Local Intrinsic Dimensionality of Hyperspectral Imagery from Non-linear Manifold Coordinates

T. L. Ainsworth[1], C. M. Bachmann[2] and R. A. Fusina[2]
[1]Code 7263, Remote Sensing Division
Naval Research Laboratory, Washington, DC USA 20375-5351 USA
[2]Code 7232, Remote Sensing Division
Naval Research Laboratory, Washington, DC USA 20375-5351 USA

Abstract—**Recent advances in non-linear dimensionality approaches to hyperspectral image analysis have renewed interest in and provided a means to determine the intrinsic dimensionality of hyperspectral data. Of the many theoretical / computational approaches to dimensionality reduction, we discuss the local intrinsic dimensionality derived from the ISOMAP manifold approach. The ISOMAP algorithm itself provides one measure of global dimensionality, the eigenvalue spectrum. However, other estimation techniques are more easily adapted to determine the local dimensionality; chief among these are methods that measure minimum spanning path lengths for subsets of the full image. While intrinsic dimensionality is inherently scene dependent, knowledge of the information content, and underlying dimensionality, can guide data analysis methods and suggest appropriate data compression limits.**

Keywords-Hyperspectral imaging, non-linear dimensionality reduction

I. Introduction

Typical hyperspectral imagery (HSI) consists of measurements of a large number (~100) of closely spaced, narrow bands that cover the visible and near infrared portion of the electromagnetic spectrum. While the range and placement of the measured bands varies somewhat from sensor to sensor it is the large number of measured bands that simultaneously complicates the analysis of HSI and suggests an approach to simplify the interpretation of HSI: dimensionality reduction. Neighboring hyperspectral bands often display strong correlations implying that the number of statistically independent parameters is less than the number of measured spectral bands. Estimating the true number of independent parameters in HSI, i.e. the intrinsic dimensionality of the HSI, becomes important for subsequent interpretation of the imaged scene.

Especially, bothersome when analyzing high dimensional data is the so-called "Curse of Dimensionality". Many computational algorithms scale exponentially with the dimensionality of the data. Merely, partitioning a D-dimensional Euclidean hypercube, $[0, 1.0]^D$, into smaller hypercubes of size $[0, 0.1]^D$ becomes problematic for large D. In this simple example, every decrease of the dimensionality, D, by one decreases the computational load by an order of magnitude. When D surpasses 100, as is typical for hyperspectral data, even seemingly simple tasks can become extremely burdensome. Therefore, applying any reasonable means to reduce data dimensionality produces significant and immediate benefits.

II. Methodology

We have recently explored non-linear manifold approaches to the dimensionality reduction problem as applied to HSI [1], [2]. During the course of these studies, we observed that the standard measure of intrinsic dimensionality from the ISOMAP algorithm provides a global estimate of the maximum number of manifold dimensions needed to characterize the entire scene [3]. Missing, in particular, is the variation of the intrinsic dimensionality that depends upon the local characteristics of that portion of the full scene. Applying ISOMAP to HSI containing both land and water indicates several inherent differences between land and water regions, including a different intrinsic dimensionality [4], [5]. Estimating the local intrinsic spectral dimensionality of features observed in HSI allows a better, more compact, description of those features. An increase of the local intrinsic dimensionality from one spatial region of interest to another signals the onset of additional spectral structure and hence a more complex set of optical scatterers.

ISOMAP provides one an estimate of the *global* intrinsic dimensionality of the data manifold. The ISOMAP eigenvalue spectrum typically shows a rapid fall off from the maximum eigenvalue, followed by a long tail of uniformly small values. The intrinsic dimensionality is the number of eigenvalues in the initial rapid fall off region [3]. However, estimating a precise dimensionality requires some judgment since the transition between the rapid fall off region and spectrum tail is not always sharp. The dimensionality derived in this manner from the eigenvalue spectrum is biased. The dimensionality derived form the eigenvalue spectrum tends to be an over-estimate the true global intrinsic dimensionality.

Other methods to estimate dimensionality have been proposed, e.g. lengths of minimum spanning trees (MST) [6], [7], or data reconstruction errors [8]. Here we emphasize the MST approaches to estimates of manifold dimensionality. The basic concept is that the length of a MST scales with the number of points (pixels) in the tree and depends exponentially on the intrinsic dimensionality of the data space spanned by the set of pixels chosen. The basic scaling results have been rigorously developed for manifolds generated by ISOMAP

1-4244-1212-9/07/$25.00 ©2007 IEEE.

techniques for an asymptotically large number of pixels [6], [7]. One estimates the intrinsic dimensionality by choosing different sized sets of pixels, computing the MST lengths, and comparing the scaling behavior to the asymptotic result. The MST length is fixed for any given set of image pixels, but in practice, we obtain estimates of the dimensionality by averaging results from a number of randomly chosen sets. To estimate the global dimensionality, the pixels for each MST computation are selected from the entire image. Choosing the set of points from a local region of interest permits one to estimate a *local* intrinsic dimensionality [7]. Here the terms global and local refer to the spectral space and spectral coordinates, not the image space and associated pixel positions.

Another possible wrinkle in the estimate of the global and local dimensionalities involves the use of approximate variants to the original ISOMAP algorithm. As originally proposed [3], the ISOMAP algorithm provided a globally optimal solution for a relatively small number of samples, i.e. image pixels. The computational techniques do not scale nicely for larger images typically employed in remote sensing applications [1], [2]. Approximate versions of ISOMAP trade global optimality for the ability to handle reasonably sized HSI. For defining manifold coordinates to represent the full spectrum HSI, the approximate forms of ISOMAP typically perform nearly as well as the original, globally optimal one. However, for some scenes comparison of the eigenvalue spectrums show differences between the globally optimal and sub-optimal variants of ISOMAP. Therefore, we have employed smaller sub-images to cross compare these ISOMAP methods.

III. Summary and Discussion

The intrinsic dimensionality is scene dependent; therefore we present an assessment of the local intrinsic dimensionality for a number of hyperspectral scenes collected by the PHILLS hyperspectral camera. We selected specific scenes of the Virginia Coastal Reserve off the Delmarva Peninsula, the Oregon Inlet off the North Carolina coast and the Indian River Lagoon in Florida due to their varied terrain and ground cover vegetation, and the proximity of water for the related bathymetric studies [4]. Specific regions of interest within a scene may display a lower intrinsic dimensionality and hence a simpler spectral structure, which becomes important for deriving interpolating look-up-tables, e.g. bathymetric retrieval tables.

We will contrast and compare the local dimensionality estimates from the eigenvalue spectrum, and different MST calculations. As part of the comparison, local dimensionalities derived from the globally optimal ISOMAP and two computationally efficient, but sub-optimal, variants of ISOMAP will be compared. The computationally efficient versions employ a set of "landmarks" that span the manifold space, however since the landmarks are a small subset of the full HSI scene a globally optimal solution does not result. Then given the set of landmarks we test two standard algorithms that extract sets of manifold coordinates. To our knowledge this is the first explicit test of the local intrinsic dimensionality of HSI imagery analyzed with non-linear dimensionality reduction techniques.

References

[1] C. M. Bachmann, T. L. Ainsworth, R. A. Fusina, "Improved Manifold Coordinate Representations of Large Scale Hyperspectral Imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, pp. 2786-2803, 2006.

[2] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting Manifold Geometry in Hyperspectral Imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, pp. 441-454, 2005.

[3] J. B. Tenenbaum, V. de Silva, J.C Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319-2323, 2000.

[4] C. M. Bachmann, T. L. Ainsworth, R. A. Fusina, M. J. Montes, J. H. Bowles and D. R. Korwan, "Bathymetric Retrieval from Manifold Coordinate Representations of Hyperspectral Imagery," *Proc. of the International Geoscience & Remote Sensing Symposium* (IGARSS'07), Barcelona, July 2007, to be published.

[5] 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, W. J. Rhea, "Modeling Coastal Waters from Hyperspectral Imagery Using Manifold Coordinates," *Proc. of the International Geoscience & Remote Sensing Symposium* (IGARSS'06), Denver, July/August 2006.

[6] J. A. Costa and A. O. Hero, III, "Geodesic Entropic Graphs for Dimension and Entropy Estimation in Manifold Learning," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2210-2221, 2004.

[7] J. A, Costa, A. Girotra and A. O. Hero, III, "Estimating Local Intrinsic Dimension with *k*-Nearest Neighbor Graphs," *IEEE Workshop on Statisitcal Signal Processing*, pp. 417-422, 2005.

[8] S. Martin and A. Bäcker, "Estimaing Manifold Dimensinality by Inversion Error," *ACM Symposium on Applied Computing*, pp.22-26, 2005.