

12-10-2010

Spherical Nearest Neighbor Classification: Application to hyperspectral Data

Dalton Lunga

Electrical and Computer Engineering, Purdue University

okan ersoy

Electrical and Computer Engineering, Purdue University, ersoy@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/ecetr>

Lunga, Dalton and ersoy, okan, "Spherical Nearest Neighbor Classification: Application to hyperspectral Data" (2010). *ECE Technical Reports*. Paper 403.

<http://docs.lib.purdue.edu/ecetr/403>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Spherical Nearest Neighbor Classification: Application to Hyperspectral Data

Dalton Lunga

Okan Ersoy

TR-ECE-10-09

December 10, 2010

School of Electrical and Computer Engineering

1285 Electrical Engineering Building

Purdue University

West Lafayette, IN 47907-1285

Spherical Nearest Neighbor Classification: Application to Hyperspectral Data

Dalton Lunga and Okan Ersoy
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907-0501
USA
{dlunga, ersoy}@purdue.edu

December 10, 2010

Abstract

The problem of feature transformation arises in many fields of information processing, including machine learning, data compression, computer vision and geosciences applications. Here we discuss an approach that seeks a hyperspherical coordinate system preserving geodesic distances in the high dimensional hyperspectral data space. A lower dimensional hyperspherical manifold is computed using a lower rank matrix approximation algorithm combined with the recently proposed spherical embeddings method. Three spherical metrics for classification that exploits the nonlinear structure of hyperspectral imagery based on the properties of hyperspherical surfaces and their relationship with local tangent spaces are proposed. As part of experimental validation, results on modeling multi-class multispectral data using the proposed spherical geodesic nearest neighbor, the spherical Mahalanobis nearest neighbor and the spherical discriminant adaptive nearest neighbor rules are presented. The results indicate that the approach yields promising and better classification accuracies especially for difficult tasks in spaces with complex irregular class boundaries.

1 Introduction

For several years, feature extraction methods in the form of *best band* combinations have been the most applied standards in the analysis of hyperspectral data. The best band approach relies on the presence of narrowband features which may be the characteristic of a particular category of interest or on known physical characteristics of broad classes of data, e.g., vegetation indices (Clark *et al.*, 1992). On the other hand, the underlying assumptions of feature extraction methods 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 and iterative spectral unmixing (Bachmann *et al.*, 2005). 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 (Sandmeier *et al.*, 1999). In land-cover applications, bidirectional reflectance distribution function 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 bidirectional reflectance distribution function effects include the optical characteristics of the canopy, canopy gap function, leaf area index, and leaf angle distribution (Sandmeier *et al.*, 1999). It has been observed that wavelengths with the smallest reflectance exhibit the largest nonlinear variations (Sandmeier *et al.*, 1999). 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. Other effects that contribute to nonlinearities include multiple scattering within a pixel and the heterogeneity of subpixel constituents. Classification of hyperspectral image data that exhibits these non-linearities poses a huge challenge to linear methods.

In this paper, we take a different approach from the feature extraction methods, in that we seek to exploit the nonlinear structure of hyperspectral imagery by using a feature transformation method. This new approach seeks a constant curvature coordinate system that preserves geodesic distances in the high-dimensional hyperspectral feature space. We then define new modified nearest neighbor rules for classification in spherical spaces on this basis. Manifold learning methods are commonly becoming a standard to embedding data onto their new transformed spaces. Many of the manifold learning methods embed objects into a lower dimensional vector-space using techniques such as Multidimensional Scaling (Cox & Cox, 2001), Diffusion Maps (Coifman & Lafon, 2006), Locally Linear Embedding (Roweis & Saul, 2000), or Principal Component Analysis (Jolliffe, 1986). Each of these approaches represents an attempt to derive a coordinate system that resides on (parameterizes) the nonlinear data manifold itself. The methods represents 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. Once embedded in such a space, the data points can be characterised by their embedding co-ordinate vectors, and analysed in a conventional manner using Euclidean space methods. Models can be developed for the low dimensional embedded data, but the challenge remains on how to interpret the geometrical characteristics of the new space so that decision making tools can take advantage of these properties. There also exists some limits to these paradigms; Euclidean distances are always definite and are intrinsically unable to represent dissimilarities which are indefinite. Recently an alternative, *spherical local embeddings (SLE)*, method that maps the dissimilarity of shape objects onto a constant curvature Riemannian manifold has been proposed in (Wilson *et al.*, 2010). It is a method that embed indefinite data onto a non-Euclidean metric space while optimizing over the kernel distance matrix of positional vectors. With data embedded onto a manifold, analytical techniques can be applied for analysis. One common method for performing such an analysis is the classification of new points according to functions or set of rules that make their decision based on the geometry of training data samples.

In a supervised classification problem, we are given C classes and N training examples. The training examples consists of d feature measurements $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and the known class labels $L_c, c = (1, \dots, C)$. The goal is to predict the class label for a given test point x_0 .

The K-nearest neighbor (K-NN) classification method is a simple and appealing approach to this problem: it finds the K-nearest neighbors of x_0 in the training set, and then predicts the class label of x_0 based on the majority voting scheme. Such a method produces continuous and overlapping, rather than fixed, neighborhoods, and uses a different neighborhood for each test vector. K-NN methods are very flexible and do not usually involve any preprocessing. Furthermore in spite of the introduction of many sophisticated classification algorithms, they remain among the most successful over very difficult classification tasks. It has been shown (Hastie *et al.*, 2009), that the 1-NN rule has asymptotic error rate that is at most twice the Bayes error rate, independent of the distance metric used. However, due to the curse-of-dimensionality, the 1-NN rule becomes less appealing with a finite set of training samples. Severely biased estimates can be introduced in the 1-NN rule in a high-dimensional input feature space with a finite number of samples. As such, the choice of a distance measure becomes crucial in determining the outcome of the NN classification. The commonly used Euclidean distance measure, while simple computationally, implies that the input space is isotropic or homogeneous. However, the assumption for isotropy is often invalid and generally undesirable for classification of multispectral images that include non-linear medium, i.e. water. As a consequence, the distance computation does not vary with equal strength or in the same proportion in all directions in the feature space emanating from the input query. Capturing the non-linear nature of scattering in land-cover classification due to bidirectional distribution function effects calls for a modified set of classification methods that adapts to the embedding space.

Our main contribution in this paper, is in proposing a set of geometrical classification rules for hyperspectral data embedded on a unit hyperspherical manifold. The main purpose of embedding the data onto a unit curvature manifold is to faithfully represent the dissimilarities between objects in a metric space in which the analysis rely on directional properties of the samples. A metric space is important because it allows for development of statistical geometrical tools and redefinition of geometric constructs such as boundaries, in contrast to a non-metric space where nonlocality is not well defined. In the next sections, we apply the recently proposed Spherical Local Embedding (SLE) (Wilson *et al.*, 2010) to establish a unit hyperspherical coordinate system that preserves geodesic distances in the high-dimensional hyperspectral data space. We then propose three classification metrics which are: (1)-the new spherical geodesic nearest neighbor (*sphknn-geodesic*), (2)-spherical Mahalanobis nearest neighbor (*sphknn-Mahalanobis*) and (3)-the spherical discriminant adaptive nearest neighbor (*sphDANN*) metric that are compatible with the Riemannian geometry for spherical manifolds. Such metrics, and hence the resulting neighborhoods, depends on the test point locations on the spherical manifold. We estimate the decision rules for the classification tasks by making use of the simple tools from Lie algebra and Lie groups to perform computations in the local tangent spaces.

2 On Riemannian Space

In non-Euclidean spaces, computations are carried out by using different tools than the standard methods used in a Euclidean space. The geometry that exist in Riemannian manifolds dictates on how these tools are formulated. On a spherical manifold, a convenient way to measure the distance between two points is no longer the straight line between the points as in the Euclidean space. Distances on spherical surfaces are defined as the length of the shortest curve between a pair of points (this defines the notion of *geodesic*). We revisit the notion of geodesic distance

and Riemannian metric in the context of *Tangent spaces* in the following sections. The method of embedding onto a constant curvature Riemannian space has recently been proposed in (Wilson *et al.*, 2010). In this section we simply revisit the spherical local embeddings (SLE) formulation. A d -dimensional Riemannian space is defined by its tensor g_{ij} in some local coordinate system u_1, u_2, \dots, u_d . This is usually related to an infinitesimal distance element in the space by

$$ds^2 = \sum_{ij} g_{ij} du_i du_j \quad (1)$$

The metric must be positive definite, and any metric tensor defines a particular Riemannian space. A simple form of a Riemannian manifold that easily relates to directional data is the elliptic manifold (Wilson *et al.*, 2010).

2.1 Constant Curvature Manifolds

An elliptic manifold is an example of a constant curvature manifold. A manifold is defined as the geometry on the surface of a hypersphere. In some cases, a hypersphere can easily be embedded in the Euclidean space, for example, the embedding of a sphere in three dimensions is

$$x = (r \sin u \sin v, r \cos u \sin v, r \cos v)^T \quad (2)$$

A spherical embedding (elliptic) implies a metric tensor of the form

$$ds^2 = dx^2 + dy^2 + dz^2 \quad (3)$$

$$= r^2 \sin^2 v du^2 + r^2 dv^2 \quad (4)$$

The embeddings of an $d - 1$ dimensional hypersphere in a d dimensional space follows from this equation. The surface of the hypersphere can be implicitly defined by the constraint

$$\sum_i x_i^2 = r^2 \quad (5)$$

The surface is curved with a constant radius of curvature $R = 1/r^2$. The geodesic distance of two points on a curved space is the length of the shortest curve lying in the space and joining the two points. For elliptic manifolds, the geodesic is a great circle on the hypersphere. The distance is the length of the arc of a great circle which joins the two points. If the angle subtended by the two points at the center of the hypersphere is θ_{ij} , then the distance between them is given by

$$d_{ij} = r\theta_{ij} \quad (6)$$

Given that the coordinate is at the center of the hypersphere, we can represent any point by a position vector x_i of length r . Since the dot product is $\langle x_i, x_j \rangle = r^2 \cos \theta_{ij}$

$$d_{ij} = r \cos^{-1} \left(\frac{\langle x_i, x_j \rangle}{r^2} \right) \quad (7)$$

2.2 Hyperspherical Manifolds

Given a distance matrix D from the Euclidean space, the goal is to find a Riemannian space kernel matrix with approximately the same distance position for each pair of sample vectors. In (Wilson *et al.*, 2010) the authors consider as first step the determination of the radius of curvature for the manifold. However we relax this requirement and fix the manifold to be a unit hypersphere. Given n objects, the goal would be to determine a $n - 1$ dimensional Euclidean space. With the freedom to choose the radius of curvature the task is then to search for a $n - 2$ dimensional space embedding in an $n - 1$ dimensional Euclidean space.

The first step is to construct a space with the origin at the center of the hypersphere. If the point positions are given by x_i , $i = 1, \dots, n$, then

$$\langle x_i, x_j \rangle = r^2 \cos \theta_{ij} = r^2 \cos\left(\frac{d_{ij}}{r}\right) \quad (8)$$

From the above equation, a matrix of positional hyperspherical vectors is defined to be X , with each position vector as a row. The goal is to match the Riemannian kernel matrix from the outer product of the positional matrix to the cosine similarity kernel Z

$$XX^T = Z \quad (9)$$

where $Z_{ij} = r^2 \cos\left(\frac{d_{ij}}{r}\right)$ and $d_{ij} \in D$. Since the embedding space has dimension $n - 1$, X consists of n points of dimension $n - 1$ and Z is a $(n \times n)$ positive semi-definite matrix with rank $n - 1$. Z has a single eigenvalue that is zero, with the rest positive. This observation led Wilson (Wilson *et al.*, 2010) to compute the radius of curvature by exploring the eigenspectrum of the kernel. Thus, Z is computed as a function of r and finding the smallest eigenvalue λ_1 determines the objective function to be minimized. Therefore, r^* is determined by minimizing the magnitude of the smallest eigenvalue as a function of r :

$$r^* = \operatorname{argmin}_r |\lambda_1\{Z(r)\}| \quad (10)$$

Given the optimal radius r^* , the embedding positions are determined through the eigendecomposition of $Z(r^*)$

$$Z(r^*) = U\Lambda U^T \quad (11)$$

We however fix the radius to $r = 1$ so that the curvature is maintained to be a unit hypersphere, and that enables our proposed approach to use some of the existing tools from directional statistics. The intuition of modeling on hyperspherical surfaces is also somehow simplified when the discussion is centered on the directional components of the positional vectors. The matrix of spherical positional vectors X is determined as

$$X = U\Lambda^{1/2} \quad (12)$$

2.3 Dimensionality Reduction

The data matrix obtained by the embedding method described above is high dimensional. We propose a method to reduce the dimension to a lower spherical manifold by computing a lower

rank approximation to the positional matrix X subject to fixing the radius of the hypersphere and the norm of each positional vector to be of unit length. We achieve this by applying a well known theorem, due to Eckart and Young (Eckart & Young, 1936), which computes for a matrix of the required rank with lowest possible Frobenius error.

Theorem 1.

$$\underbrace{\operatorname{argmin}}_{Y|\operatorname{rank}(Y)=k, \|y\|=1} \|X - Y\|_F = \|X - X_k\|_F \quad (13)$$

$$= \sqrt{\sum_{i=k+1}^{n-2} \sigma_i^2}. \quad (14)$$

Having the singular values in decreasing order $\sigma_1 \geq \sigma_2 \geq \dots$, X_k can be observed to be the best rank- k approximation to X , incurring an error (measured by the Frobenius norm of $X - X_k$) equal to σ_{k+1} . Thus the larger k is, the smaller this error becomes, but then the dimension increases with larger k . So a tradeoff has to be reached in achieving a low dimension positional matrix X_k and achieving a smaller Frobenius error.

Once the feature matrix is embedded to the required lower dimensional hypersphere, analysis can be performed. The computation of some statistics of interest to enable data analysis are made possible by applying some definitions from Lie groups and Lie algebra. In the next section, we present some notation and definitions on the relationship between a spherical surface and the local tangent planes.

2.4 Tangent Space and Curved Manifolds

A *Riemannian metric* on a manifold M is a smoothly varying inner product $\langle \cdot, \cdot \rangle$ on the tangent plane $T_p M$ at each point $p \in M$. The norm of a vector $x \in T_p M$ is given by $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$. The Riemannian distance between two points $p, s \in M$, denoted by $d(p, s)$, is defined as the minimum length over all possible smooth curves between p and s . Given a tangent vector $x \in T_p M$, there exists an unique geodesic, $\zeta_x(t)$, with x as its initial velocity. The Riemannian *exponential map*, denoted by Exp_p , maps x to the point at time one along the geodesic $\zeta_x(t)$. The exponential map preserves distances from the initial point, i.e., $d(p, Exp_p(x)) = \|x\|$. In the neighborhood of zero, its inverse is defined and is called the Riemannian *log map*, denoted by Log_p . Thus, for a point y in the domain of Log_p , the geodesic distance between p and y is given by

$$d(p, y) = \|Log_p(y)\| \quad (15)$$

2.4.1 Exponential and Log Maps

On the sphere \mathbb{S}^2 , the geodesics at the base point $p = (0, 0, 1)$ are great circles through p . If we consider the tangent vector $x = (x_1, x_2, 0) \in T_p \mathbb{S}^2$ in the $x - y$ plane, the exponential map at p is given by

$$Exp_p(x) = \left(x_1 \cdot \frac{\sin \|x\|}{\|x\|}, x_2 \cdot \frac{\sin \|x\|}{\|x\|}, \cos \|x\| \right) \quad (16)$$

where $\|x\| = \sqrt{x_1^2 + x_2^2}$. The corresponding log map for a point $s = (s_1, s_2, s_3) \in \mathbb{S}^2$ is given by

$$\text{Log}_p(s) = \left(s_1 \frac{\theta}{\sin \theta}, s_2 \frac{\theta}{\sin \theta} \right) \quad (17)$$

where $\theta = \arccos(s_3)$ is the angle from the base point p to the point s . In general, a point s on the hypersphere is mapped to the local tangent space using

$$x = \frac{\theta}{\sin \theta} (s - p \cos \theta) \quad (18)$$

while a point x on the tangent space is mapped to the hypersphere using

$$s = p \cos \theta + \frac{\sin \theta}{\theta} x \quad (19)$$

In the following section, we make use of the above relations to simplify the design of three nearest neighbor classification metrics for spherical manifolds.

3 Nearest-Neighbors in Spherical Spaces

3.1 Spherical Geodesic and Spherical Mahalanobis Metrics

The nearest neighbor is a non-parametric classifier that is memory-based. The classification rule is based on a test vector x_0 and a set of training vectors $\{x_n\}_{n=1}^N$. In a Euclidean space, k nearest points (in distance) to x_0 are chosen from the training set and their class labels are used in a majority voting procedure to select the dominating class label as the class to assign the test vector x_0 . Ties are broken at random and usually the number of K-nearest neighbors is chosen through cross validation methods before the rule is applied to test data. The classifier is one of the most straight forward to implement. If one assumes that the feature vectors are real-valued, then the Euclidean distance is usually used to compute the distance between a given test point x_0 and the potential Nearest Neighbor x . For Euclidean spaces the distance between x and x_0 is defined by

$$D_n = \|x_n - x_0\|_2 \quad (20)$$

Given its simplicity, the K-nearest neighbor rule has been applied with success in many classification problems. It is often successful where each class has many possible prototypes, and the decision boundary is very irregular. It is a classifier whose properties can be easily extended to non-Euclidean geometries. Our goal is to extend the K-NN methods to spherical manifolds by defining the tools required to carry out three most widely used Euclidean space nearest neighbor rules. The proposed metric rules will incorporate the new geometry of the transformed data. All that is required to apply K-NN in spherical manifolds is to compute the geodesic distances between a test point and its nearest neighbors. The challenge is in computing geodesic distances. Since it is not an easy task to carry out, we apply the rules and mappings presented above and compute all the requires statistics in local tangent spaces. Where necessary, the values are mapped back to the spherical spaces using the Log map function, otherwise the classification decision can be made

within the local tangent spaces. For spherical spaces, the geodesic distances are equal to their local tangent space distances. They are computed by

$$D_g(s, p) = D_t(\log_p s, \log_p p) \quad (21)$$

$$= \|\log_p s\| \quad (22)$$

Equation 22 is used as a metric for computing geodesic distances between an embedded test point and its neighboring points on the spherical manifold. The decision to assign a label to the query point is based on the majority rule from the closest (as defined by (22)) K-spherical coordinates. The Spherical K-NN distance rule of (22) can be improved by incorporating the structure of the spherical nearest neighbor points. In Euclidean spaces, this is achieved by making use of the covariance matrix to define the *Mahalanobis distance* measure. The notion of a covariance matrix can easily be extended to spherical spaces by noting that a Mahalanobis distance is a distance between a random point $\log_p s \sim \mathcal{N}(\overline{\log_p s}, \Sigma_{(\log_p s \log_p s)})$ and a (deterministic) point $\log_p p$. It is defined by

$$D_m(s, p) = (\log_p s - \log_p p)^T \Sigma^{-1} (\log_p s - \log_p p) \quad (23)$$

$$= (\log_p s)^T \Sigma_{(\log_p s \log_p s)}^{-1} (\log_p s) \quad (24)$$

Equation (24) simplifies as shown because the base point p on the manifold maps to $\log_p p$ which is the origin of the tangent space. The classification of a point is carried out in the same manner in which the standard K-NN Mahalanobis distance based classifier is done. So from the selected K-nearest neighbors, we chose the class with a dominating label from the nearest neighbor set chosen based on the spherical Mahalanobis metric.

3.1.1 Choice of the Base Point p

The choice of the base point is very critical for the classification rules discussed in this study. A mere naive implementation of these methods by which each test point has its own tangent space is computationally inefficient. To speed up computation, we propose computing the mean of each training class data as follows

$$\hat{\mu}_c = \frac{\sum_i^{n_c} x_i}{\|\sum_i^{n_c} x_i\|} \quad (25)$$

The base point is then chosen to be $p = \hat{\mu}_c$, with the proposed nearest neighbor computations performed in the C tangent spaces (a function of the number of classes) compared to the naive approach with a tangent space for each test point. The definition in equation (25) enables the computation of the training class directional mean vectors to be performed on the spherical surface since $\|\hat{\mu}_c\| = 1$ and $\hat{\mu}_c \in \mathbb{S}^{d-1}$. This approach is intuitive in the sense that better classification decisions are made for each test query using the actual training samples that are mapped to the local tangent space $T_{M_{\hat{\mu}_c}}(s)$.

Additional improvements to address some of the known challenges faced by simple metric rules defined in equation (22) and equation (24) can be incorporated. In many applications when the nearest-neighbor classification is carried out in a high dimensional feature space, the nearest neighbors of a point can be very far away, causing bias and degrading the performance of the voting rule

(Hastie *et al.*, 2009). These challenges call for adapting the metric used in nearest neighbor classification so that the resulting neighborhoods stretch out in directions for which the class probabilities don't change significantly. An extension of this method to spherical manifolds is presented.

3.2 Spherical Discriminant Adaptive Nearest-Neighbor (sphDANN) Classifier

In many high-dimensional problems, the nearest neighbor of a point can be very far away, causing bias and degrading the performance of the classification rule. This problem was addressed for Euclidean spaces in (Tibshirani & Hastie, 1996), where a *discriminant adaptive nearest-neighbor*(DANN) metric was presented. In (Tibshirani & Hastie, 1996), at each test point a neighborhood of say 50 points is formed, and the class distribution among the points is used to decide how to deform the neighborhood, meaning to adapt the rule or the metric. The adapted metric is then used in a nearest-neighbor rule at the query point. This process results in potentially different metrics for each query point based on the distribution of label boundaries near the test point. This locally discriminative procedure only demands that information contained in the local within-and between-class covariance matrices is all that is needed to determine the optimal shape of the neighborhood. An extension of the DANN metric to Riemannian spherical manifolds is simplified by taking advantage of the log-exponential mappings introduced earlier. Using the log-exponential mappings, we choose two points, s and p , on the spherical manifold and define their tangent space positions as

$$x = \log_p s \quad (26)$$

$$x_0 = \log_p p \quad (27)$$

Any points x and x_0 on the tangent space take on coordinates on the curved manifold as

$$s = \text{Exp}_p x \quad (28)$$

$$p = \text{Exp}_p x_0 \quad (29)$$

The tangent space is locally defined around x_0 , as such computing of the distance from x_0 to any other vector point x is just the norm of that vector, since x_0 forms the origin of the tangent space.

The *spherical discriminant adaptive nearest-neighbor* (sphDANN) metric at a query point $\log_p p$ is defined by

$$D_{sdann}(s, p) = (\log_p s - \log_p p)^T \Sigma (\log_p s - \log_p p) \quad (30)$$

The expression in (30) can be rewritten using the mappings from equation (27) as

$$D_{sdann}(x, x_0) = (x - x_0)^T \Sigma_{(x_0, x_0)} (x - x_0) \quad (31)$$

$$= x^T \Sigma_{(x_0, x_0)} x \quad (32)$$

The simplification is due to x_0 being the origin of the tangent space $T_{M_p}(s)$. Σ_{x_0, x_0} in equation (32) is defined by

$$\Sigma_{x_0, x_0} = S_w^{-1/2} \{ S_w^{-1/2} S_B S_w^{-1/2} + \epsilon I \} S_w^{-1/2} \quad (33)$$

S_w is the pooled within-class covariance matrix

$$S_w = \sum_{j=1}^J \pi_j S_{wj}$$

and S_B is the between class covariance matrix

$$S_B = \sum_{j=1}^J \pi_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T$$

with $\pi_j = \frac{N_j}{N}$.

$$\bar{x}_j = \frac{1}{N_j} \sum_{x \in \mathcal{C}_j} x \quad (34)$$

$$S_{wj} = \frac{1}{N_j} \sum_{n=1}^{N_j} (x - \bar{x}_j)(x - \bar{x}_j)^T \quad (35)$$

and

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (36)$$

The parameter ϵ with value set to 1 rounds the neighborhood from an infinite strip to an ellipsoid, so as to avoid using points far away from the query point. sphDANN involves choosing the nearest neighbors in the spherical manifold by first applying distance metric $D_t(x, x_0)$ defined in equation (22) or (24), to choose the initial 50 spherical nearest neighbors to a query point x_0 . The number 50 has been observed to present enough points for adapting the metric in the neighborhood of the query point. We can also consider this number to be a parameter to be chosen by cross validation methods. With all 50 points determined, equation 33 can be computed. The second part involves using the sphDANN metric in a nearest neighbor rule at x_0 . Note that the aim is to have the neighborhood of a query point stretched in the direction that coincides with the linear discriminant boundary of the classes. It is the direction in which class probabilities change the least.

4 Experiments

In this section, we briefly describe the datasets and experimental methodology used. We also discuss the performance of the three spherical nearest neighbor rules under consideration on the two datasets and compare the results to other methods that have been previously considered for the same datasets.

4.1 Datasets

We first present results on which standard hyperspectral classification has been reported to be very difficult: the Colorado dataset. The *Colorado dataset* consists of the following four data sources : (1) *Landsat MSS data* (four spectral data channels). (2) *Elevation data* (one data channel). (3) *Slope data* (one data channel). (4) *Aspect data* (one data channel). Each channel comprises an image of 135 rows and 131 columns, and all channels are spatially coregistered. There are ten ground-cover classes listed in Table 1. One class is water; the others are forest types. It is very difficult to distinguish among the forest types using Landsat MSS data alone since the forest classes show very similar spectral responses.

Table 1: Colorado dataset

Class	Type of Class	Training Samples	Testing Samples
c1	Water	408	195
c2	Colorado Blue Spruce	88	24
c3	Mountane/Subalpine meadow	45	42
c4	Aspen	75	65
c5	Ponderosa Pine	105	139
c6	Ponderosa Pine/Douglas Fir	126	188
c7	Engelmann Spruce	224	70
c8	Douglas Fir/White Fir	32	44
c9	Douglas Fir/Ponderosa Pine/Aspen	25	25
c10	Douglas Fir/White Fir/Aspen	60	39
	Total	1188	831

To further investigate the effectiveness of proposed hyperspectral feature transformation to spherical manifold and the application of the spherical nearest neighbor metrics, we generate results from the AVIRIS multispectral image. The West Lafayette image was used in the experiments. This data is a multispectral image from the Airbone/Infrared Imaging Spectrometer that was built by Jet Propulsion Laboratory and flown by NASA/Ames on June 12, 1992 (Landgrebe & Biehl, 1992). The scene is over an area that is 6 miles west of West Lafayette. It contains a subset of 9 bands from a significantly larger image with 220 bands. The bands considered have wavelengths $0.828 - 0.838$, $0.751 - 0.761$, and $0.663 - 0.673 \mu\text{m}$. The image has 17 classes (background, alfalfa, corn-notill, corm-min, corn, grass/pasture, grass/trees, grass/pasture-mowed, hay-windrowed, oats, soybeans-notill, soybean-min, soybean-clean, wheat, woods, dldg-grass-tree-drives, and stone-steel-towers). The image size is 145×145 pixels. The pixel resolution is 16 bits, corresponding to 65536 gray levels. About 3403 and 4195 pixels were selected to generate the ground-reference data for training and testing fields, respectively. In the experiments, each pixel is expressed as a vector of 9 features.

4.1.1 Dissimilarity Matrix

In order to carry out an analysis of data on a hyperspherical manifold, dissimilarity values are required between each pair of feature vectors in the collection. The structure of the embedding is sensitive to the dissimilarity measure used to compute matrix D . Some prior knowledge about the dataset could also be incorporated when selecting a dissimilarity measure. We made use of a general standard dissimilarity measure to compute the dissimilarity between each pair of vectors.

Each element of matrix D is defined by

$$d_{ij}^2 = (x_i - x_j)^T \Omega^{-1} (x_i - x_j) \quad (37)$$

where Ω is the $n \times n$ diagonal matrix whose j^{th} diagonal element is s_j^2 , where s is the vector of standard deviations. With all feature pairs embedded and their position vectors identified, we proceed to analyze the data by making use of the proposed spherical classification rules.

4.2 Results

We applied the proposed approach of spherical locally embedding of hyperspectral data and the new spherical nearest neighbor metrics to two land cover classification tasks. The performance is carried out in terms of classification accuracy and thematic maps. For each dataset, the proposed spherical nearest neighbor method's results are compared to those of algorithms that were previously applied on the same dataset.

On the first task, we experimented with the Colorado dataset and the results from the *spknn-geodesic*, *spknn-Mahalanobis* and the *sphDann* metrics were compared with those of previous methods that were used on the same dataset i.e. Minimum Euclidean Distance (ED), Maximum Likelihood (ML) and the Minimum Mahalanobis Distance (MD) (Benediktsson *et al.*, 1990).

The second tasks involved the AVIRIS dataset (Landgrebe & Biehl, 1992) and the results from applying the proposed metrics were compared with classification methods that were used on the this dataset i.e. Hierarchical Competitive Learning (HCL), Hierarchical Self-Organizing Maps and the Hierarchical Self-Organizing Global Ranking (SOGR) (Lee & Ersoy, 2007).

The following notation is used to identify the proposed classification method: *Eucliknn-mahalanobis* - is the K-nearest neighbor in the Euclidean space making use of the Mahalanobis distance metric. *spknn-geodesic* - is the spherical-nearest-neighbor on the spherical space making use of the geodesic distance defined in equation 22. *spknn-Mahalanobis* - is the spherical-nearest-neighbor on the spherical space making use of the spherical Mahalanobis distance defined in equation 24. *sphDann* - is the spherical-discriminative-adaptive-nearest-neighbor on the spherical space defined in equation 30.

To evaluate the output differences between any two algorithms, the confusion matrix was applied after each classification. The overall classification accuracy was computed as

$$OA = \frac{\sum_c N_{cc}}{N} \cdot 100\% \quad (38)$$

where N_{cc} is the number of testing samples correctly classified for class c , while N is the total number of testing samples. On the Colorado dataset, experiments were carried out using a 10-fold cross validation to determine the optimal \hat{K} value for the nearest neighbors. The final results were based on the average rate of correctly classified samples, averaged over the 20 runs.

The results in Table 2 indicate a percentage agreement with reference to each Class. It can also be observed from Table 2 that the proposed spherical metrics have a relatively high accuracy on each land cover class except for the *Douglas Fir/Ponderosa Pine/Aspen* class on which the previous methods did better. With regard to OA it is also evident that all three proposed metrics: *Sphknn-geodesic*, *sphknn-Mahalanobis*, and the *sphDann* have better results, with each attaining 73.13%, 71.64%, and 73.13% accuracy, respectively. Additionally we compared the proposed metrics to a

Table 2: Classification results on Colorado data

Class	ED	ML	MD	Sphknn-geodesic	sphknn-Mahalanobis	sphDann
c1	95	95	95	98.33	98.33	98.33
c2	0	0	0	90.9	90.9	90.9
c3	0	0	0	25	25	25
c4	28	26	26	66.67	66.67	66.67
c5	10	10	10	50	41.67	50
c6	63	63	63	78.13	78.13	78.13
c7	56	90	93	93.1	89.65	93.1
c8	0	0	0	0	0	0
c9	52	48	56	16.67	16.67	16.67
c10	0	0	10	11.11	11	11
OA	46.2	49.2	49.7	73.13	71.64	73.13
$\hat{K} - NN$	0	0	0	7	7	7

Euclidean Mahalanobis K-NN method. The results shown in Figure 2 with a varying dimension on the data indicates that the best classification results are obtained when data is embedded onto a spherical manifold and when classification carried out with any of the three proposed metrics. Figure 2 also highlights the robustness of the metrics we proposed when the dimension of the features is increased. The Euclidean K-NN method can be observed to have a degraded accuracy as the dimension increases. This is due to the bias caused by nearest neighbors that are very far in the Euclidean space. It appears that in spherical manifolds, the problem of bias is controlled and better results are achievable.

The experimental results on the AVIRIS West Lafayette image were compared to classification accuracies from methods used on the same dataset in (Lee & Ersoy, 2007). The approach in (Lee & Ersoy, 2007) combines multiple-classification results generated using different types of classifiers to form a consensual decision making procedure. It can be observed that the results obtained from the proposed sphknn-geodesic, sphknn-Mahalanobis and sphDann classifiers attain better accuracy on per class comparison with *HCL*, *HSOM*, and *HSOGR*. Take for example row 7 and row 9 from Table 3, which represents the class *grass/pasture* and *grass/pasture-mowed*, respectively. These two classes posed a 100% error rate from methods used in (Lee & Ersoy, 2007) while the spherical nearest neighbor methods achieve an accuracy of at least 33% and 45% on each class. Spherical nearest neighbor methods can be observed to achieve overall accuracies that compare to *HCL*, *HSOM*, and *HSOGR* with the spherical discriminative nearest neighbor method achieving a better *OA* result of 66.23%. The thematic maps that were generated from the spherical nearest neighbor methods are shown in Figure 4. Thematic map results gives a visual interpretation of the actual land cover results generated by the three nearest neighbor methods we presented.

5 Conclusions

In this paper, we have discussed a constant curvature nonlinear coordinate description of hyperspectral remote sensing data citing example data with a number of sources of nonlinearity such as subpixel heterogeneity and multiple scattering, bidirectional reflectance distribution function ef-

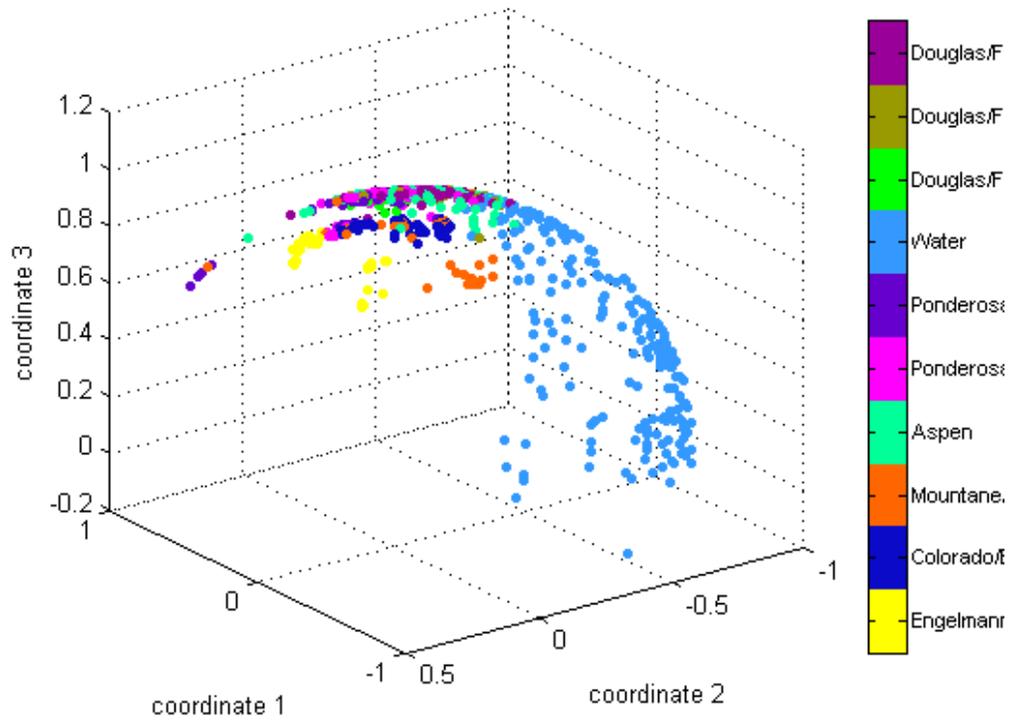


Figure 1: Embedding of Colorado 10-class dataset on a unit Sphere. Pixel coordinates color-coded on ground truth.

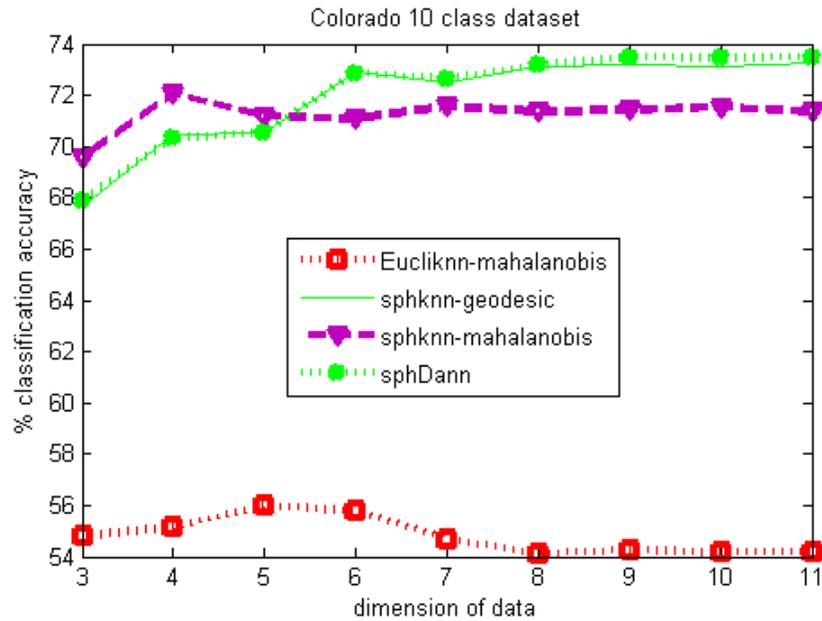


Figure 2: Classification of Colorado 10 class data set with varying dimension for $\hat{k} = 7$. Classification rates averaged over 20 runs.

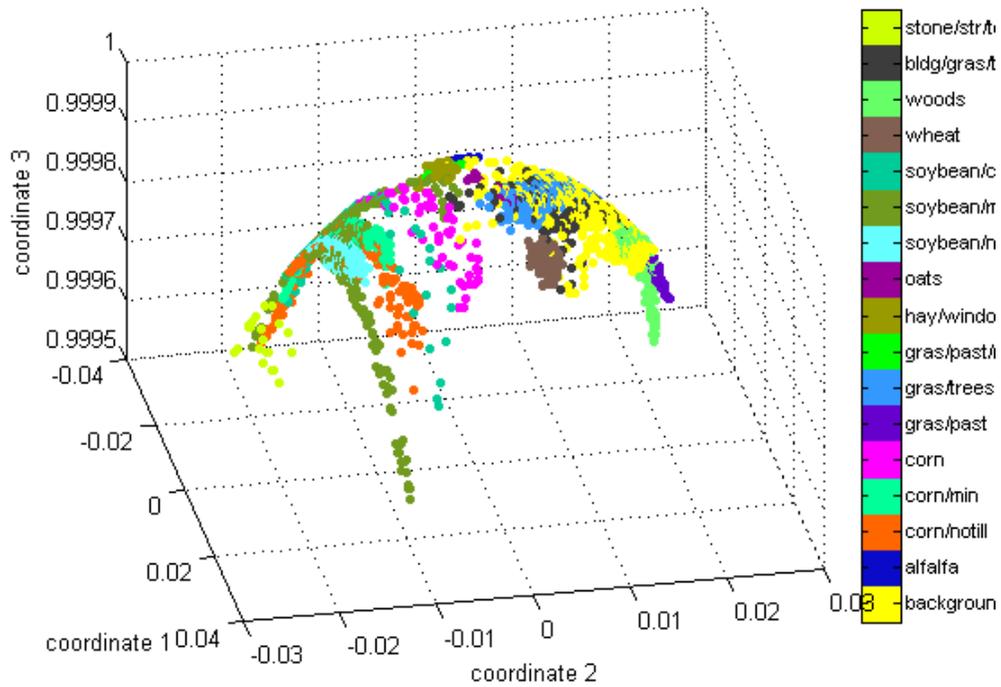


Figure 3: Embedding of AVIRIS Image features onto a unit Sphere. Pixel coordinates color-coded on ground truth.

Table 3: Classification result on AVIRIS Imagery

Class	HCL	HSOM	HSOGR	sphknn-geodesic	sphknn-Mahalanobis	sphDann
c1	68.35	92.42	97.09	54.73	55.46	56.66
c2	73.33	73.33	80.00	74.35	76.92	86.92
c3	21.41	54.82	23.53	29.58	30.69	35.69
c4	31.31	23.96	43.13	38.35	38.35	42.35
c5	42.11	22.81	75.44	41.88	44.44	44.44
c6	0.00	14.29	0.00	33.58	36.61	36.6
c7	28.68	28.68	38.97	64.05	62.59	62.59
c8	0.00	0.00	0.00	45.83	45.83	45.83
c9	100.00	100.00	100.00	98.93	98.93	98.93
c10	100.00	100.00	100.00	43.75	50.00	81.23
c11	28.72	40.00	61.54	65.13	66.08	66.09
c12	30.51	18.86	23.09	62.82	62.98	62.98
c13	35.38	39.23	36.92	36.77	39.35	39.35
c14	93.33	96.67	96.67	97.73	97.27	97.73
c15	61.80	32.02	13.48	79.73	80.46	80.46
c16	53.47	58.42	53.47	47.45	78.29	82.31
c17	80.00	80.00	80.00	72.43	79.55	90.56
OA	50.61	62.77	64.39	59.56	62.63	66.23

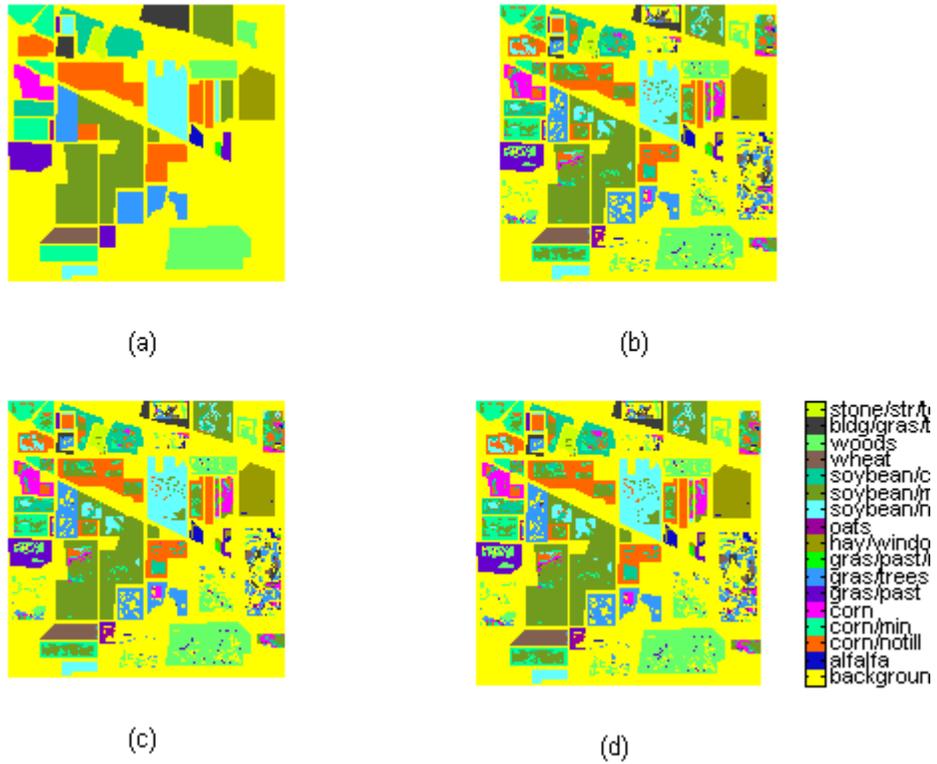


Figure 4: Thematic maps generated by the spherical nearest neighbor methods. (a) Ground reference. (b) Classification result from the sphknn-geodesic. (c) Classification result from the sph-Dann. (d) Classification result from the sphknn-Mahalanobis.

fects, and the presence of nonlinear media such as water. The direct result of such non-linearities is a fundamental limit on the ability to discriminate, for instance, spectrally similar vegetation such as forests when a linear spectral coordinate system is assumed. The experimental datasets used are very difficult to distinguish among the forest types using Landsat MSS data alone since the forest classes show very similar spectral responses to class examples such as water, and both vegetated and nonvegetated land imagery present challenges when linear spectral coordinate system is used. Using the Colorado dataset, we demonstrated that a spherical space coordinate representation coupled with novel classification rules that are tailored to incorporate the geometry of the coordinate system provides a more compact representation and discrimination of hyperspectral data than the maximum likelihood (ML), Euclidean distance (ED) and the Mahalanobis distance (MD) methods. Using the two different datasets, we showed the potential of the spherical-geodesic nearest neighbor, the spherical-Mahalanobis nearest neighbor, and the spherical-adaptive nearest neighbor methods to improve the separability of spectrally similar vegetation that typically causes false alarms when a linear coordinate system is assumed. Indications from this study suggest that multispectral features when treated as directional attributes embedded onto a unit spherical manifold increase the potential to discriminate land cover classes. In the future, we intend to build statistical models for spherical manifolds to further our investigation of directional multispectral data.

6 Acknowledgments

Dalton is thankful to the National Research Foundation of South Africa and the Oppenheimer Memorial Trust South Africa for grants received.

References

- Bachmann, C. M., Ainsworth, T. L., & Fusina, R. A. 2005. Exploiting manifold geometry in hyperspectral imagery. vol. 43.
- Benediktsson, J.A., Swain, P.H., & Ersoy, O. K. 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. *Ieee transactions on geoscience and remote sensing*, **28**(4).
- Clark, R. N., Swayze, G. A., Koch, C., Gallagher, A., & Ager, C. 1992. Mapping vegetation types with the multiple spectral feature mapping algorithm in both emission and absorption. vol. 1.
- Coifman, R.R., & Lafon, S. 2006. Diffusion maps. *Applied and computational harmonic analysis: Special issue on diffusion maps and wavelets*, **21**, 5–30.
- Cox, T. F., & Cox, M. A. A. 2001. *Multidimensional scaling*. Chapman and Hall.
- Eckart, C., & Young, G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.
- Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

- Jolliffe, I. T. 1986. *Principal component analysis*. Springer-Verlag.
- Landgrebe, D.A., & Biehl, L. 1992. 220 band hyperspectral image: Aviris image indian pine test site 3. <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/>.
- Lee, J., & Ersoy, O. K. 2007. Consensual and hierarchical classification of remotely sensed multi-spectral images. *Ieee transactions on geoscience and remote sensing*, **45**(9).
- Roweis, S. T., & Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**(5500), 2323–2326.
- Sandmeier, S. R., Middleton, E. M., Deering, D. W., & Qin, W. 1999. The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization. vol. 104.
- Tibshirani, Robert, & Hastie, Trevor. 1996. Discriminant adaptive nearest neighbor classification. *Ieee transactions on pattern analysis and machine intelligence*, **18**(6).
- Wilson, R. C., Hancock, E. R., Pekalska, E., & Duin, R. P. W. 2010 (June). Spherical embeddings for non-euclidean dissimilarities. *Pages 1903–1910 of: Cvpr conference proceedings*.