Exploiting Manifold Geometry in Hyperspectral Imagery

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

***Abstract*—A new algorithm for exploiting the nonlinear structure of hyperspectral imagery is developed and compared against the *de facto* standard of linear mixing. This new approach seeks a manifold coordinate system that preserves geodesic distances in the high-dimensional hyperspectral data space. Algorithms for deriving manifold coordinates, such as isometric mapping (ISOMAP), have been developed for other applications. ISOMAP guarantees a globally optimal solution, but is computationally practical only for small datasets because of computational and memory requirements. Here, we develop a hybrid technique to circumvent ISOMAP's computational cost. We divide the scene into a set of smaller tiles. The manifolds derived from the individual tiles are then aligned and stitched together to recomplete the scene. Several alignment methods are discussed. This hybrid approach exploits the fact that ISOMAP guarantees a globally optimal solution for each tile and the presumed similarity of the manifold structures derived from different tiles. Using land-cover classification of hyperspectral imagery in the Virginia Coast Reserve as a test case, we show that the new manifold representation provides better separation of spectrally similar classes than one of the standard linear mixing models. Additionally, we demonstrate that this technique provides a natural data compression scheme, which dramatically reduces the number of components needed to model hyperspectral data when compared with traditional methods such as the minimum noise fraction transform.**

***Index Terms*—Airborne Visible Imaging Spectrometer (AVIRIS), bidirectional reflectance distribution function (BRDF), compression, Cuprite, geodesic distance, hyperspectral, isometric mapping (ISOMAP), land-cover classification, linear mixture, local linear embedding (LLE), manifold coordinates, nonlinearity, PROBE2, Virginia Coast Reserve.**

I. Background and Introduction

FOR MANY years, linear mixing [10]–[15], [19], [29], [32], [43], [46] and "best band" combinations have been *de facto* standards in the analysis of spectral data, especially hyperspectral. The "best band" approach relies on the presence of narrowband features that may be characteristic of a particular category of interest [17], [18] or on known physical characteristics of broad classes of data, e.g., vegetation indices [45]. On the other hand, the underlying assumptions of linear mixing are that each pixel in a scene may be decomposed into a finite number of constituent endmembers, which represent the purest pixels in the scene. A number of algorithms have been developed and have become standards; these include the pixel purity index (PPI) [12], ORASIS [13]–[15], [19], N-Finder [46], and iterative spectral unmixing [43].

Manuscript received March 19, 2004; revised October 13, 2004. This work was supported by the Office of Naval Research.

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

Digital Object Identifier 10.1109/TGRS.2004.842292

Although the use of endmembers and indexes based on narrowband features have yielded very useful results, these approaches largely ignore the inherent nonlinear characteristics of hyperspectral data. There are multiple sources of nonlinearity. One of the more significant sources, especially in land-cover classification applications, stems from the nonlinear nature of scattering as described in the bidirectional reflectance distribution function (BRDF) [22], [37]. In land-cover applications, BRDF effects lead to variations in the spectral reflectance of a particular category as a function of position in the landscape, depending on the local geometry. Factors that play a role in determining BRDF effects include the optical characteristics of the canopy, canopy gap function, leaf area index (LAI), and leaf angle distribution (LAD) [37]. It also has been observed that wavelengths with the smallest reflectance exhibit the largest nonlinear variations [37]. Another source of nonlinearity, especially in coastal environments such as coastal wetlands, arises from the variable presence of water in pixels as a function of position in the landscape. Water is an inherently nonlinear attenuating medium [30]. Other effects that contribute to nonlinearities include multiple scattering within a pixel and the heterogeneity of subpixel constituents [26], [34]. In some instances, nonlinear interactions have been modeled explicitly [24], [30], [31], [44].

Recently, a number of papers that address the problem of modeling nonlinear data structure have appeared in the statistical pattern recognition literature [1], [2], [21], [33], [35], [36], [42]. Each of these approaches represents an attempt to derive a coordinate system that resides on (parameterizes) the nonlinear data manifold itself. All of these approaches are data-driven algorithms, not physical or phenomenological models. Nevertheless, they are a very powerful new class of algorithms that can be brought to bear on many high-dimensional applications that exhibit nonlinear structure, e.g., the analysis of remote sensing imagery [3], [4]. One of the first of these approaches to appear—isometric mapping (ISOMAP) [42]—determines a globally optimal coordinate system for the nonlinear data manifold, but it is only practical for small datasets because the dominant computation is based on a determination of all pairwise distances and minimal path distances between all points, which scales as $O(N^3)$, where N is the number of data samples (in our case, the number of pixels in a hyperspectral flight-line or subset). A subsequent paper [8] showed that the computational scaling can be improved to $O(N^2 \log(N))$ using Dijkstra's algorithm [20], [38]. Another significant computational challenge is that ISOMAP memory requirements scale as $O(N^2)$, because ISOMAP requires the extraction of dominant eigenvectors of an $N \times N$ geodesic distance matrix.

0196-2892/$20.00 © 2005 IEEE

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

An alternative approach, local linear embedding (LLE) [35], [36] describes the manifold by modeling each data point as a linear combination of its neighbors; this approach exploits the fundamental property that a manifold is locally linear. Like ISOMAP, it defines a neighborhood in terms of an absolute distance scale, or in terms of number of neighbors, where linearity is expected to be true. An embedding is determined by noting that the same geometric properties of neighborhood reconstruction should apply equally well to an embeded lower dimensional description up to an affine transformation (translation, rotation, and rescaling). Like the Dijkstra implementation of ISOMAP, the largest computational operations in LLE also scale as $O(N^2 \log(N))$; however, LLE is not guaranteed to discover the optimal global coordinate system and appears to be more vulnerable to noise (see response of Tenenbaum, de Silva, and Langford in [8]). The fastest available approach for estimating manifold coordinates is stochastic proximity embedding (SPE) [1], [2], [33] because its fundamental computational burden scales as $O(N)$; however, the simplifying assumptions that appear in estimating geodesic distances and the embedded approximation appear to be too weak and have lead to degenerate solutions when applied to hyperspectral data and even in some very simple, artificial problems.

In order to derive manifold coordinate systems for hyperspectral imagery on a practical scale, where the number of pixels in a typical flightline in our study area is $O(10^6)$ or greater, we adopt a divide, conquer, and merge strategy. Specifically, we first divide the scene into a set of nonoverlapping tiles where ISOMAP can derive an optimal coordinate system in a relatively short time with modest memory requirements. The global scene manifold is then obtained by merging the tile manifolds. We discuss several methods of aligning manifolds below. We also examine the relative merits of ISOMAP for both compression and classification. The remainder of our paper is organized as follows. In Section II, we first describe practical examples that motivate us to look for manifold representations and then describe the specific manifold algorithms on which our approach is based. In Section III, we demonstrate that ISOMAP provides a significantly more compact description of hyperspectral data than the minimum noise fraction (MNF) [23] transform, a widely used linear method for noise reduction and compression of hyperspectral data. Section IV shows that ISOMAP provides better discrimination than one of the standard linear mixing models [13]–[15] for an example with spectrally similar classes. In Section V, we describe ways to scale these manifold algorithms to large remote sensing scenes by aligning manifolds. Section VI shows the results of these scaling efforts. In Section VII, we show one strategy for automating manifold alignment, and in Section VIII, we describe the potential for parallel implementations of the new approach. Finally, in Section IX, we summarize and draw conclusions.

II. Manifold Coordinate Systems and Hyperspectral Data

A. Motivation for a Manifold Coordinate Sytem

In the introduction, we noted that it is the inherently nonlinear characteristic of hyperspectral data that motivates us to consider methods for identifying a set of coordinates that parameterize the data manifold. We cited a number of sources of these nonlinearities: BRDF, nonlinear media such as water, multiple scattering, and the heterogeneity of pixels. In a hyperspectral scene containing natural vegetation, the nonlinear characteristics are immediately apparent in many three-channel scatterplots. Fig. 1 illustrates this point, showing three-channel scatter plots from two subsets of a 124-channel PROBE2 hyperspectral scene described in greater detail in [5] and [7]. Choosing any three channels will provide a different view of the high-dimensional data manifold. The goal is to derive a coordinate system that resides on (parameterizes) the data manifold itself, following its intricate and convoluted structure with the hope of achieving a better data representation for classification and/or compression purposes. Fig. 1 also illustates this. In particular, we replace a linear notion of similarity, in which simple linear distance (e.g., linear distance from an endmember) describes distances between all pairs of sample points, with the concept of geodesic distance, in which distance is measured by tracing the trajectory of the data manifold. This is equivalent to finding a coordinate system that resides on the manifold and measuring similarity by estimating geodesic distance along the manifold itself. A corollary to this is that if we can discover such a manifold coordinate system, then the geodesic distance along the manifold is just the linear distance in the manifold coordinate system. In what follows, we provide examples using both spectral angle and Euclidean distance for the local metric, while past expositions of these algorithms have typically chosen only Euclidean distance as the input distance metric. Nevertheless, the same arguments apply. When estimating geodesic distances, the distance metric is only applied in a small, locally linear neighborhood. Distances to samples outside the local neighborhood of a particular sample are calculated by linking the shortest paths through points common to more than one neighborhood. Therefore, one other commonly used metric, the Mahalanobis distance, which is frequently applied in classification and detection problems in hyperspectral imagery analysis [16], cannot be applied in the context that is often used. Because the Mahalanobis distance implies the calculation of an associated scene covariance matrix, its use would require aggregation of samples that are not within the linear region of the manifold. The only way to apply it within the context of geodesic distance estimation would be to restrict the covariance matrix to those pixels lying within the locally linear neighborhood of a given pixel and producing a neighborhood covariance matrix for each point. While this is possible, it is certainly more computationally expensive. Since the distance metric is only evaluated locally, and far away geodesic distances are calculated using graph algorithms, the specific choice of local metric is less important. Regardless of the choice of local distance metric, the remainder of the algorithm as described here has the same form. Our purpose here is not to decide which is the best choice of metric for locally linear neighborhoods, but rather to present the general framework for estimating geodesic distances over an inherently nonlinear data manifold.

B. Manifolds: Formal Definitions

Either explicitly or implicitly, the most fundamental property that is usually exploited in manifold algorithms is that locally a manifold looks like a Euclidean space. This is essentially what

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

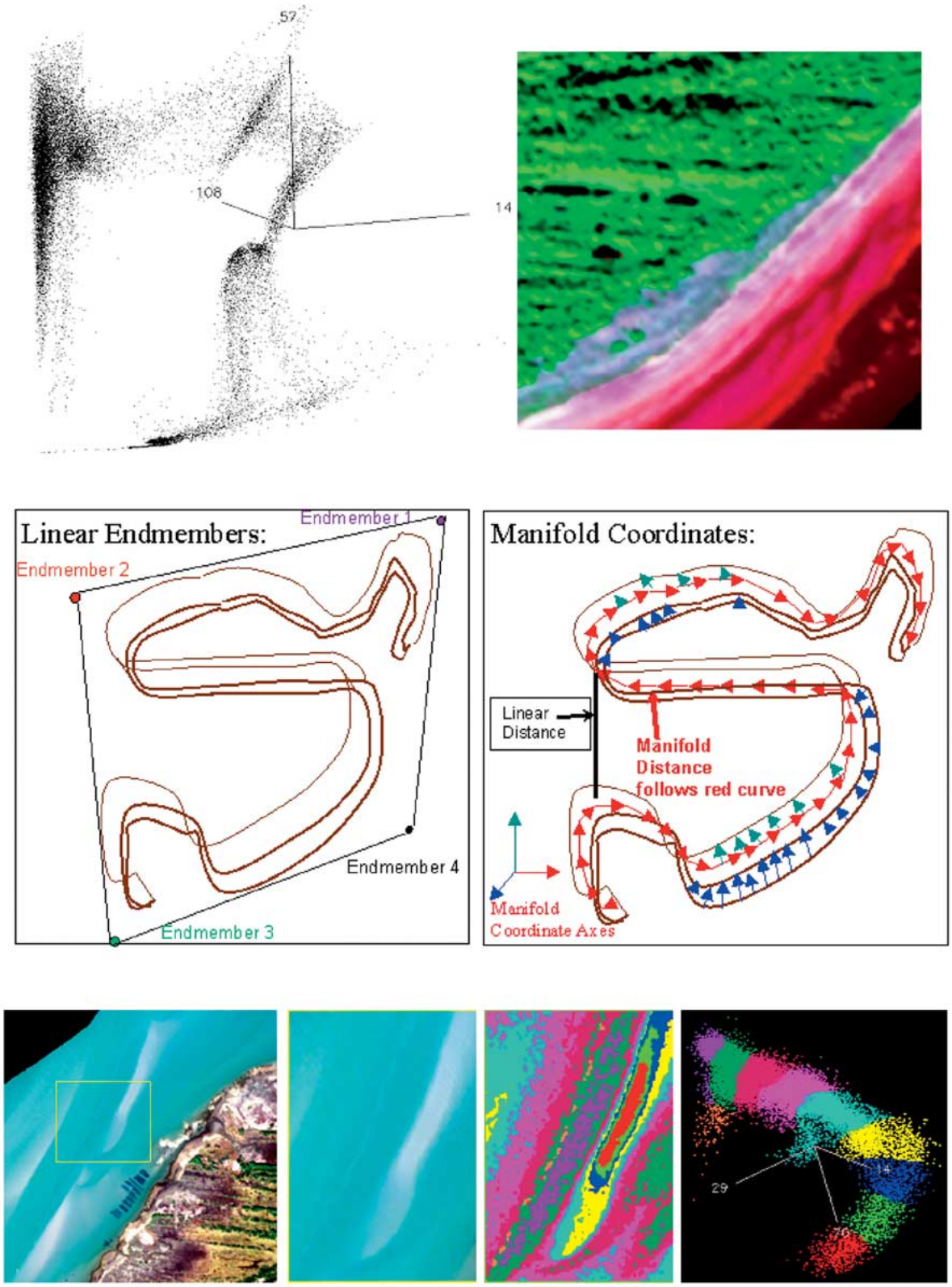

Fig. 1. (Top left) Scatter plot of channels 14, 57, and 108 (0.6278, 1.2862, 2.2165 μm) from (top right) a subset (uplands, marshes, dunes, and beach) of a 124-channel PROBE2 airborne hyperspectral flightline over Smith Island, VA, October 18, 2001. The scatterplot reveals the nonlinear structure of the data manifold. Note curves and significant gaps in the interior of the distribution. (Middle right) Conceptual view of a manifold coordinate system that resides on (parametrizes) the manifold, so that distance is measured along the trajectory of the manifold rather than the "short-circuit" linear distance typically used to measure distance and similarity. In contrast, also shown (middle left) are typical endmembers found by convex hull analysis, where linear proximity to each endmember determines percentage of each constituent in the model. (Bottom left) Lagoonal region in same scene with subtidal flat outlined (channels 15, 9, 3 shown: 0.6278, 0.5528, 0.4704 μm). (Bottom middle) Enlargement of the subtidal zone and color-coded profile corresponding to (bottom right) scatterplot of channels 14, 29, 70 (0.6278, 0.8572, 1.4962 μm) showing distribution with curvature. The color-coded arc probably corresponds to depth in subtidal flat.

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

is meant by the definition of a "d-dimensional topological manifold," where, formally, we require that for each point $\vec{x}$ in the manifold M, there is an open subset V with $\vec{x} \in V$, for which there exists a homeomorphic mapping g that maps V into a d-dimensional open subset V' in a d-dimensional Euclidean space $\Re^d$ [41]

$$V \xrightarrow{g} V'. \tag{1}$$

It is this open subset V that we identify as the "neighborhood" in algorithms included in this paper.

C. ISOMAP

Of the extant methods for estimating a global manifold coordinate system, ISOMAP [42] is the only one for which a globally optimal solution can be guaranteed. In the original ISOMAP, all Euclidean distances r_{ij} between pairs of data sample points $\vec{x}_i$ and $\vec{x}_j$ are calculated. Using the notation in [42], a connected graph of distances $d_G(i,j)$ is first determined. Whenever samples $\vec{x}_i$ and $\vec{x}_j$ lie within a neighborhood defined by an absolute distance scale ϵ (epsilon-ISOMAP) or by a set of K nearest neighbors (K-ISOMAP), $d_G(i,j) = r_{ij}$; initially, all other $d_G(i,j)$ are set to ∞. As previously discussed, while the original algorithm was introduced with a Euclidean distance metric, there is no reason in principle that other distance metrics cannot be applied. The distance metric could be any reasonable choice appropriate to the application at hand. In hyperspectral applications, we might choose the spectral angle, since this mitigates the effects of variable illumination [27]. Note that inside the neighborhood, the graph and manifold distances $d_M(i,j)$ are the same [$d_M(i,j) = d_G(i,j)$] using the fact that the manifold is assumed to be linear within a neighborhood. Once the graph has been initialized in this manner, the remaining estimated manifold distances $d_M(i,j)$ outside of the neighborhood are computed. An exhaustive iterative method known as Floyd's algorithm [28], which scales as $O(N^3)$, is optimal for dense graphs. For sparse initial graphs, often the situation in many practical applications, the scaling can be improved by using Dijkstra's algorithm [8]. The term sparse means that the number of edges E (number of edges initially $< \infty$) between pairs of samples is small compared to the square of the number of graph vertices V^2, which is also the square of the number of samples, N^2. The edges of the sparse initial graph are the distances to those points that lie within a neighborhood of each other. If the neighborhood is chosen to be small, a desirable feature to ensure local linearity, then the initial graph will be sparse and Dijkstra's algorithm is the better choice. The improvement in scaling can be quantified by considering a result found in [38]: for the all-pairs shortest path problem using Dijkstra's algorithm with a minimum priority queue implemented with a d-way heap, the processing time scales as $O(VE\log_d(V))$ with $d = E/V$, when $d > 2$.[1] Since, $V = N$, and $E = (m * N/2)$, where m is the mean number of neighbors,[2] we obtain the $O(N^2\log_d(N))$ scaling quoted earlier. Note that this is a worst case upper bound, and in practice, we have found that choosing a somewhat wider heap with $d > E/V$ may sometimes yield improved run times.

[1]For $E/V < 2$, $d = 2$, and the scaling is $O(V^2\log_2(V))$.

[2]The factor of two arises because each edge appears twice between each pair in the calculation of m.

In the Floyd implementation, the following update rule is used for all samples:

$$d_G(i,j) = \min_k(d_G(i,j), d_G(i,k) + d_G(k,j)). \tag{2}$$

The Dijkstra 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 [38], [39]. Because of the sorting that takes place in the queue, many unnecessary comparisons that are made in the Floyd algorithm are omitted by Dijkstra, and for sparse initial graphs, this provides the $O(N/\log_d(N))$ speed up in processing time. On the other hand, if the graph is initially dense, the sorting actually results in longer processing times for the Dijkstra algorithm compared to the Floyd algorithm. For the applications we consider here, the density of the graph is controlled by the neighborhood size set by the user (the ϵ in ϵ-ISOMAP or the number of neighbors K in K-ISOMAP), and in practice a sparse graph is obtained.

When all points have been exhausted, $d_G(i,j)$ will contain the shortest distances along the manifold between any two points $\vec{x}_i$ and $\vec{x}_j$ and will be the best estimate of the true manifold distances $d_M(i,j)$ for a particular choice of neighborhood size. Note that (2) and the initialization phase described above take advantage of the local properties of the neighborhood and the triangle inequality to determine the shortest path distances on the manifold. In [42], it is pointed out that one or more pockets of points may be isolated from the main distribution, in which case some distances remain infinite at the termination of the geodesic distance estimation. These disconnected points can either be omitted as outliers or attached in some self-consistent manner. In our implementation of ISOMAP, we handle disconnected pairs of points by looking for the closest disconnected pair of points (closest in the sense of the distance metric); the pair is reconnected to the graph using this distance, and then an attempt is made to attach the remaining set of disconnected pairs through the reconnected pair. This is self-consistent because an isolated pocket of points may have been fully connected within the pocket but not to the main distribution, so establishing the closest link to the main distribution allows known geodesic distances in the main distribution and in the pocket to be connected, while preserving the internal geodesic structure of each. If there are more than two disconnected pockets, the process is repeated until all points are reconnected. This process generates the minimum spanning tree that connects the isolated pockets of points. We use this approach because it ensures that no data are excluded from the analysis, and because it maintains the principle of minimum path geodesic distance between pockets of disconnected points.

Once the graph has been determined, the new coordinate system is determined by applying a multidimensional scaling algorithm to determine eigenvectors in the embedded linear space characterized by d_G. Specifically, let $\vec{v}_i$ and λ_i be the ith eigenvector and eigenvalue of the transformed distance graph matrix

$$\tau = -\frac{1}{2}H^T S H \tag{3}$$

where S is defined by

$$S_{ij} = ((d_G)_{ij})^2 \tag{4}$$

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

and H is a "centering matrix," given by

$$H_{ij} = \delta_{ij} - \frac{1}{N} \tag{5}$$

where δ_{ij} is the Kronecker delta function. Then, the matrix τ measures the second-order variation in the geodesic distance between a pairs of sample vectors, and the final embedded manifold coordinates are given by

$$\vec{M}_i = \sqrt{\lambda_i}\vec{v}_i. \tag{6}$$

Note that the eigenvectors $\vec{v}_i$ diagonalize the $N \times N$ matrix τ, so these vectors are N-dimensional, where N is the number of pixels in the dataset, and there are at most N such vectors. Typically, however, we will be interested in a substantially smaller number of components. From the eigenspectrum of the matrix τ, we will find that the intrinsic manifold dimension of our hyperspectral datasets is substantially smaller than the input dimensionality (number of samples N). Iterative eigensolvers are appropriate for large matrices, when only a small subset of the eigenvectors is actually required. We use the Jacobi–Davidson QR (JDQR) [9], [40] iterative eigensolver because it is considered to be among the more reliable methods.

D. Local Linear Embedding

Local linear embedding (LLE) was first proposed in [35]. The basic idea is to construct a locally linear model for each sample vector in terms of its neighbors. On the data manifold, any sufficiently small region will be locally linear, so we can expect to be able to reconstruct any point within a neighborhood from the set of other points within that neighborhood. For sample vector X_i (using the notation found in [35]), the reconstruction of X_i is a simple linear combination of neighboring points

$$X_i = \sum_{\text{neighbors } j} W_{ij} X_j \tag{7}$$

where

$$\sum_{\text{neighbors } j} W_{ij} = 1 \quad W_{kl} = 0 \tag{8}$$

if k and l are not neighbors. Roweis and Saul note that this form is invariant to any neighborhood-preserving transformation accomplished by rotation, translation, or rescaling. The fundamental concept is that if the manifold is lower dimensional than the original data space, then this same reconstruction should hold within the neighborhood, either in the original coordinates or the reduced manifold coordinate system. Thus, if Y_i is the manifold coordinate representation for sample vector i, then we also have

$$Y_i = \sum_{\text{neighbors } j} W_{ij} Y_j. \tag{9}$$

Once the weights in (7) are found by a least squares minimization proceedure, these same weights are used to solve for the manifold coordinates Y in (9). Additional constraints are added that ensure that the covariance matrix of the embedding vectors Y_i is normalized to unity to prevent degeneracies and that the manifold coordinates are centered about the origin. The reconstuction error in the manifold coordinate system, which is minimized to obtain the coordinates Y_i, has the form

$$\xi(Y) = \sum_i \left| Y_i - \sum_{\text{neighbors},j} W_{ij} Y_j \right|^2. \tag{10}$$

With the additional constraints, this can be written as

$$\xi(Y) = \sum_{ij} M_{ij} (Y_i^T Y_j)$$
$$M_{ij} = \delta_{ij} - W_{ij} - W_{ji} + \sum_k W_{ki} W_{kj}. \tag{11}$$

Like the Dijkstra implementation of ISOMAP, the scaling of LLE computationally is only $O(N^2 \log(N))$, but this is still large when N is the number of pixels in, for example, a large hyperspectral scene.

E. Summary

Of the main approaches to modeling manifolds, the only approach that yields a globally optimal solution is ISOMAP. Nevertheless, since its dominant computational burden scales as $O(N^2 \log(N))$ and memory requirement scales as $O(N^2)$, it is not a practical solution in its present form for typical remote sensing data, only for very small subsets. In Section V, we develop a new approach based on ISOMAP that is scalable to large remote sensing scenes. As a prelude, in Section III we compare ISOMAP with MNF, a widely used standard for removing noise and compressing data in hyperspectral datasets [23], and in Section IV, we compare the relative effectiveness of ISOMAP and one of the standard linear mixing models in distinguishing spectrally similar classes.

III. Results for Hyperspectral Data Compression: Manifold Coordinates Versus MNF

In [42], improved data compression of ISOMAP over traditional linear approaches based on principal component analysis (PCA) and multidimensional scaling (MDS) was demonstrated for several applications, including hand-written character recognition and face recognition under variable pose and illumination conditions. The principal reason for improved compression is that, by deriving an embedded coordinate system, the underlying nonlinear characteristics of the data manifold are more naturally and compactly represented by ISOMAP than by linear approaches. We have obtained similar results for hyperspectral data as we now illustrate.

MNF [23] has become a de-facto standard for reduction of noise and redundancy in hyperspectral and other remote sensing data sources and appears as a standard option in commercial global information system packages. MNF, a linear transform method, calculates a set of ordered principal axes, which are ranked by signal-to-noise ratio (image quality); it can be viewed as a noise-whitened PCA analysis. In Fig. 2, we compare eigenspectra normalized by the first eigenvalue λ_i/λ_0 for three image subsets, contrasting MNF with ISOMAP. The examples in the figure are derived for: [Fig. 2(a)] a vegetated land-cover tile (an upland and marsh zone) from the PROBE2 scene previously

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

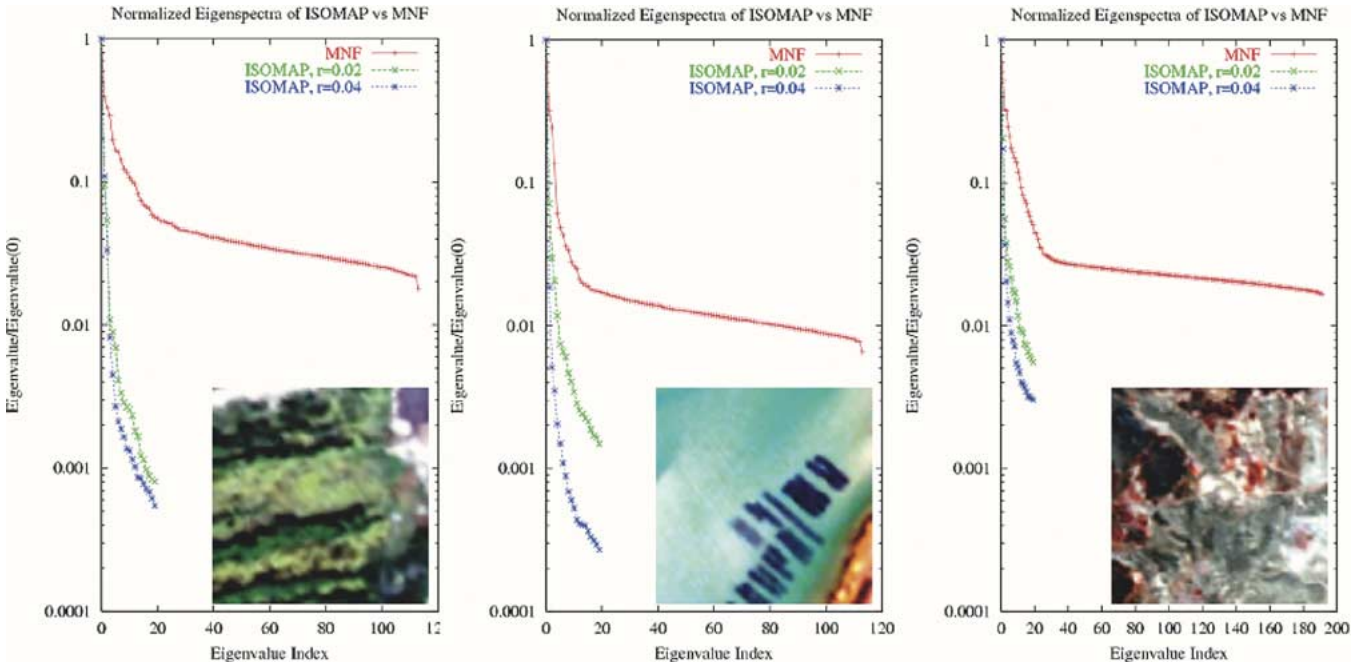

Fig. 2. Normalized eigenspectra (log scale) λ_i/λ_0 comparing classical MNF approach to ISOMAP manifold coordinates for two different neighborhood radii of 0.02 and 0.04 rad (neighborhood distance metric was spectral angle). ISOMAP curves show dramatically larger compression rates for (left) a tile with upland and marsh land-cover and (middle) a turbid water tile with submerged clam nets (middle) from PROBE2 scene with 124 spectral channels (114 channels used in analysis). (Right) The same comparison for a tile from a 1997 AVIRIS scene of Cuprite, NV (using 192 of 224 spectral channels in the analysis). (Photo insets) RGB of channels 15–9-3 (0.646, 0.553, 0.460 μm) for PROBE2 and channels 28–19-10 (0.636, 0.547, 0.458 μm) for AVIRIS.

described, [Fig. 2(b)] a shallow water region containing suspended clam nets from the same scene, and [Fig. 2(c)] a nonvegetated region in a 1997 Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) scene of Cuprite, NV. In all cases, the eigenspectra show that the vast majority of the variation in the data is captured in the first 5–15 ISOMAP components, while to achieve the same level of accuracy, the vast majority of the MNF components must be retained.

IV. Manifold Representations of Hyperspectral Scene Tiles

In this section, we turn our attention to the effectiveness of the manifold coordinate representation in separating spectrally similar classes and compare it with one of the standard linear mixing models [13]–[15], [32]. The example that we use here is the separation of spectrally similar vegetation in coastal vegetation mapping; however, as the previous section on compression demonstrated, the range of potential applications is diverse.

In earlier papers, we have described different approaches to classifying barrier island land-cover from airborne hyperspectral data emphasizing machine learning approaches and multiclassifier fusion concepts [5]–[7] on a large scale. One source of errors in these models stems from the confusion of spectrally similar vegetation such as *Phragmites australis*, an invasive wetland reed, and *Scirpus spp.*, a sedge. Although the two plants are not similar morphologically, during the fall when the PROBE2 scene described in Fig. 1 was acquired, the brown tassle of *Phragmites* and brown seed pods of *Scirpus* combine with the stalk vegetation and partially exposed subcanopy to produce a spectrally similar return (Fig. 3). A conjecture was that this might be resolved using a more traditional linear mixing approach, but this has not proved to be the case. We compared the ISOMAP algorithm with linear mixing for this problem. In Fig. 3, we show a 100×100 pixel section of the scene with superimposed differential GPS (DGPS) ground truth survey data for known locations of *Phragmites* and *Scirpus*. In the same figure, we show the result of classifying the second, third, and fourth components of both a linear mixing approach (ORASIS [13]–[15], [32]) and ISOMAP. Note that the term "component" in the case of linear mixing refers to an ordering of the demixed endmembers in terms of the degree of data variability in the direction of the endmember. We examine these components because the compression results of Section III suggest that the manifold representation may be more compact than linear methods, so we look for separation of spectrally similar classes in some of the most significant components. The classification proceeded using spectra at known locations from the DGPS survey as seeds. In each representation, we proceeded conservatively, choosing only the areas where the category seeds were isolated from each other in the linear mixture or ISOMAP representations. Adjacent points were then assigned to the class of the closest seed. The resulting classification is shown in Fig. 3. For both the ORASIS and ISOMAP representation, we chose the set of three components in which the best separation between *Phragmites* and *Scirpus* data samples was observed. The results show that adjacent points in the ISOMAP representation map directly to other points of the same category; however, the linear mixing case produces false alarms in the backdune, in a swale entirely dominated by *Scirpus* in the top of the 100×100 subset, and in the bottom left corner of the subset near one of the seed zones. We emphasize that although both ORASIS and ISOMAP scatter plots had regions where the ground truth spectra partially overlapped, the classification proceeded from seeds where there was

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

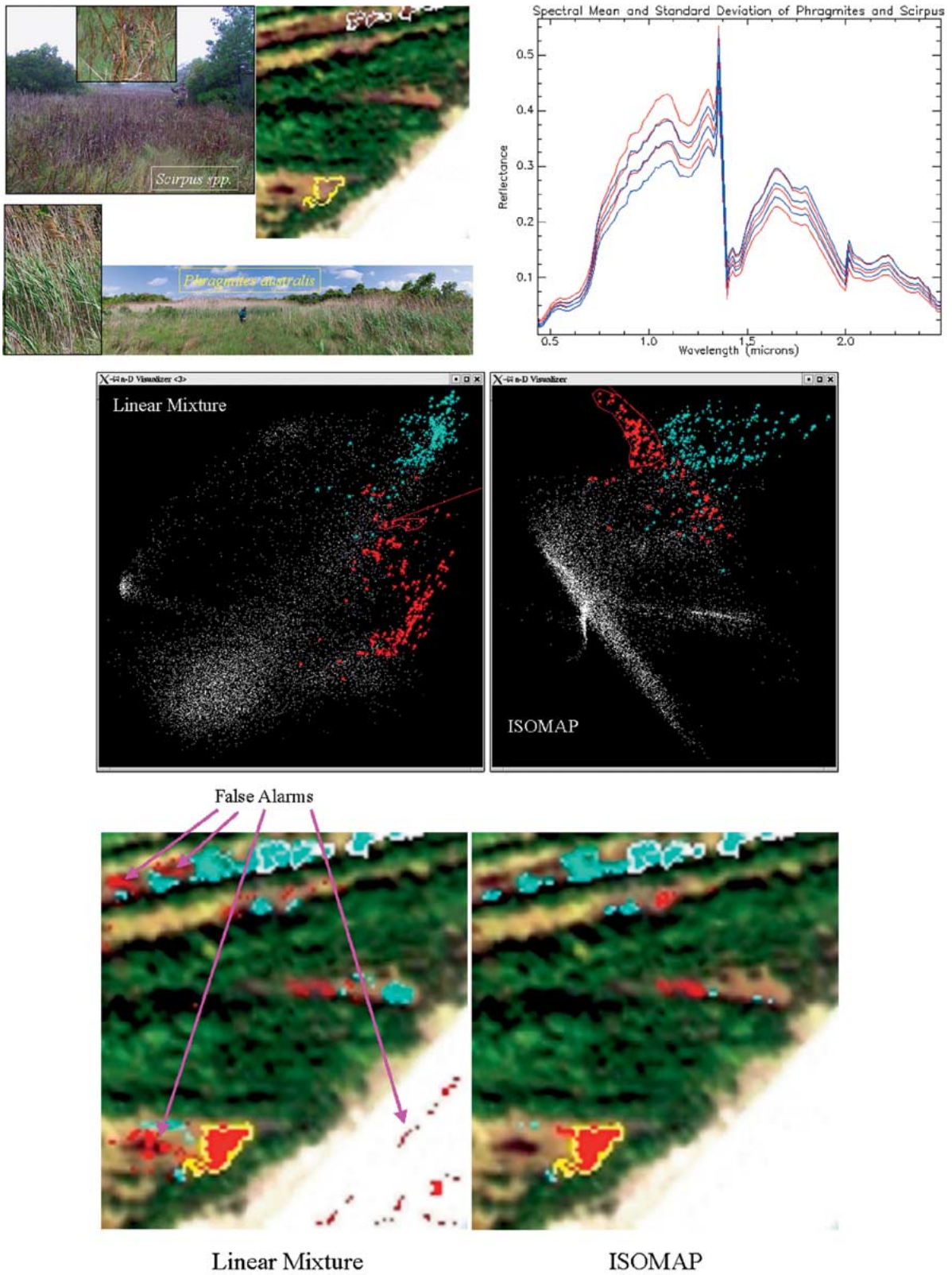

Fig. 3. Comparison of the ORASIS linear mixture model and ISOMAP for spectrally similar vegetation classes *Phragmites australis* and *Scirpus spp.* for representative hyperspectral scene tile with (top left) DGPS survey zones of (yellow lines) *Phragmites australis* and (white lines) *Scirpus spp.* superimposed. (Top right) Mean and mean plus or minus standard deviation curves for one of the seed regions for each class: (red) *Phragmites* and (blue) *Scirpus*. (Middle row) Scatterplots of components 2-3-4 for each representation with (red) *Phragmites* seed regions and (cyan) *Scirpus* seed regions. These zones were used as seed samples for classification of the linear mixture representation and an ISOMAP coordinate representation of the tile. Using the DGPS circumscribed pixels as seeds, the representations were classified only in areas where the two distributions were disjoint. Shown: intermediate stage of classification with outlines indicating subregions where points will be assigned to enclosed seed category. (Bottom row) The results were redisplayed on the original tile, showing three distinct regions with *Phragmites* false alarms generated by the linear mixing model in the upper left corner of the image in a *Scirpus*-dominated swale, in the backdune and beach (lower right corner of image), and adjacent to the seed *Phragmites* pixels in the lower left. ISOMAP eliminates these errors.

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

TABLE I
SCALING OF ϵ-ISOMAP, PENTIUM IV PROCESSOR, 3.06 GHz

Graph Alg.	Spatial Dim. (V=N)	NxN Geodesic Memory Req.	ϵ	Avg. No. Neighbors	d=$\frac{E}{V}$	Time (minutes):Geodesic Graph Only	Total Time (minutes)
Memory Requiremtns and Total Processing Time to Extract First 10 Manifold Coordinates							
Typical Land Tile, PROBE2 Hyperspectral Data, Using 114 of 124 spectral channels							
Dijkstra	50x50	0.050 GB	0.02	150.4	75	0.3	0.53
Dijkstra	75x75	0.253 GB	0.015	59	30	0.9	3.2
Dijkstra	75x75	0.253 GB	0.02	168.2	84	0.67	3.8
Dijkstra	75x75	0.253 GB	0.03	489.9	290*	3.61	5.32
Dijkstra	100x100	0.8 GB	0.02	302.9	155*	8.23	15.37
Floyd	50x50	0.050 GB	0.02			1.77	2.02
Floyd	75x75	0.253 GB	0.02			22.73	25.13
Typical non-vegetated tile, AVIRIS, Cuprite 1997, Using 192 of 224 channels							
Dijkstra	75x75	0.253GB	0.015	8.8	5	0.53	2.78
Dijkstra	75x75	0.253GB	0.02	48.9	24	1.17	3.15
Dijkstra	75x75	0.253GB	0.03	333.8	167	2.85	5.07
* Indicates, d chosen $> \frac{E}{V}$.							

no overlap between the two categories. Thus, the false alarms in classifying the ORASIS representation occur because adjacent pixels from other land-cover types are immediately adjacent to the seed pixels, while in the ISOMAP representation, the pixels immediately adjacent belonged to the same class. Therefore, the ISOMAP classification did not have the same problem with false alarms. This implies that nonlinear correlations between spectral channels exist and are better separated by a representation that directly parameterizes these nonlinearities. While this example can hardly be deemed a complete proof that one representation is better than another, solving a previously unsolved problem on a small scale suggests that the goal of scaling a manifold representation up to the level of the scene is worthy of investigation. In Section V, we turn our attention to the problem of scaling the manifold representation to scales that are practical for remote sensing applications.

V. SCALING TO LARGE-SCALE REMOTE SENSING DATASETS

It is scaling that prevents the direct application of the globally optimal ISOMAP algorithm to large remote sensing datasets. We want the optimal coordinates of ISOMAP but not its computational complexity and especially not its memory requirements, which make direct application impractical for scenes of even modest size. Table I gives some typical times and memory requirements for the dominant calculations in the ISOMAP approach for two hyperspectral tiles decribed in Section III, the vegetated land tile derived from the PROBE2 scene, and the nonvegetated tile from the 1997 AVIRIS Cuprite scene. The strategy that we adopt here is one of dividing, conquering, and merging. ISOMAP can be applied readily to small tiles within a scene to achieve a globally optimal coordinate system within each tile. The challenge, however, is to derive a globally consistent manifold coordinate system, so that manifold distances from points in one tile to points in another disjoint tile will be accurate.

One approach would be to augment each tile with a small common set of data tie-points. These tie-points would then allow the manifold from one tile to be transformed into the manifold coordinate system of any other tile. Let M_i be the matrix of coordinates of the tie-points in the ith tile manifold, and let M_j be their coordinates in the jth tile manifold. Then, the coordinate transformation from the jth tile to the

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

ith tile coordinate system can be rapidly estimated using the pseudoinverse [25]

$$P = (M_i^T M_i)^{-1} M_i^T M_j. \tag{12}$$

Note that the size of both matrices is $N_c \times p$, where p is the dimensionality of the derived manifold coordinate system. If an offset is allowed, then the matrices are $N_c \times (p+1)$.

A second approach might be to define a small decimated data cube from the original scene by random sampling. The hope is that such an approach would provide enough samples that represent the global space of possible spectral samples, so that the tile manifolds could be spliced onto this "backbone." By design, we can insist that the backbone input data cube be formed by obtaining a fixed number of random samples s from each tile. The common points then act as guideposts that will allow a change of coordinates from the local tile manifold to the globally coordinated backbone manifold. Let M_i be the coordinates of the ith tile manifold, and let B be the coordinates of the same points in the backbone manifold. Then, the coordinate transformation from the tile manifold coordinate system to the backbone manifold coordinate system, can be rapidly estimated using the pseudoinverse as before

$$P_B = (M_i^T M_i)^{-1} M_i^T B. \tag{13}$$

In practice, we find that sampling is the limiting factor for the "backbone" approach. For sparse samples, it is necessary to choose large neighborhood radii in order to get a consistent backbone manifold; with such a manifold, when the tiles are warped to the backbone coordinate system, the local features found in tile manifolds tend to be blurred in the first coordinates, and lower coordinates may be lost. The only way to overcome these limitations is to ensure that the backbone has a large enough sample space. However, we want to avoid the scaling problems associated with deriving manifold coordinates, so this limits the number of samples that we can allow, and therefore the number of tiles from which we can hope to derive a sufficient sample size. In practice, this does not preclude us from using a backbone concept to align the manifolds of a smaller number of tiles on a more local basis, but practical application over a large scene could be computationally expensive.

A third alternative strategy returns us to the fundamental concept behind the manifold, namely that on the manifold the data is locally linear. Indeed, we saw LLE took advantage of this local linearity in both the manifold coordinate system and in the original data space, to construct local linear embedding transformations that preserve neighbood geometry. Now, suppose that we have optimized manifold coordinates for tile T_j using ISOMAP and that we extract a random fraction f_i of samples from an adjacent tile T_i that has also been optimized by ISOMAP. For sample n from tile T_i, let m_{ni} be the manifold coordinate and s_{ni} the coordinates in the full spectral space (the original spectrum). Presuming that adjacent tiles have enough data of similar type, we can expect that a good fraction of these points will have neighbors in the full spectral space. We search for K neighbors of s_{ni} in the full spectral space of tile T_j. Let these K neighbors of s_{ni} in the full spectral space of tile T_j be denoted by s_{njk}. Whenever we encounter less than $K_{\min}$ neighbors in the full spectral space, we reject this point for later use in the mapping that we want to construct from the manifold coordinates of tile T_i to tile T_j. For

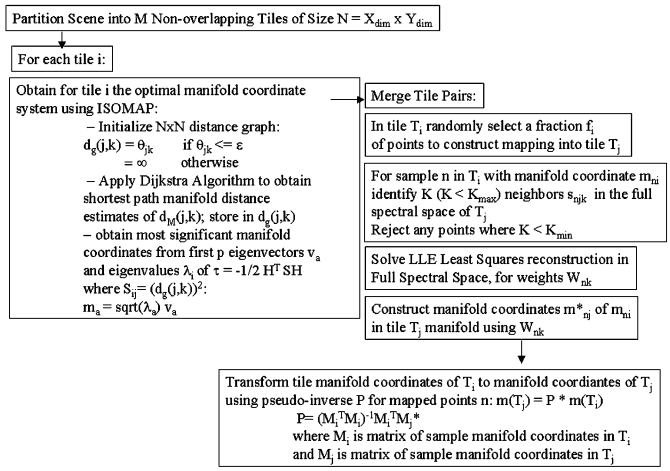

Fig. 4. Hybrid approach to merging manifold coordinates. Tiles derived from the scene are optimized separately using ISOMAP. Merging of a pair of tiles T_i and T_j uses the local reconstruction properties of LLE to map a selected sample of points from one manifold coordinate system to the other. A pseudoinverse is then constructed to map the entire manifold from one tile to the other.

each point with a sufficient number of neighbors, we now solve for the least squares reconstruction weights W_{nk} of s_{ni} in the full spectral space of tile T_j using the neighboring samples s_{njk}, as was done in the LLE algorithm described earlier. As we saw earlier, because LLE is a neighborhood-preserving transformation, we can now construct the manifold coordinates of m_{ni} in tile T_j, which we denote by m^*_{nj}, using the weights W_{kn} derived for the neighbors in the full spectral space. Once we have done this for a sufficient sample of points, we can then construct a transformation matrix from the manifold coordinates of tile T_i, $m(T_i)$ to those of tile T_j, $m(T_j)$

$$m^* = m(T_i) = P * m(T_j) \tag{14}$$

again using the pseudoinverse P

$$P = (M_i^T M_i)^{-1} M_i^T M_j^* \tag{15}$$

where M_i is the matrix of manifold coordinate samples m_{ni} from tile T_i, and M_j^* is the corresponding set of coordinates constructed in the manifold coordinate system of tile T_j using LLE insertion. These steps are summarized in Fig. 4.

VI. Results: Aligning Tile Manifolds

A 124-channel PROBE2 airborne hyperspectral scene of Smith Island, VA acquired on October 18, 2001 [5], [7] was partitioned into a set of 75×75 pixel tiles. For each tile, ISOMAP was used to derive an optimal tile manifold coordinate system. All three approaches to merging tile manifold coordinates were evaluated using this data.

The advantage of the approach that uses an LLE reconstruction of a limited set of points and then the pseudoinverse is that it scales as $O(n_{f_i}^2)$, where n_{f_i} is the number of points needed to construct the mapping from one tile to an adjacent one, and this is typically a small number. In the tests that we performed with 75×75 tiles, we were successful in merging tiles using 300 pixels in the mapping. In contrast, the backbone approach requires significantly more samples to work because, in effect, we must create an entirely new representative manifold. For the backbone approach, we found that nearly 50% of the samples

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

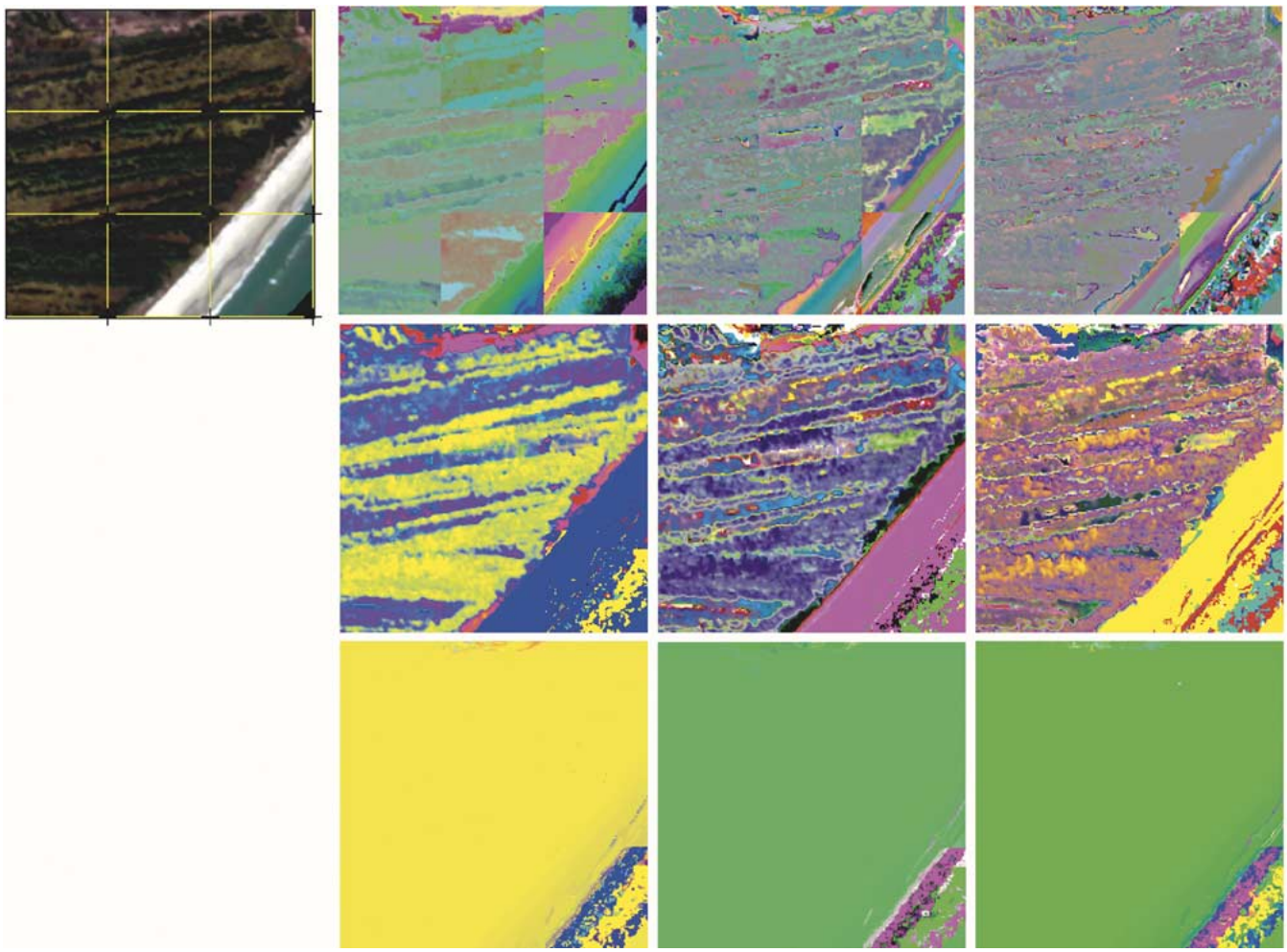

Fig. 5. Merging manifold coordinates. (Upper left) RGB subset derived from original PROBE2 hyperspectral scene, Smith Island, VA, showing nine adjacent tiles. (Upper right, top row) RGB of manifold coordinates derived for each tile using ISOMAP prior to alignment, from left to right: coordinates 1-2-3, 4-5-6, 7-8-9. (Middle row) RGB with stretch, showing same manifold coordinates after alignment using the pseudoinverse LLE mapping. (Bottom row) Same without stretch highlighting surf zone.

from each tile in a two-tile merge were required to ensure a merge with a minimum of discontinuities. Because the backbone approach is sample-limited, and optimization of the backbone manifold scales the same way that the tiles do computationally, this makes it less attractive for merging a large number of tiles. The tie-point approach suffered from similar scaling issues. If a sparse set of tie-points samples are employed, this tends to create isolated pockets in the distance graph, which ultimately led to poor matching. Overcoming this requires a large number of tie-points, which leads to long computational times.

Fig. 5 shows a set of nine tiles derived from the hyperspectral flightline. For each of these nine tiles, a manifold coordinate system was derived using ISOMAP. As the top row of the figure shows, the original tile manifolds show a rich structure, but the coordinate systems are not aligned. When the pseudoinverse LLE processing is applied to these using roughly 300 points, an excellent fit of the coordinate systems results. In this example, all nine tiles ultimately were aligned to the tile in the upper right corner. This proceeded in two stages. If the tiles are numbered by consecutive rows, then, in the first stage, the sixth and ninth tiles were each converted to the manifold coordinate system of the eighth tile. These are the only tiles that show any beach, dune, or surf. Then these two transformed tiles and all the remaining tiles were aligned to the the manifold coordinate system of the tile in the upper right corner. The logic behind this relates to the types of land-cover found in specific tiles. In order for the pseudoinverse LLE processing to be able to merge tiles pairwise, the fraction of samples extracted from one tile in the pair must have a sufficient number of neighbors in the original full spectral space in order to construct the mapping. Note that in the case of the ninth tile, there were not sufficient samples found to construct a mapping that aligned the pixels in the surf zone, which were not represented in the other tiles to any great degree. For this reason, the aligned manifolds are shown with two different stretch factors because these points could not be mapped in a continuous manner and a larger number of pixels were rejected by the pseudoinverse LLE processing during the construction of the mapping. The surf zone is a distinct class that is not spectrally close to any other class in the set of nine tiles. This large distance is preserved when the tiles are merged. The surf zone appears as an anomaly with respect to data in the other tiles. The rejection of a large number of pixels by the pseudoinverse LLE processing also suggests how we can go about constructing a metastrategy for merging individual tiles. When a large rejection rate occurs, an automated version of this processing is then instructed to search for other neighboring tiles

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

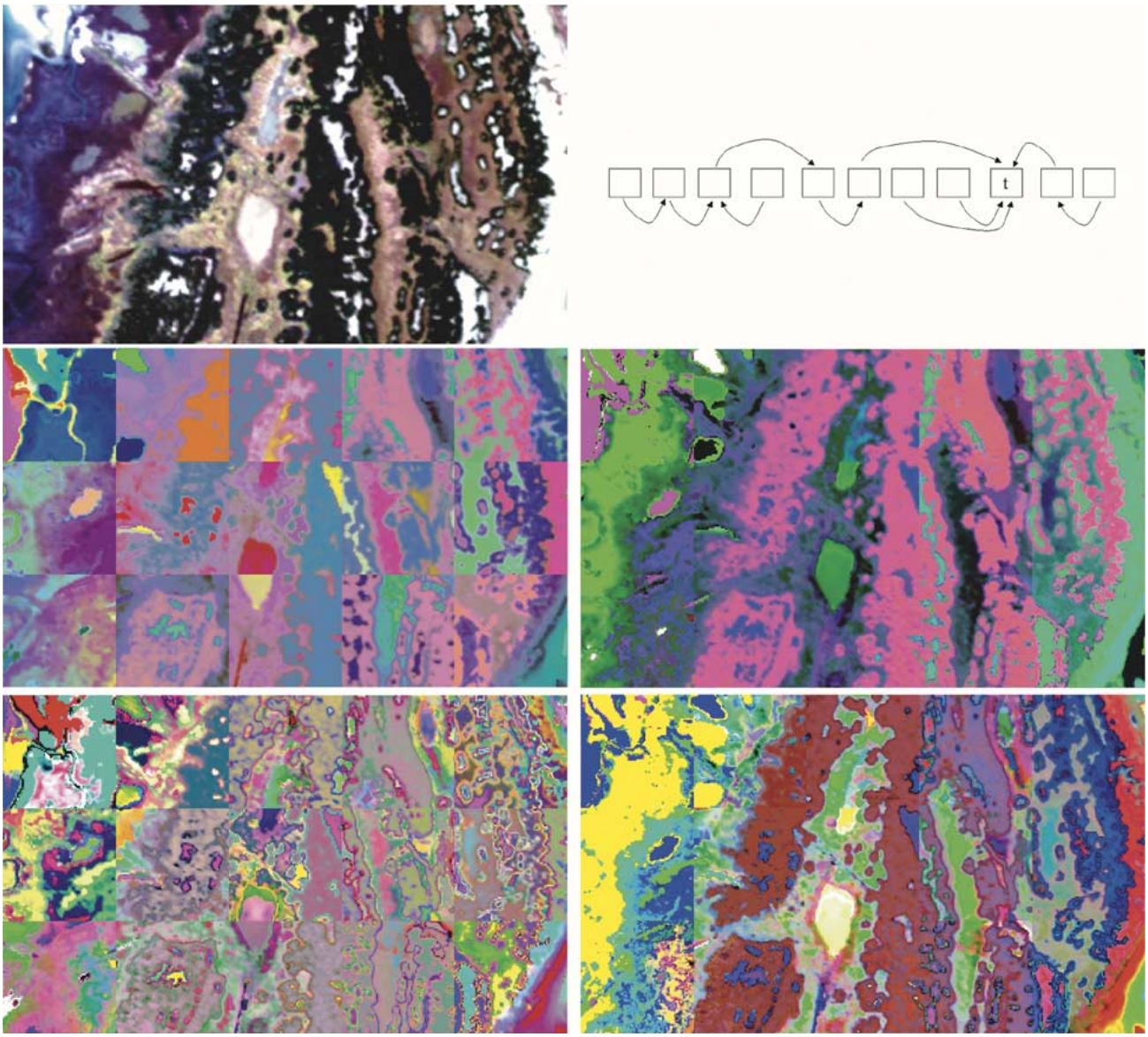

Fig. 6. Automated alignment of manifold coordinates. (Upper left) Subset derived from PROBE2 hyperspectral scene of Hog Island, VA, from October 18, 2001, showing a cross section of the island with salt marsh, uplands, brackish and fresh water marshes, dune grasslands, and beach. The subset was partitioned into fifteen 75×75 tiles. (Upper right, top row) Conceptual view of a series of alignment steps made between tile manifold coordinates. The iterative algorithm proceeds by attempting to align the leading edge tile of each graph (series of connected alignments) to the target tile, labeled "t." (Middle and bottom rows) ISOMAP coordinates before (left column) and after alignment (right column) for coordinates 1-2-3 (middle row) and coordinates 4-5-6 (bottom row).

in the immediate vicinity where a more reasonable number of points could be used to merge. Conversely, if a merge is allowed, the discontinuity in the manifold mapping allows for the mapping of anomalies in a scene.

VII. Automating the Alignment of Tile Manifolds

We focus on one possible strategy, showing its strengths and weaknesses and how improvements can be made. The basic idea is to look for a series of alignment transformations that take each tile manifold coordinate system to that of a target tile. If a tile differs significantly from the target tile, a series of alignment steps must be made between intermediate tiles that share common properties. We first look for a direct mapping to the target tile. If the tile point reconstruction does not pass thresholds for valid number of neighbors K in the target tile and average reconstruction error, then we look for the best match in a backbone directory of random points taken from each tile. The same kinds of criteria are applied at this stage, except that we allow for the possibility of expanding the matching neighborhood radius between the tile and potential intermediate tiles to ensure at least one valid "hop." The procedure records each hop for each source tile. As the procedure continues, a series of chains, some of which ultimately link to each other, is formed with various pathways to the target tile. These are a series of linear transformations described in (14) and (15). Fig. 6 shows a conceptual view of this alignment procedure.

Note in Fig. 6 that the majority of the scene is aligned quite well. There are some small discontinuities at the interface between a few of the tiles. This occurs primarily because along the manifold alignment trajectory for any particular tile, the tile may not encounter a sufficient number of constraints to the manifold coordinate orientation and scales for all possible land-cover types. It also suggests strategies for improving the approach de-

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

scribed in Fig. 6 by enforcing manifold continuity constraints between adjacent tiles. Another alternative is also discussed in Section VIII.

VIII. Computational Scaling and Optimization

Subdividing a large hyperspectral flightline into a set of smaller tiles permits application of (suboptimal) nonlinear dimensionality techniques to the entire image. This paper is driven by the desire to process large remote sensing images and, simultaneously, overcome memory and computational scaling obstacles. Precise values for the memory requirements and computation times of the present algorithm are provided in Table I. Here, we present a simple scaling analysis of the basic computation and memory requirements.

Consider as an example an N pixel problem divided into G tiles, each containing N/G pixels. The manifold computation for each tile scales as $O((N/G)^2 \log(N/G))$ and requires memory of $O((N/G)^2)$ bytes. The size of the calculated distance graph determines the overall memory requirement. Once all of the individual tiles are processed, splicing them together scales computationally as $O(G^2(N/G))$, where N/G is the number of pixels per tile, and in the worst case, $O(G^2)$ pairs of tiles are tested to obtain a best-fit final result. Minimizing the total run time and dropping constants and slowly varying log terms yields $N \sim G^3$. Using a typical value of $N = 10^6$ yields $G \approx 10^2$. Therefore, an optimal tile size is $N/G \approx 10^4 \approx 100^2$. As seen in Table I, a 1-GB memory restriction limits the maximum tile size N/G to a square tile approximately 110 pixels on a side. In this example, the optimal tile size happens to nicely balance typical memory and runtime constraints.

A simple parallelization involves a distributed architecture, wherein each tile is processed by a separate processor. Using the example above, this requires about 100 processors each with access to ~ 1 GB of memory. The runtime to complete all of the individual tiles is $O((N/G)^2 \log(N/G))$, and to splice the tiles to one another is, in the worst case, $O(G^2(N/G))$. Assuming the optimal number of tiles is chosen, the total parallel runtime scales roughly as $O(N^{4/3} \log(N))$. Again, only the leading term is kept. Employing the same image size as before ($N = 10^6$) implies that the tiling procedure improves total runtime by $O(N^{2/3}) \approx 10^4$ as compared to the entire-image-at-once calculation. Similarly, the memory requirements of the tiling approach are reduced by a factor of $G^2 \approx 10^4$. Thus, subdividing large images makes these calculations feasible for large hyperspectral scenes.

The caveats to the tiling approach and to the optimization calculations involve sampling issues. First, the tile size N/G must be large enough to accurately represent the manifold of the local region, and, second, each tile must sample enough of the global manifold to represent all of the features of the full scene. At present, the latter restriction appears more critical. Features that are spatially correlated and appear in only one or two tiles may not map nicely when the individual tiles are spliced together. This is the source of some of the minor spatial discontinuities in Fig. 6. A method that better samples the global manifold involves subdividing the image by decimating rather than tiling. However, the tiling approach employed here with well-defined tile boundaries permits an easy, visual analysis of the continuity of the manifold results across tile boundaries.

IX. Conclusion

We have motivated the need for a nonlinear coordinate description of hyperspectral remote sensing data, citing a number of sources of nonlinearity such as subpixel heterogeneity and multiple scattering, BRDF effects, and the presence of nonlinear media such as water. The direct result of these is a fundamental limit on the ability to discriminate, for instance, spectrally similar vegetation such as *Phragmites australis* and *Scirpus spp.* when a linear spectral coordinate system is assumed. Providing examples from water and both vegetated and nonvegetated land imagery, we demonstrated that the manifold coordinate representation provides a more compact representation of hyperspectral data than MNF, a standard linear transform used in the analysis of hyperspectral imagery. On a small dataset derived from an airborne hyperspectral scene, we compared ISOMAP with one of the standard linear mixing models and showed the potential of ISOMAP to improve the separability of spectrally similar vegetation that typically causes false alarms when a linear coordinate system is assumed. We have described a number of data-driven algorithms for deriving nonlinear data manifold coordinate systems, e.g., ISOMAP and LLE. The only globally optimal algorithm is ISOMAP, which scales computationally as $O(N^2 \log_d(N))$ with memory requirements of $O(N^2)$, making it impractical in its original form for large remote sensing datasets. We have described a number of practical strategies for overcoming these scaling issues. All of these approaches relied on the ability to divide, conquer, and merge smaller subsets of the remote sensing data. Each of the approaches to scaling the representation began by dividing the scene into a large number of computationally tractable tiles. An optimal tile manifold coordinate system was then obtained using ISOMAP, and then the goal was to align the manifold coordinate system to a consistent set of global coordinates. All of the approaches then relied on some set of common or similar points to align the manifold coordinates systems. The most computationally flexible and scalable approach to aligning these coordinate systems aligned tiles in pairs, using a set of random samples from one tile to construct a mapping of these points to a nearby tile. This approach used the LLE reconstruction algorithm to find the best mapping of one manifold coordinate to another. Other alternatives that were explored included optimization of a decimated backbone manifold to which tile manifolds were aligned. The LLE pseudoinverse merge of the ISOMAP optimized tiles scaled the best, since it scales as $O(n_{f_i}^2)$, where n_{f_i} is the number of points needed to construct the mapping from one tile to an adjacent one, and $n_{f_i} \ll N$. The overall approach was demonstrated on nine adjacent tiles in an airborne hyperspectral scene. A metastrategy for autoalignment of manifold tiles was also outlined and demonstrated on a 15-tile example; areas for improvement (small artifacts) were identified, and suggestions for mitigation were described and will be the subject of future investigation. General scaling analysis of our nonlinear dimensionality reduction algorithm revealed computation times of $O(N^{4/3} \log(N))$ for a parallel implementation using a distributed memory architecture. Optimal tile size implies reasonable memory

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

requirements that are available on present-day parallel computers. Even serial implementations on a single CPU could be potentially useful in some applications with present-day workstations. Deriving 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

The authors gratefully acknowledge computing resources provided by the Department of Defense High Performance Computing Modernization Program, including the Maui High Performance Computing Center, the Space and Missile Defense Command, and the Army Research Laboratory's Major Shared Resource Center.

REFERENCES

[1] D. K. Agrafiotis, "Stochastic proximity embedding," *J. Comput. Chem.*, vol. 24, pp. 1215–1221, 2003.

[2] D. K. Agrafiotis and H. Xu, "A self-organizing principle for learning nonlinear manifolds," *Proc. Nat. Acad. Sci.*, vol. 99, no. 25, pp. 15 869–15 872, Dec. 2002.

[3] T. L. Ainsworth and J. S. Lee, "Optimal image classification employing optimal polarimetric variables," in *Proc. IGARSS*, vol. 2, Toulouse, France, 2003, pp. 696–698.

[4] ——, "Optimal polarimetric decomposition variables—Non-linear dimensionality reduction," in *Proc. IGARSS*, vol. 2, Sydney, Australia, 2001, pp. 928–930.

[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. J. 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] 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. 7a, Jan. 2002.

[9] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. M. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. (1994) Templates for the solution of linear systems building blocks for iterative methods. SIAM, Philadelphia, PA. [Online]. Available: http://www.netlib.org/templates/Templates.html.

[10] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1, Pasadena, CA, 1993, JPL Pub. 93–26, pp. 11–14.

[11] ——, "Analysis, understanding and visualization of hyperspectral data as convex sets in n-space," *Proc. SPIE*, vol. 2480, 1995.

[12] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Summaries 5th JPL Airborne Earth Science Workshop*, vol. 1, 1995, JPL Pub. 95-1, pp. 23–26.

[13] J. Bowles, P. Palmadesso, J. Antoniades, M. Baumback, and L. J. Rickard, "Use of filter vectors in hyperspectral data analysis," *Proc. SPIE*, vol. 2553, pp. 148–157, 1995.

[14] J. Bowles, M. Daniel, J. Grossman, J. Antoniades, M. Baumback, and P. Palmadesso, "Comparison of output from ORASIS and pixel purity calculations," *Proc. SPIE*, vol. 3438, pp. 148–156, 1998.

[15] J. Bowles, D. Gillis, and P. Palmadesso, "New improvements in the ORASIS algorithm," in *Proc. IEEE Aerospace Conf.*, vol. 3, 2000, pp. 293–298.

[16] C.-I Chang and S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[17] R. N. Clark, G. A. Swayze, C. Koch, A. Gallagher, and C. Ager, "Mapping vegetation types with the multiple spectral feature mapping algorithm in both emission and absorption," in *Summaries of the 3rd Annu. JPL Airborne Geosciences Workshop*, vol. 1, 1992, AVIRIS Workshop JPL Pub. 92–14, pp. 60–62.

[18] "Mapping minerals with imaging spectroscopy," U.S. Geol. Surv., Office of Mineral Resources, Reston, VA, Bull. 2039, 1993.

[19] D. Gillis, P. Palmadesso, and J. Bowles, "Automatic target recognition system for hyperspectral imagery using ORASIS," *Proc. SPIE*, vol. 4381, pp. 34–34, 2001.

[20] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Num. Math.*, vol. 1, pp. 269–271, 1959.

[21] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci.*, vol. 100, no. 10, pp. 5591–5596, May 2003.

[22] 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.

[23] 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.

[24] K. Guilfoyle, M. L. Althouse, and C.-I. Chang, "Further investigations into the use of linear and nonlinear mixing models for hyperspectral image analysis," *Proc. SPIE*, vol. 4725, pp. 157–167, 2002.

[25] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

[26] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[27] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ*, vol. 44, pp. 145–163, 1993.

[28] V. Kumar, A. Grama, A. Gupta, and G. Karypis, *Introduction to Parallel Computing, Design and Analysis of Algorithms*. Redwood City, CA: Benjamin/Cummings Publishing Company Inc., 1994.

[29] M. Lewis, V. Jooste, and A. A. de Gasparis, "Discrimination of arid vegetation with airborne multispectral scanner hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 7, pp. 1471–1479, Jul. 2001.

[30] C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*. San Diego, CA: Academic, 1994.

[31] 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. 19 419–19 425, Aug. 1998.

[32] A. Plaza, P. Matrinez, R. Perez, and J. Plaza, "A quantitative and comparitive analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

[33] D. N. Rassokhin and D. K. Agrafiotis, "A modified update rule for stochastic proximity embedding," *J. Molecul. Graph. Modeling*, vol. 22, pp. 133–140, 2003.

[34] D. A. Roberts, M. O. Smith, and J. B. Adams, "Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data," *Remote Sens. Environ.*, vol. 44, pp. 255–269, 1993.

[35] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[36] L. K. Saul and S. T. Roweis. (2001, Jan.) An introduction to locally linear embedding. [Online]. Available: http://www.cs.toronto.edu/~roweis/lle/publications.html.

[37] 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. 1999.

[38] R. Sedgewick, *Algorithms in C++, Part 5: Graph Algorithms*. Reading, MA: Addison-Wesley, 2002.

[39] T. H. Cormen, L. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.

[40] G. Sleijpen and H. van der Vorst. (1994) Jacobi–Davidson Methods. [Online]. Available: http://www.cs.utk.edu/~dongarra/etemplates/node136. html
[41] N. Straumann, *General Relativity and Relativistic Astrophysics*. Berlin, Germany: Springer-Verlag, 1984.
[42] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, Dec. 2000.
[43] F. Van Der Meer, "Iterative spectral unmixing," *Int. J. Remote Sens.*, vol. 20, no. 17, pp. 3431–3436, 1999.
[44] W. Verhoef, "Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model," *Remote Sens. Environ.*, vol. 16, pp. 125–141, 1984.
[45] M. M. Verstraete and B. Pinty, "Designing optimal spectral indexes for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 5, pp. 1254–1264, Sep. 1996.
[46] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. Applied Geologic Remote Sensing*, vol. II, Vancouver, BC, Canada, 1999, pp. 337–344.

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.

While at Brown University, he participated in interdisciplinary research in the Center for Neural Science, investigating adaptive models related to neurobiology and to statistical pattern recognition systems for applications such as speech recognition. 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 from 1994 to 1996. In 1997, he moved to the Remote Sensing Division, where he is currently Head of the Coastal Science and Interpretation Section of the new 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 that focused on coastal land-cover from hyperspectral and multisensor imagery. His research interests include image and signal processing techniques and 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 as these apply to environmental remote sensing, especially wetlands and coastal environments.

Dr. Bachmann is a member of the American Geophysical Union, the Society of Wetland Scientists, and the Sigma Xi Scientific Research Society. He is the recipient of three NRL Alan Berman Publication Awards (1994, 1996, and 2003) and an Interactive Session Paper Prize at IGARSS '96.

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.

He has held academic research and teaching positions at the University of Arizona, Tucson, the University of Illinois, Urbana, Forschungszentrum Jülich (KFA), Jülich, Texas A&M University, College Station, and the State University of New York, Stony Brook. His areas of research include computational many-body physics, quantum fluids, nuclear matter equation of state, superfluidity, and the structure and properties of neutron stars. Since joining the Remote Sensing Division, Naval Research Laboratory, Washington, DC, in 1993, his research interests are primarily focused on the analysis and applications of polarimetric and interferometric SAR imagery.

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

He has been with 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.

Authorized licensed use limited to: Univ of Puerto Rico Mayaguez - Library. Downloaded on January 27, 2009 at 11:10 from IEEE Xplore. Restrictions apply.