m}_i = \sqrt{\lambda_i} \vec{v}_i. \quad (5)$$

Note that the original ISOMAP algorithm did not suggest a self-consistent means of dealing with isolated points at the conclusion of the geodesic distance calculation. In [8], we presented one means of doing this; however, the method that we present in Section II-E is a simpler, although equivalent, method.

C. Scalable Algorithms

In [8], we addressed the computational and memory scaling issues associated with manifold coordinate calculations at remote-sensing scales where $O(10^6) - O(10^7)$ pixels in a single hyperspectral scene are not uncommon. One of the principal limiting factors in the original ISOMAP algorithm was memory, which scaled as $O(N^2)$ because of the need to store d_G . Computationally, Dijkstra's algorithm with a minimum priority queue implementation [50] ensures that the graph calculation scales as $O(N^2 \log(N))$. Because of the computational and memory requirements, we developed a scaling strategy [8] in which large hyperspectral scenes are divided into a computationally tractable set of subsets or "tiles" for which manifold coordinates can be optimally computed followed by an alignment phase during which the embedded manifold coordinates for each tile subset are aligned to a common manifold coordinate system.

In [8], several strategies for alignment were proposed. These included 1) splicing a set of common samples onto each tile that could serve as guide posts for manifold alignment; 2) partitioning the scene into tiles by random or active sampling, and computing optimal manifold coordinates for each tile followed by an alignment stage; and 3) a direct reconstruction technique in which, following partitioning and optimization of manifold coordinates for each tile, a subset of full spectral samples from one tile (derived from the original scene or a decimated subset) was reconstructed in the spectral space of another tile. For 3), using the locally linear property of the manifold, these same weights were applied to transform the corresponding manifold coordinates of the samples. By reconstructing enough samples where there is sufficient data in each tile for accurate reconstruction, a linear coordinate transformation can be derived. We used the optimal least-squares coordinate transformation for which the matrix solution is the pseudoinverse [26], [30], i.e.,

$$P = (M_i^T M_i)^{-1} M_i^T M_j^* \quad (6)$$

where M_i is the matrix of the manifold coordinate sample vectors (columns defined by (5) from tile T_i), and M_j^* is the

corresponding set of coordinates reconstructed in the manifold coordinate system of tile T_j . When no sufficiently accurate reconstruction was possible to a prechosen target tile, a series of alignment hops was used between intermediate tiles possessing common features of source and target tiles. Another alternative strategy that we also proposed in [8] included the use of a "backbone" in which random samples were drawn from within the tiles comprising the scene, and then the manifold coordinates for each tile were aligned to the backbone manifold by a change of coordinates of the form in (6). Note that in one way or another, all of the methods that we previously suggested in [8] ultimately built up enough mapping pairs between coordinate systems of subsets so that a suitable pseudoinverse transformation of the form in (6) could be defined.

In [8], the reconstruction method was determined to be the most effective of the manifold alignment strategies. The other methods we proposed, such as the variation involving a backbone manifold described above, were considered to be less useful primarily because of sampling limitations that resulted from restrictions on tile size imposed by memory and computational scaling of the original ISOMAP algorithm. In all of the methods we proposed earlier, the alignment errors that appeared were the result of incomplete constraints on the manifold coordinate transformation between tiles. The latter stemmed from the limited number of samples available in each tile. Additionally, with the earlier approach, there is always the risk that lower order manifold coordinates might not be adequately sampled in the set of points entering the pseudoinverse mapping, especially when these lower order dimensions are primarily spanned by more sparsely populated categories. Likewise, variations in the eigenspectra between small subsets can lead to a torquing of the final manifold coordinates of subsets, no matter what method of subset partitioning is used. For contiguous tiles, the resulting artifacts will be seams, while for random sampling, the artifacts will appear as graininess.

In the sections that follow, however, we incorporate several new improvements to ISOMAP that allow us to work with significantly larger subsets. One of these improvements allows the $N \times N$ matrix d_G to be replaced by a significantly smaller but representative matrix that mitigates the memory burden and some of the computational burden. In addition to lower memory requirements, the method described in the next section also streamlines other computational issues such as the eigensolution of τ , eliminating iterative eigensolvers in favor of more reliable exact solvers appropriate for smaller matrices [31], and also requires fewer geodesic distance calculations. The $O(N^2 \log(N))$ Dijkstra calculation is now replaced with an $O(LN \log(N))$ calculation with $L \ll N$. We also overcome another significant bottleneck that arises both in the calculation of manifold coordinates using ISOMAP and in the reconstruction step, namely the calculation of neighborhood lists and distances. The naive approach to calculating neighborhoods, which is usually employed, scales as $O(N^2 \log(N))$. In Section II-B2, we describe a way to streamline the calculation of the neighborhoods to improve the scaling to $O(N \log^2(N))$, which is nearly linear for the range of N s that we consider. With the new methods described in the next sections, all of the methods previously described in [8] will yield better overall

TABLE II
 ENH-ISOMAP PROCESSING STEPS AND SCALING WITH NUMBER OF SAMPLES N

	Process	Description	Computational Scaling	Memory Scaling	Comments	Citation
1	Construct VPF's	one or more radii r	$O(N \log(N))$	$O(N \log(N))$	prepare for neighborhood search	this paper
2	Find Neighbors, each sample	sparse initial graph	$O(N \log^2(N))$	$O(f(r)N)$, $f(r) \ll N$	query VPF's	with VPF's, this paper
3	Patch isolated points to graph	FastPatch	$O(sN \log(N))$, $s \ll N$	$O(sN)$ $s \ll N$	repeated Dijkstra	this paper
4	Select landmarks			Skeleton: $O(LN \log(N))$ Random: $O(L)$		Skeleton: this paper Random: [23]
5	Calculate geodesic distances to landmarks	LxN matrix	$O(LN \log(N))$, $L \ll N$	$O(LN)$	Dijkstra algorithm	[24],[50]
6	Calculate variance landmarks geodesic distances		$O(L)$	$O(LN)$	input to eigendecomposition	
7	Eigendecomposition, manifold coordinates	exact eigensolver	$O(L^3 + LN)$	$O(L^2)$		[31]

products because we can process larger more representative subsets before a coordinate alignment or reconstruction must occur. These steps will help mitigate alignment errors that appeared originally in [8].

In this paper, we propose a modified version of the backbone method that eliminates the artifacts that appeared in our previous approach [8]. The modified version will take advantage of several specific improvements from other researchers as well as our own improvements to ISOMAP. The new approach is applicable to significantly larger data sets and, in particular, allows a vastly larger backbone manifold to be constructed. Likewise, the new approach eliminates the pseudoinverse in favor of direct insertion of all points into the final manifold via reconstruction. The reconstruction step takes advantage of a new fast neighborhood calculation algorithm that we develop in this paper. We also incorporate the fast neighborhood calculation into our enhanced version of the ISOMAP algorithm, which is used to compute the backbone manifold coordinates. With local reconstruction of samples in the backbone manifold coordinate system, we eliminate torquing artifacts and provide a more stable self-consistent description because only one eigenspectrum is computed. The result is that if errors do occur, they will be localized. The other consequence is that as more samples are added to the backbone, the quality of the manifold coordinates improves.

The remainder of this paper is structured as follows: In Section II, we first introduce the enhanced ISOMAP (ENH-ISOMAP) algorithm that will be the cornerstone of the overall improved method that we present in this paper. Our ENH-ISOMAP algorithm incorporates several new features such as 1) a very fast neighborhood construction, using a new algorithm, Vantage Point Forests, and which scales nearly linearly with N and is based on Vantage Point Trees (VP-Trees) [55]; 2) a new patching algorithm to ensure all points separated from the main distribution by more than a neighborhood radius are included in a self-consistent manner in the geodesic distance calculation; 3) an approximation [23] that streamlines the geodesic distance calculation, improving scaling of both computational and memory requirements for geodesic distance calculations; and 4) a more flexible neighborhood definition than that previously used in the original ISOMAP algorithm. The flexible neighborhood is significant because it better accommodates differences in the scale of intersample spacing that naturally arise, for example, when land and water regions

appear in the same scene, as they do in coastal environments. In Section II-G, we discuss the effect of noisy data on manifold coordinate optimization and propose at least one solution. In Section II-H, we develop a new solution to estimating the global manifold for large hyperspectral scenes. This solution incorporates ENH-ISOMAP and a reconstruction principle made faster also by using Vantage Point Forests. We call this new approach the ‘‘Modified Backbone’’ method because it is an improved version of one of the manifold coordinate approaches that we initially developed in [8]. In Section III, we present results of the Modified Backbone approach, showing a seamless set of manifold coordinates for a coastal hyperspectral scene with nearly two million pixels. In Section III-A, we show quantitatively enhanced land-cover category separability in a coastal HSI scene, contrasting the traditional minimum noise fraction (MNF) preprocessing used in the PPI end-member selection algorithm with manifold coordinates derived from the MNF representation. Finally, in Section IV, we provide a summary and draw conclusions.

II. ENH-ISOMAP

A. Overview of ENH-ISOMAP

Even with the use of Dijkstra’s shortest path algorithm, the $O(N^2)$ memory and $O(N^2 \log(N))$ computational demands that it imposes still limit the size of scene subsets that can be effectively processed. In addition, as the size of the data set increases, the computational demands of calculating neighborhoods quickly become a principle bottleneck since the scaling of this critical step is $O(N^2 \log(N))$. To improve solutions for the global manifold coordinates over those methods that we originally developed in [8], we will need to increase the size of data subsets for which we can achieve an optimal or near-optimal manifold coordinate solution. A cornerstone of this is the improvement of the original ISOMAP algorithm.

To achieve this, we have combined several new algorithms in a new computational framework (Table II). The approach incorporates a new Vantage Point Forest algorithm for improving the speed of the neighborhood calculation, a more flexible neighborhood definition, an improved patching algorithm to ensure that points separated from the main distribution by more than a neighborhood are incorporated self-consistently in the graph calculation, and finally the use of Landmarks ISOMAP [23], an approximation to improve the computational

scaling and memory requirements of the geodesic graph calculation. We term the amalgamation of these improvements ENH-ISOMAP. Another refinement in ENH-ISOMAP is a more judicious landmark selection algorithm that ensures that the geometric structure of the manifold will be more completely represented than would be the case with random sampling. Table II details the fundamental processing steps that we have defined. Comparing it to the original ISOMAP implementation (Table I) with the underlying Dijkstra shortest path calculation (Fig. 1), we can see that the overall flow of processing has been modified significantly, resulting in dramatically better scaling. In the following subsections, we detail the processing in each of the steps in Table II.

B. Improved Scaling of Manifold Neighborhood Construction With Vantage Point Forests

1) *Neighborhood Scaling Problem:* Although the use of Landmarks processing described in Section II-C solves two important bottlenecks in the required memory and the calculation of geodesic distances, it does not solve another fundamental scaling problem that arises as the tile or subset size increases. This emergent scaling issue is the computational time needed to calculate the neighborhood lists and neighbor distances of each point in the manifold. This is a critical step in the initialization of the sparse graph used in calculating the manifold geodesic distances. To date, the algorithms that have been presented surrounding ISOMAP and its variants have not explicitly addressed this key issue. For each sample, the exhaustive approach employed in past implementations of ISOMAP requires calculation of all metric distances to all other samples and the determination of which ones lie within the required neighborhood distance or K-limit depending on which form of ISOMAP is used. The exhaustive approach, therefore, scales as $O(N^2 \log(N))$, which quickly becomes impractical for the scale of typical remote-sensing data sets that contain $O(10^6) - O(10^7)$ pixels or more in a scene. Therefore, once landmarks are employed as per the above discussion, the use of the exhaustive approach to establish the neighborhoods quickly becomes the dominant computational burden even at $O(10^5)$ pixels. Another important point to emphasize here is that for two of the divide-conquer-and-merge strategies [8] for computing global manifolds, neighbors of samples from one subset must be found in another subset in the full manifold coordinate space. These were required to produce a coordinate transformation that applies equally in the full spectral space as well as the manifold coordinate space to be aligned. While this step is not $O(N^2 \log(N))$, it is nevertheless quite time consuming if distances must be calculated to all possible samples in the other subset, as they would be in the exhaustive approach. Vantage Point Forests, therefore, also alleviate this burden downstream when optimized manifold coordinate subsets are to be merged or samples are to be inserted in a representative “backbone” manifold coordinate representation, as we do in the present work (Section II-H).

2) *Fast Neighborhood Calculation Using Vantage Point Forests:* Here, we develop a solution to the problem of neighborhood calculation that does not require us to examine all

possible distances from each sample to all others. The latter exhaustive approach scales as $O(N^2 \log(N))$, whereas we seek an algorithm that scales as close as possible to $O(N)$. Note that in what follows, we develop an approach to finding exact neighbors, not approximate neighbors. The importance of this lies in the fact that if approximate geodesic distances are used, these inaccuracies may distort classification results when classes of interest are only subtly different.

The typical extant approaches to fast neighbor search algorithms in the literature proceed by partitioning the input data space but, to our knowledge, have not been previously applied to the construction of manifold neighborhoods. These fast search algorithms produce a tree-like structure with successive partitioning proceeding iteratively until all points are exhausted. The beauty of such an approach is that, at each sample, it eliminates the necessity of examining all other samples to determine the neighborhood. In particular, it eventually focuses the search on a narrow subset of end “leaves” in the tree where possible neighbors could exist. One other limitation of previous approaches using VP-Trees is that at the top level of the tree search, where the first set of comparisons are made, the search does not eliminate as many possibilities as it might. The solution that we develop overcomes this limitation by constructing a Vantage Point Forest, which tends to focus the search in a narrower portion of the high-dimensional space at the beginning of the search process.

A significant amount of research has been published on the subject of calculating neighborhoods [55] and approximate neighborhoods, although not per se in the context of manifold algorithms. Extraction of appropriate samples of the full data set is critical to all approximate methods. Many discuss the “curse of dimensionality” to highlight the poor scaling with D , which is the dimensionality of the data. Generally, the sampling should represent/cover the entire data set at sufficient resolution to allow accurate construction of the manifold coordinates. Most nearest-neighbors theorems and computational experiments were originally based on uniformly or Gaussian-distributed data in R^D . Several papers compare various methods and make recommendations based on internal data structure, tradeoffs between the costs of searching, and generating and storing tree structures [38], [41]. Specific tree structures have been proposed, e.g., K-Dimensional Tree (KD-Tree) [28], Sphere/Rectangle Tree (SR-Tree) [40], VP-Tree [55], Principal Axis Tree (PA-Tree) [41], and improved upon [19]. General schemes to efficiently search tree structures in logarithmic time have also been proposed [36], [38]. In [41] and [55], it is observed that the VP-Tree is more adaptable than the KD-Tree when the data distribution is unknown in advance. The PA-Tree proposed in [41] optimizes the data partitioning by recursively partitioning the data along the (local) principal axis. The multi-VP-Tree [19] was developed specifically to address high-dimensional data spaces ($D > 20$). This approach is related to our m -way VP-Tree structure (Vantage Point Forest). In contrast, the original VP-Tree recursively splits the pixels into two equal-numbered groups producing a “binary” VP-Tree. Based on the above considerations, we chose the VP-Tree structure as the basis for developing an algorithm that would be suitable for hyperspectral applications since it

is among the set of better extant algorithms. The absolute best choice of algorithm appears to be data dependent [38].

Our use of a forest of VP-Trees helps with two aspects of nearest-neighbors searches in high-dimensional ($D > 100$) hyperspectral data sets. First, the general VP-Tree depends upon the choice of vantage points, a.k.a. pivots. Good pivots can be hard to find, and at least one qualitative analysis suggests that the pivots should be well spaced [20]. Various authors employ random pivot selection, at least as a test case. However, hyperspectral data of large littoral scenes display a wide range of pixel density in spectral space. For instance, land areas tend to tight clumps or bands, while water is generally more diffuse. At the same time, in the transition zone (tidal zone or transition into the marsh in lagoonal areas), the pixel density may be sparse. Selecting a single random set of pivots across the scene would concentrate too many on the land and essentially none in the land–water transition region. In a Vantage Point Forest, trees are well spaced. Each tree is centered on a data pixel and covers a local region of the spectral space: trees do not overlap, and enough are generated to cover the data set. No (computational) time or trees are wasted in empty parts of the spectral space. The forest adapts to the data and provides an efficient nearest-neighbors search structure.

The second advantage is harder to quantify and, while present, may not always provide much savings in computational time. The complexity of nearest-neighbors searches depends strongly (exponentially) on the dimensionality of the data. The various tree structures aim to reduce this dependency; however, it remains at some level. Each individual tree in the forest sees only the local dimensionality that can be significantly less than D , i.e., the full dimensionality. If the local dimensionality varies across the data set, then each tree in the forest-of-trees approach will adapt to the local dimensionality and thereby speed-up the nearest-neighbors calculations as appropriate. In contrast, traversing branches of a single tree will proceed at various rates that depend on the branch’s local dimensionality. This may offset some of the forest’s advantage. However, one poorly placed vantage point, or branch, in a single tree, i.e., in the middle of a high-dimensional pocket of data, could significantly slow the entire nearest-neighbors search. In contrast, a poorly placed tree in a Vantage Point Forest only affects that tree and its immediate neighbors, not the entire nearest-neighbors search. The Forest approach appears to minimize the computational time of the worst case, rather than an average runtime, although the scaling of the Vantage Point Forest is not appreciably different from the scaling laws obtained for those that minimize the average runtime but ignore worst-case scenarios.

We make one additional note concerning PA-Trees [41]. The PA-Tree is probably preferred whenever the actual data dimensionality is much less than the spectral dimensionality and the principal axis analysis is meaningful. In those cases, PA-Trees should also provide a good covering of the data set. If, however, the local dimensionality is low, but the data are embedded nontrivially (nonlinearly) in the high-dimensional spectral space, then the PA-Tree approach loses effectiveness. A forest of PA-Trees may recover some of the performance lost due to nonlinearly embedded data. We have not yet experimented with a forest of PA-Trees on hyperspectral

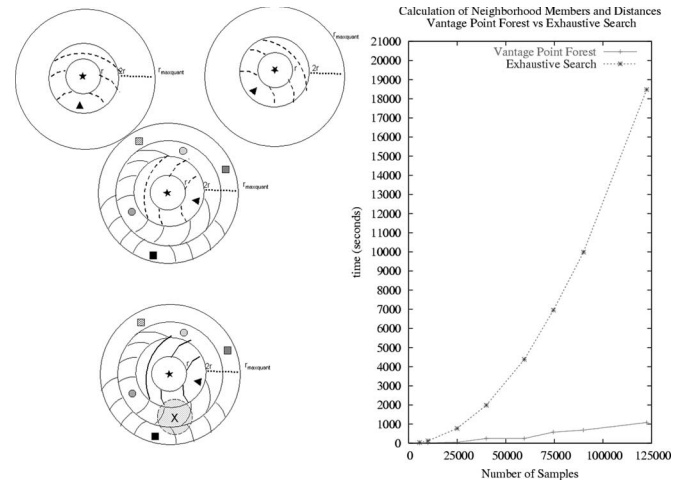


Fig. 2. (Top left) Schematic diagram of Vantage Point Forest concept showing top-level vantage points denoted by stars and associated hyperspherical shells and subvantage points (squares, circles, and triangles) at the next level in the forest; for some of these, we indicate the associated rings constrained to the shell. (Bottom, left) One of the VP-Trees in the forest overlapped by a query sample. (Right) Scaling of Vantage Point Forests algorithm for calculating neighborhoods versus sample size compared to exhaustive search.

data, but we can expect that the large N scaling will remain $O(N \log(N))$.

Below, we develop a new process for determining the neighborhoods that scales, at worst, as $O(N \log^2(N))$. Note that over the range of N that we will consider here, this is nearly linear and significantly better than the exhaustive neighborhood search that occurs in the original ISOMAP algorithm and scales as $O(N^2 \log(N))$.

Construction of the Vantage Point Forest (Fig. 2) consists of a set of iterative steps. In each iteration, the first step is to pick the top-level vantage points that form the tops of the trees in the Vantage Point Forest. We have implemented the selection of tree tops in more than one way. Obviously, this can be done at random, but a more appropriate way from the standpoint of covering the varying structure and local dimensionality of the data may be to “skeletonize” the data. We describe the process of “skeletonization” at greater length in Section II-D.

The iterative construction process proceeds as follows: Assume that r is the characteristic radius within which the manifold is locally linear, and pick a maximum quantization radius $r_{\max\text{quant}}$ for each tree. A value of two to five times r has worked well in the hyperspectral data sets that we have considered thus far. Also note that, ultimately, $r_{\max\text{quant}}$ determines the maximum number of rings present in each tree. The next step is to select the first top-level (forest-level) vantage point (the first tree), compute all distances to all other samples (note that the only time we make this kind of $O(N)$ exhaustive calculation is for a small set of samples that form the top-level vantage points of each Tree in the Forest), and then, using a heap, sort them in increasing distance out to $r_{\max\text{quant}}$. For v VP-Trees, this latter step computes in $O(N \log(N))$ time, so it is a very fast calculation. Note that v is a function of the underlying neighborhood radius assumed and the maximum quantization radius $v = v(r, r_{\max\text{quant}})$. Those samples within $r_{\max\text{quant}}$ are removed from the available sample list for all other trees that are subsequently generated at this level of the

Forest to ensure an exclusive quantization with each sample belonging to only one tree. The sorted lists of distances from each top-level vantage point are divided into a set of hyperspherical shells or “rings” of radius $r, 2r, \dots, r_{\max\text{quant}}$. Within each tree and within each of these shells, the same set of steps is repeated with subvantage points (branches) being selected at random; however, this time, the distances must only be calculated to the samples available at this sublevel of the tree, i.e., to those samples in the shell; as before, these are sorted out to $r_{\max\text{quant}}$. The process proceeds in this manner, iteratively stopping when all samples are within r of the parent tree or branch.

Fig. 2 shows a schematic view of the top level of a portion of a Vantage Point Forest. In this example, there are three top-level trees in the Forest. The centers of these trees are the vantage points indicated by stars. The center neighborhood of the vantage point and a set of “rings” (hyperspherical shells) of thickness r are indicated at radii $r, 2r, \dots, r_{\max\text{quant}}$; recall from the above description that sorted lists of these distances at each level of each tree in the Forest are maintained. For two of the trees in the example, the data occur in only the first two rings, while for the third, the data appear in all of the rings. Subvantage points in the rings are indicated symbolically by triangles in the first rings, by circles in the second ring, and by squares in the third rings. These subvantage points also quantize their respective shells from $r, 2r, \dots, r_{\max\text{quant}}$ or less if the size of the ring is smaller. We have indicated only some of these within-shell quantizations at the next level down to avoid an excessively crowded figure. At each level, quantization within a shell by a subvantage point removes its associated data from the list of available samples within the shell so that when additional vantage points are added to the shell, they must only quantize and sort the residual points in the shell. As indicated above, this process iterates so that each shell is eventually divided into subshells until all data resides within r of a subvantage point. In an average sense, we can expect that there will be $O(\log(N))$ levels in a typical tree in the Forest.

Once a Vantage Point Forest is constructed, the query to find the neighbors of a sample vector first identifies which hyperspherical shells of the top-level trees in the Vantage Point Forest could be overlapped by a hypersphere of radius r about the query sample (Fig. 2). For example, Fig. 2 shows an enlargement of one of the top-level trees and its associated rings. For this tree, a query denoted by an X is shown along with its associated hyperspherical neighborhood radius. From the figure, we can see that the neighborhood of X overlaps three of the first-level rings of a tree. Within these shells, the query then iteratively determines which overlapped shells are within less than r of their corresponding subvantage point. For those that are, the query determines which subset is actually within r of the query sample using the distance metric. Otherwise, for shells that are overlapped and are further than r away from the corresponding subvantage point, the query determines which sub-subvantage points are overlapped at the next sublevel down. Thus, gathering of the samples in the neighborhood of the query sample occurs at each ending branch, wherein only samples within r of the subvantage point reside and are potentially overlapped with a small number of comparisons to actually be made between the samples in these shells and the

query sample. In Fig. 2, at the next level down, note that there are eight subvantage point overlapped regions. Of these, only one contains points all within r of their respective subvantage points, so seven of these regions will have to be explored recursively. In summary, all overlapped rings must be explored recursively through the forest until all overlapped regions are within r of their respective subvantage points.

Fig. 2 shows the typical scaling of this Vantage Point Forest algorithm as the number of samples N in the hyperspectral subset is varied. For comparison, Fig. 2 also shows the $O(N^2 \log(N))$ behavior of the previously used exhaustive search algorithm that compares each query sample with all other samples to determine which are within the neighborhood of the query sample. Note that r and $r_{\max\text{quant}}$ were held fixed for all points on the curve in Fig. 2. Thus, for the Vantage Point Forests, the number of top-level trees varied with the particular random samples used to choose source nodes for each tree in the forest and with the density of points that changed as the data set size grew. The variation of the number of trees in the forest was not over a large scale. Fig. 2 shows that the Vantage Point Forest scales nearly linearly for most of its range. Curves are derived from subsets of a PROBE2 HSI scene of Smith Island, VA, on October 18, 2001, using 114 of 124 bands in the analysis with $r = 0.02$ and $r_{\max\text{quant}} = 0.08$. Timing results were obtained on an Athlon64 3000 + 2.0-GHz processor.

The following scaling argument suggests why the Vantage Point Forest scales nearly linearly. We choose $r_{\max\text{quant}}$ relative to a given r such that $v \ll N$, but not so small that $v \sim 1$. If necessary, we can also control v to make sure that it is independent of N by adjusting $r_{\max\text{quant}}$. The algorithm is constructive at each level, so for instance, for the first tree at the top level of the forest, we can expect to do the sorting in $O(N \log(N))$ time. For the next top-level tree, we are left with, on average, $O(N - N/v)$ points, so the sorting takes $O((N - N/v) \log(N - N/v))$ time. Likewise, for the third tree at the top level of the forest, the time is $O((N - 2(N/v)) \log(N - 2(N/v)))$, and so forth. Therefore, taken together, the processing time for the v trees at the top level of the forest is, on average, $O(N \log(N))$ time since $N/v \ll N$. At the next level down, i.e., for each of the trees in each ring (shell), we have only N/v^2 samples in each ring on average. Across the forest, we assume that each of the v vantage points at this level across all of the trees will each have in turn v sub-VP-Trees, although in practice the further down we go in the Forest levels, the number of subvantage points will eventually die out as we reach the bottom of the Forest and exhaust all points. Nevertheless, in an average sense, at the second level of the forest, we can expect $O(v^2)$ trees; therefore, repeating the same argument used at the first level, we would expect to be able to sort all of these second-level trees in the rings in $O(v^2(N/v^2) \log(N/v^2)) \sim O(N \log(N))$. At the next level down, repeating the same argument, we find $O(v^3)$ trees each with $O(N/v^3)$ samples in each, so we can expect to sort at this third level of the forest in $O(v^3(N/v^3) \log(N/v^3)) \sim O(N \log(N))$ time, and so forth. As noted earlier, we can expect that the number of levels in the Forest will be $n_{\text{levels}} \sim O(\log(N))$; therefore, the total computational time will be $O(N \log^2(N))$. For the range of samples that we will need to process, which will be of $O(10^5)$

in the “modified backbone” presented in Section II-H, we find $O(N \log^2(N)) \approx O(N)$, as shown in the curve in Fig. 2. We note that the nearest-neighbors search of the Vantage Point Forest is exact. The time saving arises from the tree structure that rapidly excludes distant pixels from further consideration.

C. Improved Scaling of Geodesic Distance Graph Calculations

An improvement to the processing speed and memory requirements associated with ISOMAP was described in [23]. The improved method chooses a set of “landmarks” (L-ISOMAP) from which all of the manifold geodesic distances d_G are calculated. This forms an $L \times N$ geodesic distance matrix with $L \ll N$. The symmetric submatrix d_L of distances between landmarks is an $L \times L$ matrix whose eigenvalues and eigenvectors form the basis of the embedding of the manifold coordinates. Note that so long as the sampled landmarks span the space of the embedded manifold coordinates, the landmark distances are sufficient to calculate the manifold coordinate system. Note also that the eigenvector and eigenvalue problem of a large $N \times N$ matrix has been replaced by a smaller $L \times L$ problem. As before, the second-order variation in d_L is computed according to

$$\tau_L = -\frac{1}{2} H^T S_L H \quad (7)$$

where $(S_L)_{ij} = ((d_L)_{ij})^2$. Once the most significant eigenvalues and eigenvectors of τ_L have been determined, the manifold coordinates of the remaining nonlandmark samples can be computed by a simple linear transformation since their distances to the landmark positions are all known, i.e.,

$$M_i(\vec{x}) = P_L * (\bar{\Delta}_i - \Delta_i) \quad (8)$$

where $M_i(\vec{x})$ is the embedded manifold coordinate of the spectral sample \vec{x} , P is a matrix whose i th row is

$$(P_L)_i = \frac{(\vec{v}_L)_i}{\sqrt{(\lambda_L)_i}} \quad (9)$$

where $(\vec{v}_L)_i$ and $(\lambda_L)_i$ are the i th eigenvector and eigenvalue of τ_L , $\bar{\Delta}_i = E_{L_j}(((d_L)_{L_i L_j})^2)$ is the mean squared distance from the i th landmark to all other landmarks, and $\Delta_{ij} = ((d_G)_{L_i j})^2$ is the squared distance from sample j to the i th landmark.

The Landmarks approach allows two major simplifications of ISOMAP: First, the $N \times N$ distance matrix is reduced to an $L \times N$ distance matrix with the number of landmarks $L \ll N$. Memory usage now scales as $O(LN)$ rather than $O(N^2)$, and CPU time scales as $O(LN \log(N))$ rather than $O(N^2 \log(N))$. Second, the iterative eigensolver is replaced by a more stable “exact” eigensolver [31]. CPU time to extract the manifold coordinates drops from $O(N^2)$ to $O(LN + L^3)$, where $L \ll N$. Therefore, considerable (CPU) time and effort can be put into

finding appropriate landmarks without significantly affecting the total CPU time or memory storage requirements. Several landmark selection algorithms were compared [20] and were implemented within our data structure.

D. New Landmark Selection Criteria: Skeletonization

The selection of landmarks becomes a critical problem in large remote-sensing scenes. In previous publications related to other applications [23], the selection of landmarks is done randomly. There is potentially a fundamental difficulty with this approach because it implies that high-density regions will be predominant when a small landmark set is chosen. It is, after all, the goal to use as few landmarks as possible from both computational and memory storage perspectives. In HSI, such as in the examples presented in this paper, there are large variations in the local density of points. This is particularly true for instances when terrestrial and water samples are compared. In some HSIs with which we have experimented, we have found local spectral densities that vary by as much as an order of magnitude in spectral angle. From first principles, the manifold coordinate system should cover the geometric form of the data rather than its local density. Otherwise, low-density dimensions may be missed, and these may contain important information for classification purposes.

One way to achieve this is to choose landmarks in such a way that they cover this geometric structure in a spatially uniform way. We call this approach “skeletonization” and implement it by constructing a list of all samples, then iteratively selecting a random point from the data set, finding all neighbors from within a prespecified radius, and removing these neighbors from the list. As before, note that we can achieve this quickly in $O(LN \log(N))$ time for the L landmarks by constructing and querying a Vantage Point Forest.

Another risk of using a random selection criterion is that sparse regions of the data space may be unsampled or undersampled and therefore not spanned in the final manifold coordinate system obtained. The process of skeletonization overcomes this by ensuring that even sparsely populated manifold coordinate dimensions will be spanned. Another advantage of the skeletonization method of selecting landmarks is that anomalous spectra will tend to be spread further away precisely because the full convoluted structure of the manifold will be found. Likewise, dissimilar spectra will be farther apart, and there is less risk of the so-called “short circuits” in which a portion of the manifold folds on itself due to inaccurate path estimation.

Note that there are some other potential ways to implement skeletonization. In particular, we have also used a density-based approach in which at each step we remove up to a maximum number of samples from the list at each iteration. By setting r sufficiently large in the Vantage Point Forest, we can ensure that each landmark will have constant density around it. Note that this uses a concept similar to what is found in conformal ISOMAP [23], where the distance metric is replaced by a conformal definition that assumes the distribution has constant density in the transformed data prior to further processing by ISOMAP. The uniform assumption may be too restrictive

in some applications, and it is unclear whether it is justified for HSI imagery. Nevertheless, it is a way to address data sets such as HSI in which there can be large variations in local spectral density.

E. FASTPATCH: New Method for Patching Isolated Samples in Geodesic Distance Calculations of ISOMAP and Variants

The original ISOMAP method had no self-consistent way of dealing with samples that remain disconnected from the geodesic distance graph after the Dijkstra algorithm has completed, except possibly to ignore these samples altogether. This is not an option for remote-sensing applications that demand that the whole scene be processed. We have developed a patching algorithm that ensures that all points will be self-consistently connected to the geodesic distance graph. The method consists of inserting a step prior to the general geodesic distance calculation. All samples are self-consistently attached to the first sample (the choice of source point is arbitrary). In this method, we recursively connect the closest disconnected point not on the graph via one or more links and then call the Dijkstra shortest paths algorithm. We repeat this process until all points are connected to the first sample via shortest paths. We call this patching algorithm FASTPATCH. Note that in the ENH-ISOMAP method description, we can choose the first landmark to be the source point. This procedure augments the sparse neighborhood graph d_G by a minimal set of new connections sufficient to interconnect the full set of N samples.

F. More Flexible Neighborhood Definition

In HSI, variations in density naturally arise, leading to larger intersample spacings for some categories than for others. This occurs in our coastal HSI, where, for instance, the spectral separation between pairs of water samples is, on average, larger than for many land categories. A logical objective, therefore, is to establish a more flexible definition of the neighborhood. One solution is a conformal version of ISOMAP described in [23]. On the other hand, a simple alternative consists of combining the absolute- r and k -neighbor limits first specified in [53]. A hybrid neighborhood definition (Table II) that we have used successfully to accommodate these variations in coastal HSI is defined as follows: Define a nominal search radius r , usually close to the radius of the smallest neighborhoods (i.e., categories for which spectral separation is small) to avoid losing detail. To accommodate data categories where there is high density, limit the number of neighbors to the k_{\max} closest. Likewise, for low-density areas, if necessary, expand the neighborhood radius definition to encompass at least k_{\min} neighbors. Note that to implement the latter, we can create several Vantage Point Forests at different radii, e.g., $r, 2r, 3r, \dots$, etc. If no matches are found within the nominal Vantage Point Forest at radius r , try the next Forest at radius $2r$, and so forth. If none of the Forests returns at least k_{\min} neighbors, then, as a last resort, perform an exhaustive search for that sample. Although this falls short of a precise estimate of the true size of the linear region about each point (neighborhood definition), it is nonetheless an improvement.

G. Variation: Accounting for Noisy Data

Noisy data pose a risk to any methodology involving multidimensional scaling, which, of course, is at the heart of ISOMAP and its variants. In HSI imagery, sensor noise can degrade performance, especially when landmarks are used in the estimation of the manifold coordinates. In linear mixing, it is common to use some methodology to reduce noise prior to estimating end-members. For example, as implemented in the commercial software package ENVI/IDL [48], the pixel purity index method of identifying end-members [16]–[18] is usually preceded by the application of the MNF [33] transformation of the data cube. Naturally, with the use of landmarks in ENH-ISOMAP and the extension to full scenes presented later in this paper, we are led to consider the utility of preceding the manifold calculation by some means of smoothing. Although not the only method available, one possibility is to use the MNF transform and then calculate the manifold coordinates of the MNF-transformed data using ENH-ISOMAP. A question that naturally arises is: given that MNF reduces the noise and achieves a certain level of separability of data categories, does the hybrid approach yield a better level of separation of the data? We address this in Section III-A. We note that smoothing prior to applying the manifold estimation algorithm has been considered in other applications such as face recognition using another estimation technique similar to ISOMAP, known as the Laplacian Eigenmap [34].

H. Estimating the Global Manifold: Modified Backbone Method of Scaling and Merging Manifold Coordinates

Although the pseudoinverse procedure described in [8] produces a practical solution to the problem of scaling manifold coordinate representations to the required scale of $O(10^6)$ pixels or greater, it does sometimes produce artifacts ([8, Fig. 6] and Fig. 3 of this paper), such as visible seams or slight tonal shifts between merged tiles. There are several causes of these artifacts in the “pseudoinverse reconstruction” method.

For example, the subsets to be merged may have slightly different underlying manifold embedding eigenspectra, even for subsets constructed with random sampling. The effect of this is the potential for a small but visually apparent torquing of the manifold sheets from the two different subsets, which appears as a seam if the subsets are contiguous tiles [8], or as graininess if random sampling is used. Another problem with the pseudoinverse method is that it relies on a smaller subset of points to be mapped from one subset to the other. For sparsely populated dimensions, this is a potential problem because lower order manifold coordinates will have more artifacts, a feature that we have observed. Although these problems can be mitigated somewhat with larger data sets, a better solution exists that overcomes these artifacts. The solution is the modified backbone method that we now describe.

The modified backbone method involves first obtaining a sufficiently large subset of the data that represents the typical spectral variation present in the data set. This can be obtained by random sampling, decimation, active sampling, or simply by manually extracting representative scene subsets. In practice, we find that between 10% and 33% of the hyperspectral scene

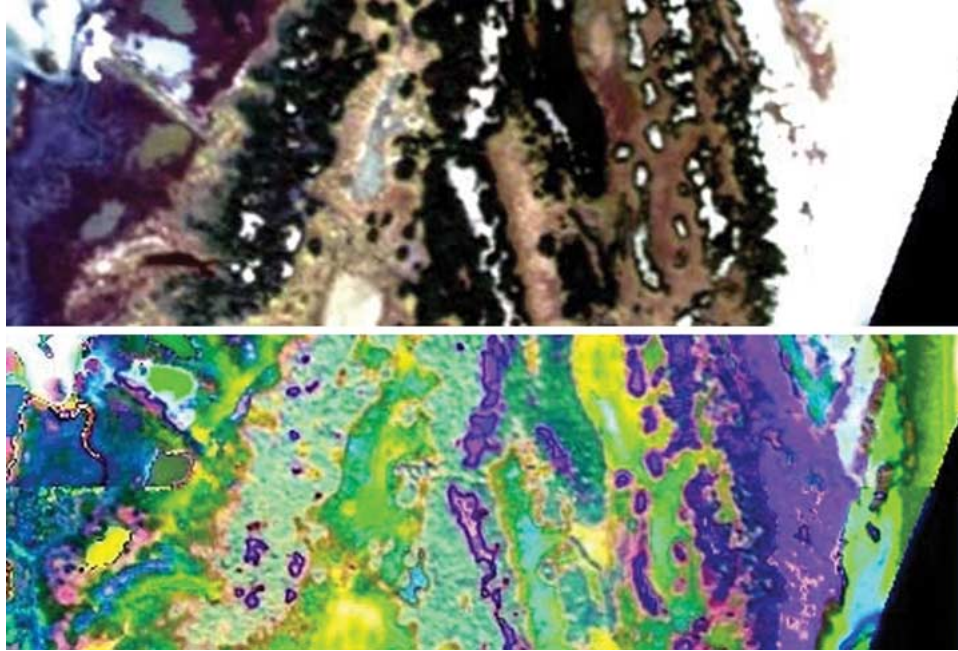


Fig. 3. (Bottom) Manifold coordinates for a subset of (top) a PROBE2 hyperspectral scene of Hog Island, VA, on October 18, 2001, revealing details about species-level spatial distributions. Shown: manifold coordinates 10-11-12. Manifold coordinates were optimized using Landmarks ISOMAP for two tiles, which were rows of size 75×825 and 75×750 pixels (only the land subset of these tiles is shown here). The manifold coordinates were aligned using the pseudoinverse method described in [8]. Note the presence of small horizontal seams at the boundary between the two tiles in the middle of manifold coordinates 10-11-12 toward the edge of the scene.

may be sufficient to produce a usable product; it depends on the scene complexity and the level of detail required by the application. The next step is to obtain a set of manifold coordinates for the backbone subset using ENH-ISOMAP and then insert the remaining data into the backbone manifold coordinate system using a reconstruction principle similar to one used in local linear embedding (LLE) [47]. The difference here is that we insert into a globally optimized manifold coordinate system derived from a large representative backbone. The reconstruction principle uses the fact that the manifold is locally linear so that, for sample vector X_i not in the backbone, its reconstruction, denoted X'_i in the original full hyperspectral input data space, is a simple linear combination of neighboring points in the backbone, denoted B_j , i.e.,

$$X'_i = \sum_{\text{neighbors } j} W_{ij} B_j. \quad (10)$$

The same transformation set of weights must apply in the reduced manifold coordinates

$$m'(X_i) = \sum_{\text{neighbors } j} W_{ij} m(B_j). \quad (11)$$

In (10), we use the fast Vantage Point Forest method that we also incorporated into ENH-ISOMAP.

We obtain the reconstruction weights W in a different manner than that originally described in [47]. In our approach, we directly obtain the weights W by inverting (10) using the eigenvectors and eigenvalues of the covariance matrix formed from

the difference vectors between the sample to be reconstructed and its neighbors, i.e.,

$$\vec{d}_j = \vec{X}_i - \vec{B}_j. \quad (12)$$

With this definition, the covariance matrix of interest is

$$C_{jk} = \vec{d}_j * \vec{d}_k \quad (13)$$

for which the solution to the reconstruction weights is

$$W_i = \frac{\sum_k C_{ik}^{-1}}{\sum_j \sum_k C_{jk}^{-1}} \quad (14)$$

with

$$C_{jk}^{-1} = V Q V^T \quad (15)$$

where V is a matrix whose columns are the eigenvectors of the matrix C , and Q is a matrix that is zero on the off diagonal, with diagonal entries equal to the inverse of the eigenvalues

$$Q_{ii} = \frac{1}{\lambda_i} \quad \text{and} \quad Q_{ij} = 0 \quad \text{for } i \neq j. \quad (16)$$

The solution to (12)–(16) can be obtained from a variety of standard numerical methods for determining eigenvalues, eigenvectors, and inverse matrices, e.g., singular value decomposition, QR decomposition, or LU decomposition.

We term the above method, in which a large representative backbone manifold coordinate is obtained using ENH-ISOMAP, and all other points are inserted into the backbone via reconstruction, the modified backbone method. Note that with the modified backbone approach, only a single eigenspectrum

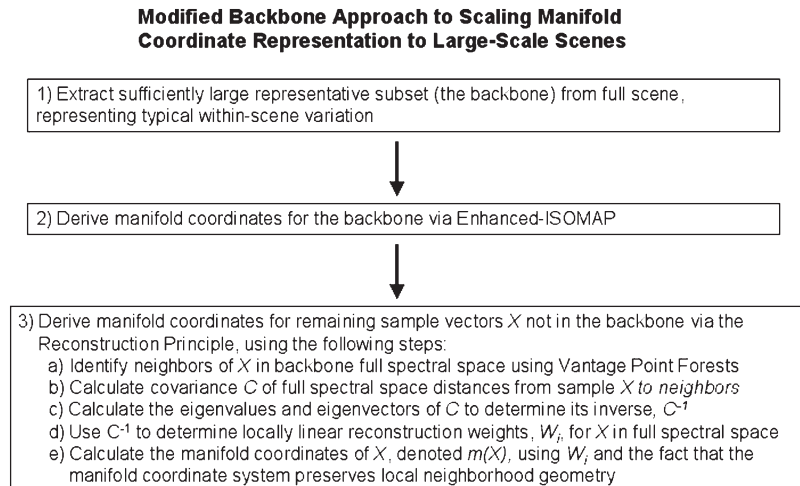


Fig. 4. Our “modified backbone” approach incorporating ENH-ISOMAP.

is computed, so statistical variations in sample-to-sample eigen-spectra are eliminated as a potential source of artifacts in the global manifold product. Also, by reconstructing all remaining points via neighbors in the backbone, errors, if they do occur, will be localized rather than producing global seams, as seen in the pseudoinverse approach. With the reconstruction-based modified backbone, the overall quality of the global manifold improves as the sample size of the backbone increases.

We have now defined the critical steps necessary to achieve a solution to the problem of constructing the manifold coordinates for large hyperspectral scenes of $O(10^6)$ or greater. The approach is the modified backbone algorithm described above and shown schematically in Fig. 4. There are three critical steps, namely 1) extraction of a sufficiently representative subset of the data (in practice 10%–33% is sufficient), which can be quite large ($\sim 100\,000$ – $300\,000$ or greater); 2) the use of our new ENH-ISOMAP method to optimize the manifold coordinates rapidly and without exhausting typical computer memory; and 3) the insertion of the samples not in the backbone into the backbone manifold coordinate system using the reconstruction principle first described in (10) and (11). To obtain the reconstruction weights, we construct the inverse of (10) by inverting the covariance matrix from its eigenvalues and eigenvectors [as described in (12)–(16)].

III. RESULTS

A. Example: Classification of Coastal Land-Cover

In [8], we demonstrated that the original ISOMAP algorithm could be used to separate two spectrally similar categories (*Scirpus spp.* and *Phragmites australis*, an invasive plant species) in HSI imagery of Smith Island, VA. The problem of choosing the best classifier has also been addressed with the original ISOMAP algorithm in [21] in Hyperion imagery and in other application areas such as phoneme recognition, handwritten text recognition, and text classification [15]. As noted earlier, none of these previous works address the issue of scaling to large sample scenes, which is the subject of this paper.

In this section, we return to a PROBE2 hyperspectral scene of Smith Island, VA, from October 18, 2001, and ground truth that we obtained through extensive field validation [5]–[7]. This data set was the basis of results we obtained for the classification of coastal land-cover using machine learning methods. To evaluate ENH-ISOMAP, this time, we used a larger set of surveyed categories than in [8]. In what follows, we have chosen a large set of regions of interest (ROIs) determined from the differential GPS (DGPS) survey data. In this first example, we focus on the southern end of the island, where a large portion of the ground truth was obtained. Of the 19 categories used in our earlier classification experiments [5]–[7], 18 are represented in our survey data for the southern end of the island. In addition, subsequent survey efforts that we have made have allowed us to divide one of the previous categories into two separate ones in the upland (Pine versus Pine/Hardwood complex). One additional category was identified in a follow-up survey and is included in the classification experiment of this subsection. The 20 categories used in our test are shown in Table III.

We compared the separability of the surveyed regions of the first 19 categories (the original categories described in earlier publications) in Table III for MNF, ENH-ISOMAP, and an MNF/ENH-ISOMAP hybrid. The hybrid was used to first reduce noise before optimizing the data manifold coordinates. For the MNF case, 30 components were extracted, and for the MNF/ENH-ISOMAP hybrid, 30 MNF components were extracted and then used as the basis for constructing a set of 30 manifold coordinates. In [8], we used small data subsets, roughly 10 000 samples, with the original ISOMAP algorithm to show the potential for improved data compression over MNF. In contrast, in this paper, we extracted manifold coordinates for the entire southern end of the island (excluding the lagoon and ocean), a total of 261 000 samples.

The separability measure chosen was the Jeffries–Matsushita (JM) distance [43], [52] between the set of surveyed regions of each category treated as group and the group of ROIs for each of the other categories. The JM distance between two classes is given by

$$J_{ij} = \sqrt{2(1 - e^{-\alpha})} \quad (17)$$

TABLE III
 TERRESTRIAL CATEGORIES FOR PROBE2 HSI IMAGERY TESTS, SMITH ISLAND, VA, OCTOBER 18, 2001. BASED ON GPS AND DGPS SURVEYS. FORMAT: LATIN NAME, COMMON NAME, NUMBER OF SPECTRAL SAMPLES IN TESTS

(1) <i>Phragmites australis</i> (Common Reed): 1066	(2) <i>Spartina alterniflora</i> (Smooth Cordgrass) : 1459
(3) <i>Spartina patens</i> (Salt Hay): 358	(4) <i>Salicornia virginia</i> (<i>Perennial glasswort</i>): 534
(5) <i>Borrchia frutescens</i> (Sea Ox-eye): 517	(6) <i>Juncus roemerianus</i> (Spike Needle Rush): 990
(7) Inland Water Body: 243	(8) <i>Distichlis spicata</i> (Saltgrass): 795
(9) <i>Scirpus spp.</i> (Saltmarsh Bulrush): 592	(10) "wrack": 1083
(11) Mudflat/saltflat : 236	(12) Beach/sand : 997
(13) <i>Ammophila breviligulata</i> (American Beachgrass) : 152	(14) <i>Andropogon spp.</i> (Broomsedge): 290
(15) <i>Uniola paniculata</i> (Sea Oats): 36	(16) <i>Myrica cerifera</i> (Bayberry) thicket: 436
(17) Pine/hardwood complex: 1685	(18) Pine only: 219
(19) <i>Iva frutescens</i> (Marsh-elder): 319	(20) <i>Typha spp.</i> (Cattail): 69

with

$$\alpha = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|C_i + C_j|}{2\sqrt{|C_i||C_j|}} \right) \quad (18)$$

where C_i is the covariance matrix of category i , $|C_i|$ is its determinant, and μ_i is its associated mean vector.

We consider here the question of what happens once the manifold coordinates are applied to an MNF-transformed HSI data. For the present example of 261 000 samples, we obtained a rank graph from worst to best separation for the first three components of the MNF representation and of the ENH-ISOMAP manifold coordinates when the data had been presmoothed by MNF. Fig. 5 shows the rank graphs using the JM distance measure for the first three components of MNF and the MNF/ENH-ISOMAP hybrid. Note that the ~ 60 least separable pairs of categories are more separable under this measure for MNF/ENH-ISOMAP than for MNF alone when three components are retained. MNF/ENH-ISOMAP always achieves better separation of the categories than MNF alone, and the effect is larger when fewer components are retained. This is an important benefit in cases where compression of data may be a consideration. We performed a set of simple classification experiments with a maximum-likelihood classifier for both MNF and MNF/ENH-ISOMAP representations for all 20 categories in Table III. The number of retained components was varied, and trials were formed by constructing a 50% stratified random sample across the 20 categories; we made two exceptions to the percentages due to smaller sample size in two of the categories. The choice of classifier was made for the sake of simplicity, not through an exhaustive search of possibilities. There are some potential limitations of this choice. For instance, we implicitly assume that the individual category representations do not contain disjoint subdistributions or multiple modes. Nevertheless, we can use this widely accepted method to indicate some general trends in the representation.

We found that when fewer components were retained (the dominant components in each representation), the MNF/ENH-ISOMAP hybrid correctly identified a significantly larger fraction of pixels. For example, with just three components retained, the fraction of pixels correctly classified was $56.15\% \pm 0.41\%$ for the hybrid compared with $46.68\% \pm 0.40\%$ for MNF. This is consistent with the earlier observations made about category separability in the dominant components for the JM distance measures used earlier. Likewise, it is consistent

with our earlier observations about the potential for hyperspectral data compression with the original ISOMAP algorithm [8], [11], and [12]. As the number of components increased, the average performance of the two methods begin to merge. At the other end of the continuum, for 30 components, the MNF result is slightly higher than the hybrid: $88.88\% \pm 0.31\%$ (MNF) versus $85.39\% \pm 0.34\%$ (MNF/ENH-ISOMAP). The smaller loss in performance for a larger number of components in the hybrid is probably due to three factors. The first is the approximation of geodesic distances using landmarks. The second is the relatively simple assumption about neighborhood size that exists within all ISOMAP implementations, including ENH-ISOMAP. Both of these factors contribute to inaccuracies in our estimate of the true curvature of the geodesic surfaces of the original data space. The approximation errors are smaller effects, so they are more likely to appear in the lower order components. The third factor is the choice of classifier and is related to the fact that some distributions may appear with multiple modes.

B. Global Manifold Products

An example of the modified backbone approach applied to the full PROBE2 airborne hyperspectral scene [5], [7] of a barrier island on Virginia's eastern shore is shown in Fig. 6. A subset of this scene was used in the previous feature analysis and classification results.

The full input hyperspectral scene contained 124 spectral channels, and there were $\sim 1.8 \times 10^6$ spectral samples. Examples of RGB plots of combinations of the 20 manifold coordinates extracted in this example are shown in Fig. 6 for the full scene. In this case, the backbone manifold coordinates were constructed from a subset containing roughly 12% of the scene. There were 219 357 spectral sample vectors each of 124 dimensions in the backbone, and using ENH-ISOMAP on an Athlon64 3000+ 2.2-GHz processor, the time to calculate the manifold coordinates was 40.4 min, roughly two-thirds of an hour. The time to insert the samples not in the backbone into the backbone by deriving their manifold coordinates via reconstruction was 3.75 h, again on the same computer. Thus, the total processing time on a single CPU in this example was 4.4 h.¹

¹In the examples presented here, we also reconstructed the original backbone, which added roughly an additional 17 min of processing time

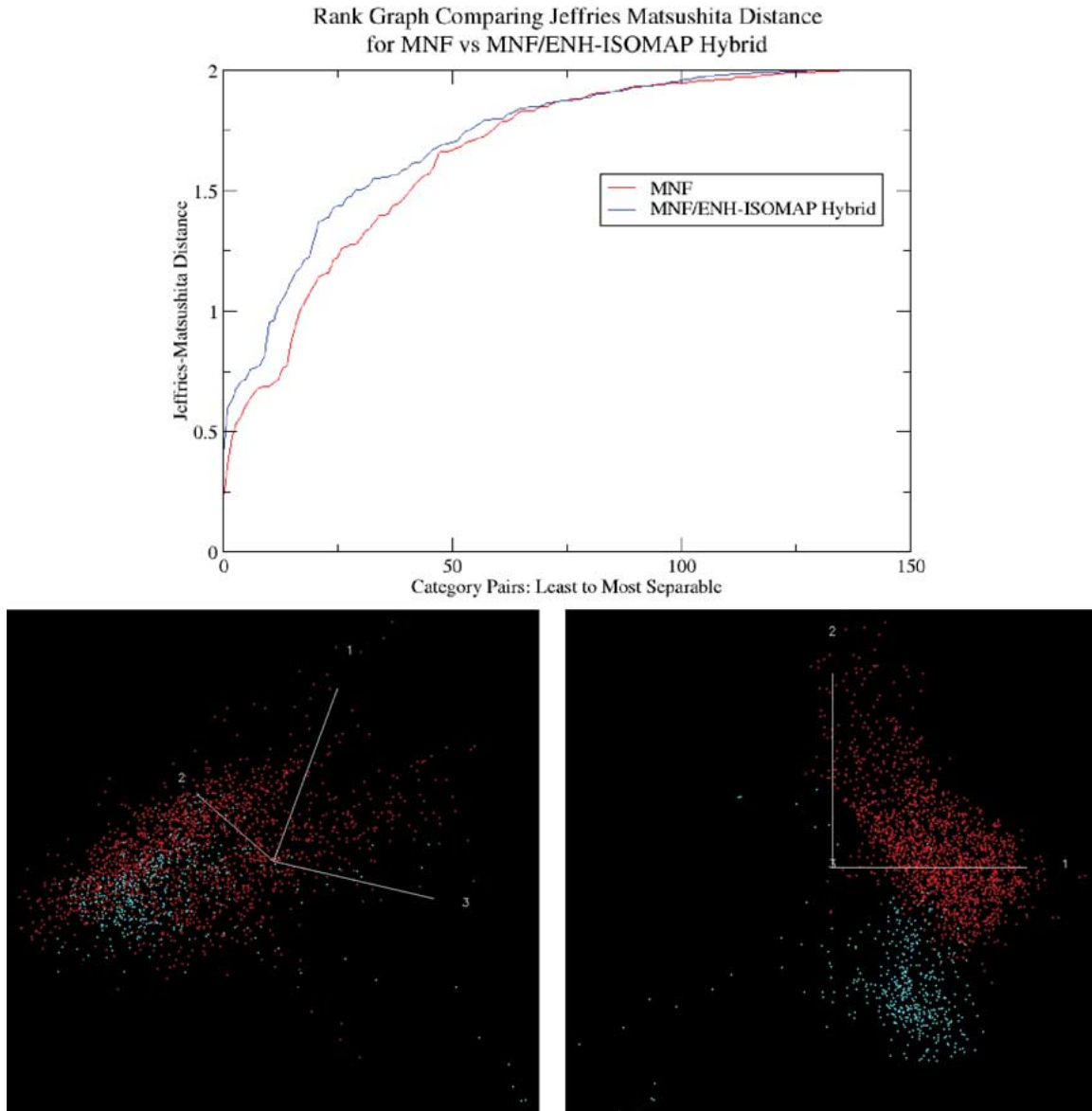


Fig. 5. Rank graphs of JM distances for category pairs in Table III using surveyed regions from PROBE2 HSI scene. Along x axis: worst to best separable pair. (Top) First 130 ranked of pairs of 171 possible pairs, MNF versus MNF/ENH-ISOMAP hybrid for the first three components. Scatterplots: corresponding separation of one of the pairs, Pine/Hardwood Complex (red) versus *Myrica cerifera* shrub (cyan). (Bottom, left) MNF, JM distance = 0.539. (Bottom, right) MNF/ENH-ISOMAP Hybrid, JM distance = 1.387.

Note that in a parallel processing implementation, the time to insert the samples into the backbone, i.e., to derive their coordinates in the backbone manifold coordinate system, would scale as $\sim 1/N_{\text{CPUs}}$ since the backbone itself is not modified during the reconstruction. This would be achieved by sending copies of the backbone full spectral cube, its associated Vantage Point Forests, and its corresponding manifold coordinates to each processor. From the portion of the data not already calculated in the backbone, each processor would also receive only one subset of the full spectral data of the scene, disjoint from the other full spectral subsets received by the other CPUs. Each CPU would be responsible for reconstructing just that single subset in the backbone manifold coordinate system. Thus, for example, the 3.75 h that it took to complete the calculation of the manifold coordinates of the nonbackbone samples on a single CPU computer would

take approximately 3.5 min on a 64-processor multiprocessor system with the same CPU speed. This estimate assumes that all CPUs were used and ignores the overhead needed to send copies of the various subsets to each processor and retrieve the results.

A detail from some of these manifold coordinate combinations is shown in Figs. 7 and 8, which depict, respectively, coordinates 1-2-3 and 7-8-9 for both the manifold coordinate representation and the MNF representation. Note that unlike the earlier classification example that dealt with only the terrestrial portion of the data, the manifold in this example was not preprocessed by MNF. Figs. 7 and 8 show the RGBs of these coordinate representations for a subset of the full scene near a delta at the inlet to a tidal creek, as well as the surrounding marsh, dunes, grasslands, beach, and ocean and lagoonal waters. Figs. 6-8 show the highly significant

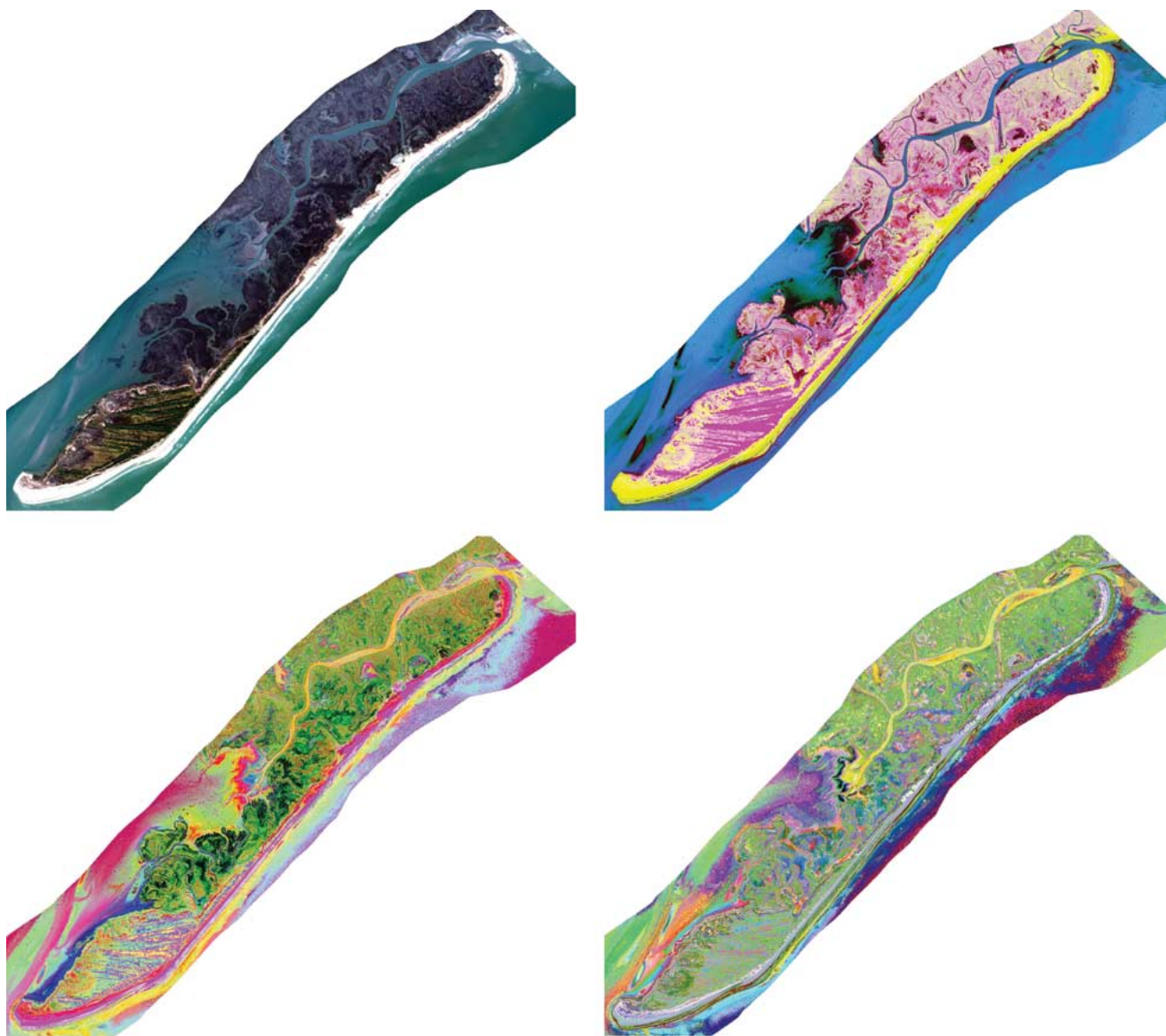


Fig. 6. Full-scene manifold coordinates derived using the modified backbone approach. (Top, left) RGB from PROBE2 scene of Smith Island, VA, October 18, 2001. (Top, right) Manifold coordinates 1-2-3. (Bottom, left) Manifold coordinates 7-8-9. (Bottom, right) Manifold coordinates 13-14-15.

structure revealed in the water, land, and land/water margin by the manifold coordinate representations. Note the rich detail in the water column, which may be related to both bathymetric structure as well as in-water constituents [12], [13]. Note that the manifold coordinates provide greater detail, especially in the delta, the shallow waters near the lagoonal shore, and in the adjacent tidal creeks. This is consistent with similar results that we have obtained with PHILLS [25] HSI in the Indian River Lagoon, FL [12]. In that study, we showed that significantly greater detail could be found with the manifold coordinate representation when compared with MNF, and the resulting coordinates strongly correlated with the in-water properties such as bathymetry. Subsequently, we have shown that parameterization and prediction of bathymetry [13] with global manifold products can be obtained with good accuracy. Indeed, in the present example, it appears likely that bathymetric structure, bottom type, and in-water constituents may all be represented in the manifold coordinates. The land

portion of the manifold coordinates also provides a particularly detailed delineation of marsh, dune, scrub–shrub, and grassland structure.

IV. CONCLUSION AND SUMMARY

Our prior efforts [8] to develop a global manifold for hyperspectral scenes of size $O(10^6) - O(10^7)$ pixels contained some fundamental limitations due to both scaling and sampling. We had relied primarily on a pseudoinverse reconstruction technique that aligned the estimated manifold coordinate systems from small subsets. Although feasible, the approach produced some artifacts, which motivated us to develop the improved methodology described in this paper. In this paper, we took a number of specific steps that included improving the scaling of the ISOMAP algorithm itself by incorporating a new method, i.e., Vantage Point Forests, for the rapid initialization of sparse graph neighborhoods used in the geodesic distance

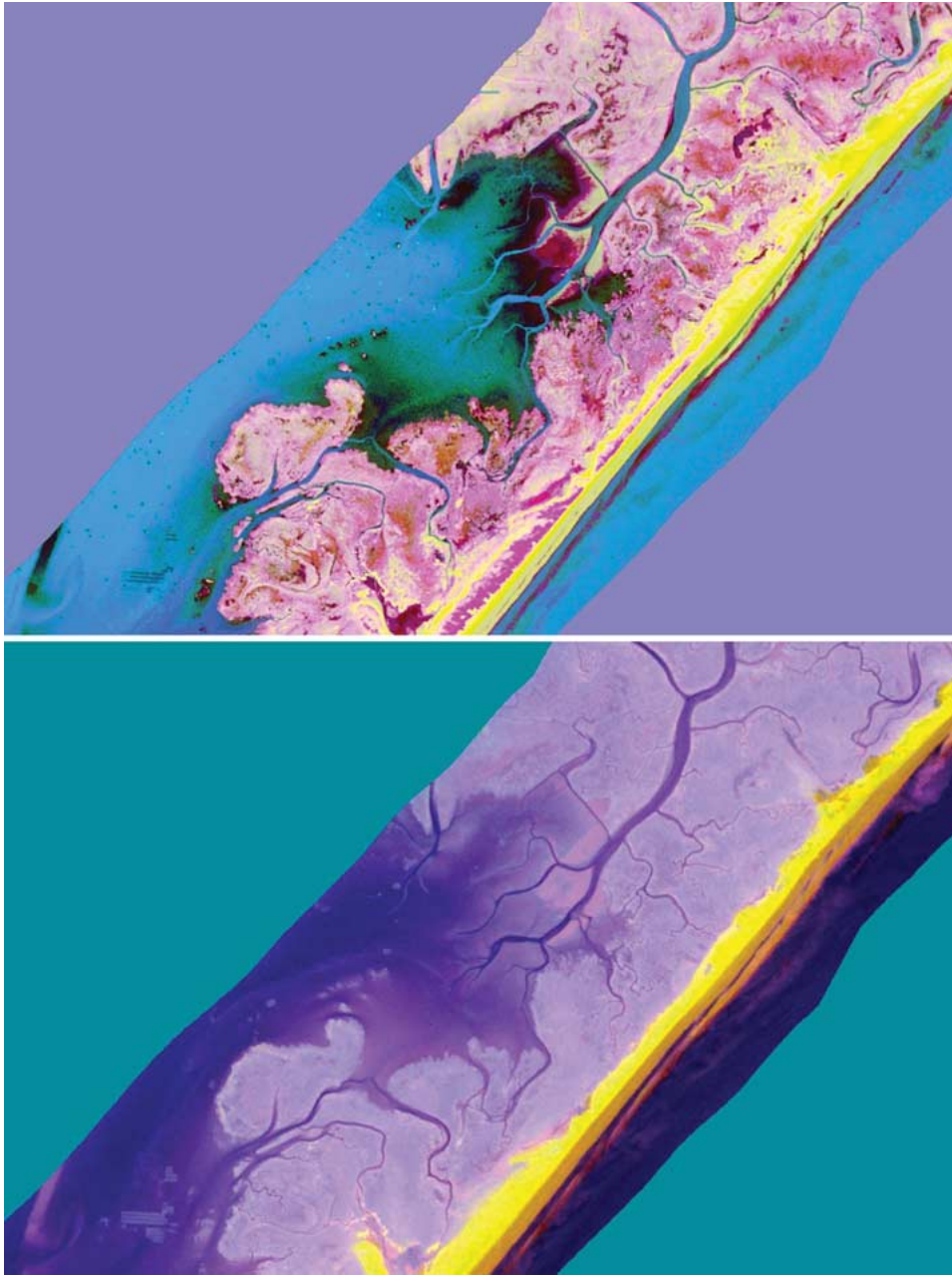


Fig. 7. (Top) Detail from Fig. 6 showing manifold coordinates 1-2-3 in the vicinity of a mid-island delta near tidal creek, beach, dunes, scrub-shrub, grassland vegetation, lagoonal waters, and ocean. (Bottom) Linear MNF coordinates 1-2-3 for the same region.

calculation. This guarantees computational scaling no worse than $O(N \log^2(N))$, but in practice, in many instances, it achieves nearly linear scaling behavior. The improvements also incorporated previous efforts described in [23], where an approximate means of circumventing the high computational and memory cost of calculating all $N \times N$ geodesic distances was described. The latter replaced the full $N \times N$ geodesic distance problem with a smaller embedded problem of finding all geodesic distances to a set of “landmarks.” Our ENH-ISOMAP algorithm incorporated our Vantage Point Forests algorithm, landmarks ISOMAP, and other modifications. One of these was related to patching isolated points, our FASTPATCH algorithm, in a self-consistent manner prior to undertaking the geodesic distance calculation,

and a more flexible neighborhood definition that incorporated aspects of the absolute radius and K-neighbor definitions previously used in [53]. Likewise, we have also described an improved landmark selection process, i.e., skeletonization. With ENH-ISOMAP, it is now possible to calculate manifold coordinates for subsets with $O(10^5)$ samples, which is an improvement of an order of magnitude. With the backbone reconstruction (modified backbone algorithm) that we also developed in this paper, this solution for $O(10^5)$ samples can readily be extended to usual scene sizes of $O(10^6)$ samples or greater.

One potential drawback to our overall solution is the use of landmarks, which may not fully span the original space if too few landmarks are used. This is again a sampling issue because for the landmarks approximation to work, we must be

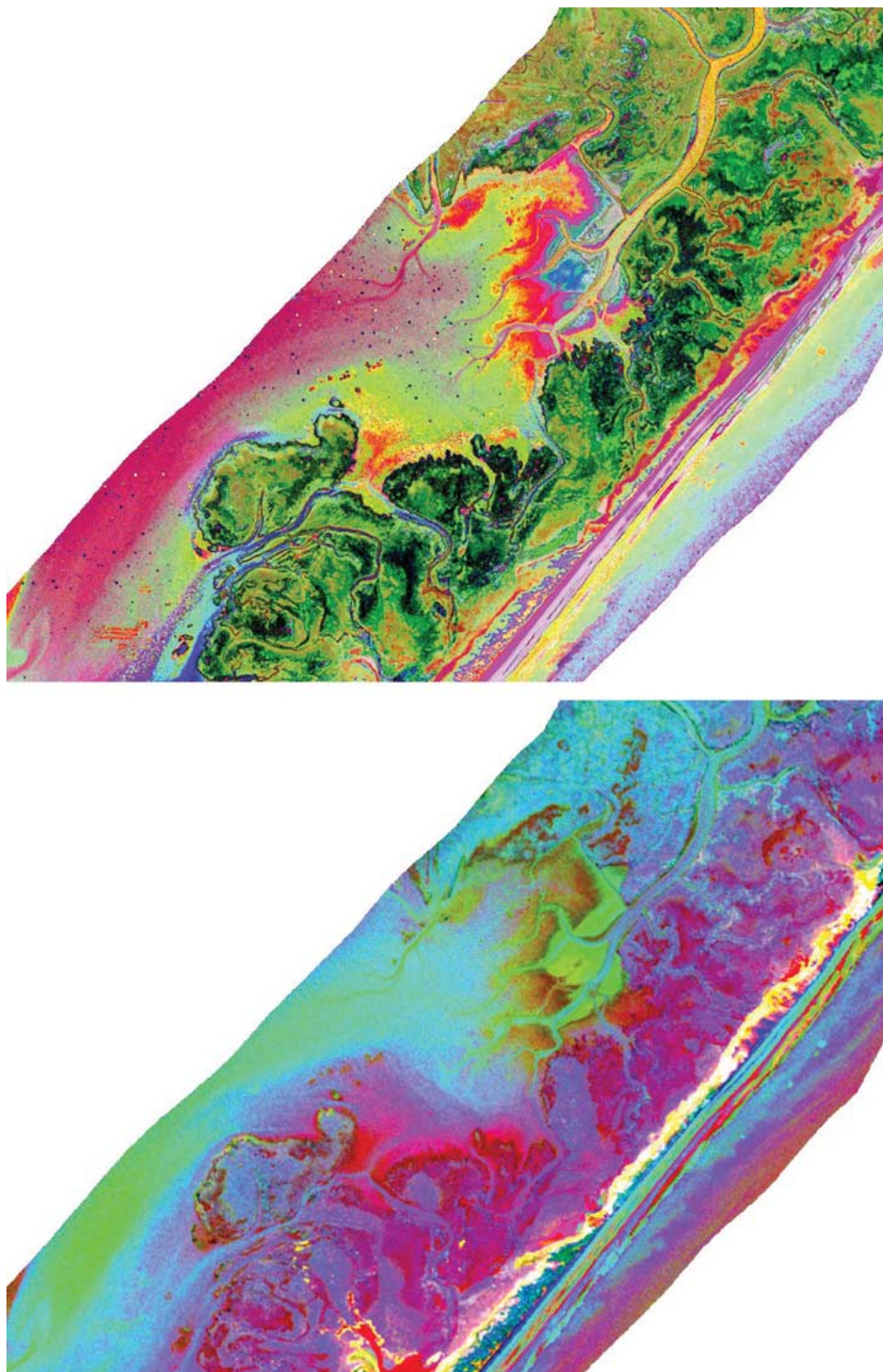


Fig. 8. (Top) Detail from Fig. 6 showing manifold coordinates 7-8-9 in the vicinity of a mid-island delta near tidal creek, beach, dunes, scrub-shrub, grassland vegetation, lagoonal waters, and ocean. (Bottom) Linear MNF coordinates 7-8-9 for the same region.

able to guarantee that even sparsely populated dimensions will be fully spanned by the landmarks in the manifold coordinate space. Failure to span these dimensions properly will result in poor performance of subdominant categories in classification experiments. It also suggests a strategy for the future, which is to define a method for landmark selection that results in a guarantee that all dimensions in the linear embedding space are fully spanned. The skeletonization procedure described in

this paper is one such landmark selection approach, but there may be ways to improve it. Another potential drawback is the sampling associated with the backbone. In the examples that we have provided, a reasonable result was produced, but the degree of accuracy required in the reconstruction process may depend on the downstream use of the manifold product. Nevertheless, the concrete results that we have produced for full scene manifolds of size $O(10^6)$ suggest the tremendous potential of

our new method now that it can be applied effectively to these scales and larger.

Our specific examples demonstrated that the manifold coordinate system provides a very powerful delineation of in-scene constituents. Of particular note was the fact that the shallow waters in our coastal scene example were much more effectively delineated when compared with traditional linear processing such as MNF [33], revealing greater information potentially related to bathymetric structure and in-water constituents. The latter is not surprising given the results that we have previously described in relation to the improved compression capabilities of manifold coordinate representations over that of traditional linear methods [8], [12]. The approach to developing full scene manifolds presented in this paper has also been used as the basis of estimating bathymetry from HSI imagery [13].

In this paper, we also described the results of some simple land-cover classification experiments from coastal HSI imagery using maximum likelihood. We found that when only the dominant features of the ENH-ISOMAP representation were used, significantly better classification performance was achieved compared with MNF. For experiments retaining a larger number of components, MNF was only slightly better. The latter was probably principally the result of 1) the approximation errors introduced by the use of landmarks, 2) the imperfect estimate of local neighborhood size present in all ISOMAP variants, and 3) possibly the choice of classifier and the assumption of unimodal category distributions. The improved results with ENH-ISOMAP for fewer components also served to further illustrate the potential for better compression that we first addressed in [8] with the original ISOMAP algorithm.

Calculation of a manifold coordinate system from the geodesic distances between pixels of the underlying nonlinear hyperspectral data manifold provides a compact representation of the hyperspectral information. These new coordinates inherently incorporate the correlations between the various hyperspectral channels and, therefore, provide an efficient and convenient set of coordinates for image classification and target and anomaly detection.

ACKNOWLEDGMENT

C. M. Bachmann and R. A. Fusina would like to thank the DOD High Performance Computing Modernization Program for computing resources at the Maui High Performance Computing Center, the Army Research Laboratory's Major Shared Resource Center, and SMDC.

REFERENCES

- [1] D. K. Agrafiotis, "Stochastic proximity embedding," *J. Comput. Chem.*, vol. 24, no. 10, pp. 1215–1221, Jul. 2003.
- [2] T. L. Ainsworth and J. S. Lee, "Optimal polarimetric decomposition variables—Non-linear dimensionality reduction," in *Proc. IGARSS*, Sydney, Australia, 2001, vol. 2, pp. 928–930.
- [3] —, "Optimal image classification employing optimal polarimetric variables," in *Proc. IGARSS*, Toulouse, France, 2003, vol. 2, pp. 696–698.
- [4] —, "Polarimetric image classification: Exploiting optimal variables derived from multi-image data sets," in *Proc. IGARSS*, 2004, pp. 188–191.
- [5] C. M. Bachmann, "Improving the performance of classifiers in high-dimensional remote sensing applications: An adaptive resampling strategy for error-prone exemplars (ARESEPE)," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2101–2112, Sep. 2003.
- [6] C. M. Bachmann, T. F. Donato, G. M. Lamela, W. J. Rhea, M. H. Bettenhausen, R. A. Fusina, K. DuBois, J. H. Porter, and B. R. Truitt, "Automatic classification of land-cover on Smith Island, VA using HYMAP imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2313–2330, Oct. 2002.
- [7] C. M. Bachmann, M. H. Bettenhausen, R. A. Fusina, T. F. Donato, A. L. Russ, J. Burke, G. M. Lamela, W. Joseph Rhea, B. R. Truitt, and J. H. Porter, "A credit assignment approach to fusing classifiers of multiseason hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2488–2499, Nov. 2003.
- [8] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
- [9] —, "Modelling data manifold geometry in hyperspectral imagery," in *Proc. IGARSS*, 2004, pp. 3203–3206.
- [10] —, "Improvements to land-cover and invasive species mapping from HSI in the Virginia Coast reserve," in *Proc. IGARSS*, 2004, pp. 4180–4183.
- [11] —, "Improved manifold coordinate representations of hyperspectral imagery," in *Proc. IGARSS*, 2005, pp. 4307–4310.
- [12] C. M. Bachmann, T. L. Ainsworth, D. B. Gillis, S. J. Maness, M. J. Montes, T. F. Donato, J. H. Bowles, D. R. Korwan, R. A. Fusina, G. M. Lamela, and W. J. Rhea, "A new data-driven approach to modeling coastal bathymetry from hyperspectral imagery using manifold coordinates," in *Proc. MTS/IEEE Oceans Conf.*, Washington, DC, Sep. 19–23, 2005, pp. 2242–2249.
- [13] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Modeling coastal waters from hyperspectral imagery using manifold coordinates," in *Proc. IGARSS*, 2006, to be published.
- [14] M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva, and J. C. Langford, "The isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, p. 7, Jan. 4, 2002.
- [15] M. Belkin and P. Niyogi, *Semi-Supervised Learning on Riemannian Manifolds*. Norwell, MA: Kluwer, 2004.
- [16] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. 4th Annu. JPL Airborne Geosci. Workshop*, Pasadena, CA, 1993, vol. 1, pp. 11–14. JPL Pub. 93-26.
- [17] —, "Analysis, understanding and visualization of hyperspectral data as a convex sets in n-space," *Proc. SPIE*, vol. 2480, pp. 14–22, 1995.
- [18] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data: In summaries," in *Proc. 5th JPL Airborne Earth Sci. Workshop*, 1995, vol. 1, pp. 23–26. JPL Publication 95-1.
- [19] T. Bozkaya and M. Ozsoyoglu, "Distance-based indexing for high-dimensional metric spaces," in *Proc. ACM SIGMOD*, 1997, pp. 357–368.
- [20] B. Bustos, G. Navarro, and E. Chavez, "Pivot selection techniques for proximity searching in metric spaces," *Pattern Recognit. Lett.*, vol. 24, no. 14, pp. 2357–2366, Oct. 2003.
- [21] Y. Chen, M. M. Crawford, and J. Ghosh, "Applying nonlinear manifold learning to hyperspectral data for land cover classification," in *Proc. IGARSS*, Seoul, Korea, 2005, pp. 4311–4314.
- [22] T. H. Cormen, L. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.
- [23] V. de Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2002, pp. 705–712.
- [24] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.
- [25] C. O. Davis, J. Bowles, R. A. Leathers, D. Korwan, T. V. Downes, W. A. Snyder, W. J. Rhea, W. Chen, J. Fisher, W. P. Bissett, and R. A. Reisse, "Ocean PHILLS hyperspectral imager: Design, characterization, and calibration," *Opt. Express*, vol. 10, no. 4, pp. 210–221, Feb. 2002.
- [26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [27] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976.
- [28] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An Algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 209–226, Sep. 1977.
- [29] T. Han and D. G. Goodenough, "Nonlinear feature extraction of hyperspectral data based on local linear embedding (LLE)," in *Proc. IGARSS*, 2005, pp. 1237–1240.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining Inference, and Prediction*. New York: Springer-Verlag, 2001.

- [31] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
- [32] D. G. Goodin, J. Gao, and G. M. Henebry, "The effect of solar illumination angle and sensor view angle on observed patterns of spatial structure in Tallgrass Prairie," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 154–165, Jan. 2004.
- [33] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
- [34] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–346, Mar. 2005.
- [35] D. H. Kim and L. H. Finkel, "Hyperspectral image processing using locally linear embedding," in *Proc. 1st Int. IEEE EMBS Conf. Neural Eng.*, 2003, pp. 316–319.
- [36] B. S. Kim and S. B. Park, "A fast nearest neighbor finding algorithm based on ordered partition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 761–766, Nov. 1986.
- [37] V. Kumar, A. Grama, A. Gupta, and G. Karypis, *Introduction to Parallel Computing, Design and Analysis of Algorithms*. Redwood City, CA: Benjamin Cummings, 1994.
- [38] S. Lubiarsz and P. Lockwood, "Evaluation of fast algorithms for finding the nearest neighbor," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, 1997, vol. 2, pp. 1491–1494.
- [39] Web site for the University of Virginia's Long Term Ecological Research Program. [Online]. Available: <http://www.vclter.virginia.edu>
- [40] N. Katayama and S. Satoh, "The SR-tree: An index structure for high-dimensional nearest neighbor queries," in *Proc. ACM SIGMOD*, 1997, pp. 369–380.
- [41] J. McNames, "A fast nearest-neighbor algorithm based on a principal axis search tree," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 964–976, Sep. 2001.
- [42] C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*. San Diego, CA: Academic, 1994.
- [43] T. Murakami, S. Ogawa, N. Ishitsuka, K. Kumagai, and G. Saito, "Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains, Japan," *Int. J. Remote Sens.*, vol. 22, no. 7, pp. 1335–1348, 2001.
- [44] J. F. Mustard, L. Li, and G. He, "Nonlinear spectral mixture modeling of lunar multispectral data: Implications for lateral transport," *J. Geophys. Res.*, vol. 103, no. E8, pp. 19419–19425, Aug. 25, 1998.
- [45] Web site for NASA EOS Land Validation Core Site at the Virginia Coast Reserve. [Online]. Available: <http://landval.gsfc.nasa.gov/MODIS/coresite.php?SiteID=27>
- [46] D. A. Roberts, M. O. Smith, and J. B. Adams, "Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 255–269, May/June. 1993.
- [47] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 22, 2000.
- [48] ENVI and IDL are products of ITT Visual Information Solutions. [Online]. Available: www.itvis.com/index.asp
- [49] S. R. Sandmeier, E. M. Middleton, D. W. Deering, and W. Qin, "The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization," *J. Geophys. Res.*, vol. 104, no. D8, pp. 9547–9560, Apr. 27, 1999.
- [50] R. Sedgewick, *Algorithms in C++, Part 5: Graph Algorithms*. Boston, MA: Addison-Wesley, 2002.
- [51] G. Sleijpen and H. van der Vorst. (1994). *Jacobi-Davidson Methods*. [Online]. Available: <http://www.cs.utk.edu/dongarra/etemplates/node147.html>; <http://www.cs.utk.edu/dongarra/etemplates/node136.html>
- [52] P. H. Swain and S. M. Davis, Eds., *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill, 1978.
- [53] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 22, 2000.
- [54] Site description of the Virginia Coast Reserve barrier islands owned by the Nature Conservancy. [Online]. Available: <http://nature.org/wherework/northamerica/states/virginia/preserves/art1244.html>
- [55] P. N. Yianilos, "Data structure and algorithms for nearest neighbor search in general metric spaces," in *Proc. ACM-SIAM Symp. Discr. Algorithms*, 1993, pp. 311–321.



Charles M. Bachmann (M'92–SM'04) received the A.B. degree from Princeton University, Princeton, NJ, in 1984, and the Sc.M. and Ph.D. degrees from Brown University, Providence, RI, in 1986 and 1990 respectively, all in physics.

In 1990, he joined the Naval Research Laboratory (NRL), Washington, DC, as a Research Physicist in the Radar Division, serving as a Section Head in the Airborne Radar Branch between 1994 to 1996. In 1997, he moved to the Remote Sensing Division, where he is presently Head of the Coastal Science and Interpretation Section of the Coastal and Ocean Remote Sensing Branch. He has been a Principal Investigator for projects funded by the Office of Naval Research, and more recently for an internal NRL project focused on coastal land-cover from hyperspectral and multisensor imagery. His research interests include image and signal processing techniques, adaptive statistical pattern recognition methods, and the instantiation of these methods in software. His research also focuses on specific application areas such as multispectral and hyperspectral imagery, field spectrometry, SAR, and multisensor data fusion, as these apply to environmental remote sensing, especially wetlands and coastal environments.



Thomas L. Ainsworth (M'97–SM'02) received the A.B. degree in mathematics and mathematical physics from Brown University, Providence, RI, the M.Sc. degree from the University of North Carolina, Chapel Hill, and the Ph.D. degree in theoretical nuclear physics from the State University of New York, Stony Brook.

Since joining the Remote Sensing Division, Naval Research Laboratory, Washington, DC, in 1993, his interests have primarily focused on the analysis of and novel applications of polarimetric and interferometric SAR imagery. Current interests include SAR polarimetry, polarimetric SAR calibration, polarimetric interferometry, polarimetric Image classification, joint time-frequency analysis, and nonlinear approaches to hyperspectral data analysis.



Robert A. Fusina (M'01) received the B.S. degree from Manhattan College, New York, and the M.S. and Ph.D. degrees from the State University of New York at Albany, all in physics.

He has worked in the Remote Sensing Division, Naval Research Laboratory, Washington, DC, since 1993. His current research involves land cover classification, hyperspectral remote sensing and data fusion. His previous work included calculation of radar scattering from ocean waves.