# Unsupervised Classification of Hyperspectral Images on Spherical Manifolds

Dalton Lunga and Okan Ersoy

Purdue University, West Lafayette, IN 47907-0501, USA
{dlunga,ersoy}@purdue.edu

**Abstract.** Traditional statistical models for remote sensing data have mainly focused on the magnitude of feature vectors. To perform clustering with directional properties of feature vectors, other valid models need to be developed. Here we first describe the transformation of hyperspectral images onto a unit hyperspherical manifold using the recently proposed spherical local embedding approach. Spherical local embedding is a method that computes high-dimensional local neighborhood preserving coordinates of data on constant curvature manifolds. We then propose a novel von Mises-Fisher (vMF) distribution based approach for unsupervised classification of hyperspectral images on the established spherical manifold. A vMF distribution is a natural model for multivariate data on a unit hypersphere. Parameters for the model are estimated using the Expectation-Maximization procedure. A set of experimental results on modeling hyperspectral images as vMF mixture distributions demonstrate the advantages.

**Keywords:** spherical manifolds, mixture models, directional data, hyperspectral image clustering.

## 1 Introduction

For several years, spectral unmixing techniques have been widely used for hyperspectral data analysis and quantification. Many novel applications have been developed from the unmixing point of view, including surface constituent identification for land use mapping, geology and biological process analysis[1]. Feature extraction methods in the form of *best band* combinations have been the most applied standards in such analysis. 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 [2]. 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 [3]. Although the use of endmembers and indexes based on narrowband features have yielded very useful results, these approaches largely ignore the inherent

P. Perner (Ed.): ICDM 2011, LNAI 6870, pp. 134–146, 2011.

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 [4]. 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 [4]. It also has been observed that wavelengths with the smallest reflectance exhibit the largest nonlinear variations [4]. 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. Classification of hyperspectral image data that exhibits these non-linearities poses a huge challenge to linear methods. Therefore increased better modeling of such data can be aided by use of better transformation methods. Recently, there has been ongoing work in the field of manifold learning to develop methods that capture the low dimensional embedding of high-dimensional data from which the non-linearity properties of observed data can easily be captured and incorporated into the model with all the redundant information eliminated.

Many of the manifold learning methods embed objects into a lower dimensional vector-space using techniques such as Multidimensional Scaling[5], Diffusion Maps [7], Locally Linear Embedding [8], or Principal Component Analysis [10]. Recently, a new method for embedding data onto a spherical manifold was proposed in [11]. The spherical embedding approach maps the dissimilarity of shape objects onto a constant curvature spherical manifold. It embeds data onto a metric space while optimizing over the kernel distance matrix of positional vectors. Each of these approaches represents an attempt to derive a coordinate system that resides on (parameterizes) the nonlinear data manifold itself. The methods represent 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 characterized by their embedding co-ordinate vectors, and analyzed in a conventional manner using traditional tools. Models can be developed for the low dimensional embedded data. However, 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.

In this paper we exploit the nonlinear structure of hyperspectral imagery using the spherical embedding method as a feature transformational tool. The approach seeks a constant curvature coordinate system that preserves geodesic distances in high-dimensional hyperspectral feature spaces. A With data embedded onto a spherical manifold, modeling techniques can now be developed. We first outline the intuition and motivation explaining why a spherical manifold is relevant for remote sensing data. Traditional supervised and unsupervised

classification algorithms involve multivariate data that is drawn from $\mathbb{R}^d$ with all emphasis attached to the magnitude of the feature vectors while the directional element of the feature vectors is usually not considered. For some non-linearities observed in remote sensing imagery data,e.g. presence of water in pixels viewed as a function of position in the landscape, it makes sense to transform the observed data onto manifolds on which the coordinate system allows for the directional nature of the features to be significant. It has been observed that for most high-dimensional remote sensing feature vectors, the *cosine similarity* measure which is a function of an angle between a pair of vectors, performs better than the *Euclidean distance* metric [13]. Such an observation suggests pursuing a directional model for hyperspectral images. With the above insight, we develop a novel von Mises-Fisher (vMF) distribution based approach for unsupervised classification of hyperspectral images on spherical manifolds. This is an approach for unsupervised classification of embedded hyperspectral data based on a mixture model, where the distribution of the entire data is considered to be a weighted summation of the von-Mises Fisher class conditional densities. The vMF distribution is a generalization of the von Mises distribution to higher dimensions [15,16]. This distribution arises naturally for directional data with few parameters requiring estimation.

The main aim of this study is to introduce spherical manifolds to remote sensing data using the spherical embedding approach and also to propose a model for identifying cluster components of similar land cover usage. Unsupervised classification of AVIRIS data is performed with each pixel allocated a class label with the highest posterior probability. Cluster components are mapped to corresponding classes using the best permutation mapping obtained from the Kuhn-Munkres algorithm [6]. In the next section, we first discuss the embedding space and the method of transforming hyperspectral images to a constant curvature manifold. We then present the model based clustering on a spherical manifold. Experimental results are provided with discussions on why spherical manifolds with neighborhood preserving properties have a potential impact on future models for hyperspectral images. The last section concludes with a brief discussion and ideas future work.

## 2 Spherical Embedding of Image Pixels

A spherical manifold defines the geometry of a constant curvature surface. The spherical embedding procedure we apply has neighborhood preserving properties meaning that transformed feature vectors of similar pixel vectors are embedded in the neighborhood of each other. The outline of the embedding algorithm as recently proposed in [11], is shown in Figure 1. In the following sections, we first set up the Bayes rule for a single component model based approach for classifying image pixels on a spherical surface and then we will apply the same rule to a spherical mixture model for image pixels.

**Input:** Dissimilarity matrix $D_{n\times n}$, where $n$ is the number of pixels.
**Output:** $X^\star$, whose rows are pixel coordinates and whose inner-product $X^\star X^{\star T}$ has the same neighborhood as $D$. Procedure:

1. If the spherical point positions are given by $X_i$, $i = 1, \cdots, n$, then $\langle X_i, X_j\rangle = r^2 \cos\beta_{ij}$, with $\beta_{ij} = \frac{d_{ij}}{r}$.
2. If $X$ in unknown, compute for $X$ such that $XX^T = Z$, where $Z_{ij} = r^2 \cos\beta_{ij}$ and $d_{ij} \in D$. Find the radius of sphere as $r^\star = \arg\min_r \lambda_1\{Z(r)\}$. $\lambda_1$ is the smallest eigenvalue of $Z(r)$.
3. Set $\hat{Z} = \frac{Z}{r^\star}$ and $X^\star = \arg\min_{X, x^T x=1} \|XX^T = \hat{Z}\|$
4. Decompose $\hat{Z}$, $\hat{Z} = U\Lambda U^T$. Set the embedding positional matrix to be $X^\star = U_{n\times k}\Lambda^{1/2}_{k\times k}$, where $k$ is chosen such that the elements of $U_{n\times k}$ corresponds to the largest $k$ eigenvalues of $\Lambda_{k\times k}$.

**Fig. 1.** Outline of Spherical Embedding

## 3 von Mises-Fisher Model and Bayes Rule

The Bayes rule approach to supervised classification is a fundamental technique, and it is recommended as a starting point for most pattern recognition applications. The rule bases its classification in terms of probabilities. As such all probabilities must be known or estimated from the data. We adapt this rule and apply it on data that has been mapped to a spherical manifold. Traditional Gaussian models cannot be applied on spherical manifolds as the properties of the data have been manipulated to have a unit magnitude while the feature angles are different. The analysis of such data will require models that can only depend on the direction of the vectors and not their magnitudes. Such models for handling directional data have been used in literature [15]. We assume that each embedded pixel vector was generated from a von Mises-Fisher distribution.

Given directional data sample $\{x_i\}_{i=1}^n$ such that each $x_i$ has the property, $\|x\| = 1$, that is, $x_i \in \mathbb{S}^{d-1}$, with $\mathbb{S}$ a unit hypersphere of dimension $(d-1)$, the assumed corresponding von Mises-Fisher density is defined by

$$f(x|\mu, \kappa) = \frac{\kappa^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} I_{\frac{d}{2}-1}(\kappa)} \exp\{\kappa\mu^T x\} \tag{1}$$

where $I_r(\cdot)$ denotes the modified Bessel function of the first kind and order $r$. The parameters $\mu$ and $\kappa$, denotes the mean direction and concentration parameter of the distribution, respectively. The greater the value of $\kappa$, the higher the concentration of the distribution around the mean direction $\mu$. The distribution is uni-modal for $\kappa > 0$, and is uniform on the sphere for $\kappa = 0$. The posterior probability for choosing class membership is defined by

$$P(c_j|x_0),\ j = 1, \ldots, J \tag{2}$$

The above equation describes the probability that the test vector belongs to the j-th class given the observed feature vector $x$. Making use of the Bayes' Theorem we can find the posterior probabilities by

$$P(c_j|x_0) = \frac{P(c_j)P(x_0|c_j)}{P(x_0)} \tag{3}$$

where $P(x_0) = \sum_{j=1}^{J} P(c_j)P(x_0|c_j)$. $P(c_j)$ can be inferred from prior knowledge of the application, estimated from the data by defining it to be $P(c_j) = \frac{N_j}{N}$, where $N_j$ is the number of training samples with class label $j$ and $N$ is the total number of training samples. The class conditional $P(x_0|c_j)$ represents the probability distribution of the features of each class. Thus, for parametric density estimation, one has to assume a form of distribution for the class conditionals and then proceed to estimate the parameters for that distribution. As noted above, we have made the assumption that the feature vectors were generated from a von Mises-Fisher distribution. The next task is then to estimate the parameters of a von Mises-Fisher distribution for each class of the labeled data.

### 3.1 Maximum Likelihood Estimation

Maximum likelihood estimation on a spherical manifold is simply carried out in a conventional manner, i.e given a sample space $\mathcal{X}$ of unit random pixel vectors drawn independently according to $f(x|\mu, \kappa)$, the likelihood of the sample space is given by

$$L(\mathcal{X}|\mu, \kappa) = \prod_{i=1}^{n} f(x_i|\mu, \kappa) \tag{4}$$

We can write the above in the log-likelihood form to get

$$\log L(\mathcal{X}|\mu, \kappa) = n \log c_d\kappa + n\kappa\mu\bar{x} \tag{5}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $c_d\kappa = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}$. To obtain the maximum likelihood estimates of $\mu$ and $\kappa$, we maximize equation (5) subject to the constraint $\mu^T\mu = 1$ and $\kappa \geq 0$. For a classification task, we consider the training instances of each class separately in estimating the model parameters. Given $j = 1, \ldots, J$ classes, the derivations of the $MLE$ solutions $\hat{\mu}_j$ and $\hat{\kappa}_j$ for each class conditional are given by

$$\hat{\mu}_j = \frac{\bar{x}_j}{\|\bar{x}_j\|} \tag{6}$$

and

$$A(\hat{\kappa}_j) = \frac{I_{d/2}(\hat{\kappa_j})}{I_{d/2-1}(\hat{\kappa_j})} = \bar{x}_j \tag{7}$$

where $\bar{x}_j = \sum_{x_i \in C_j} x_i$ and $\|\bar{x}_j\|$ is the length of the average resultant vector for class $j$. A closed form solution of equation (7) is not possible and one can use numerical techniques to solve for $\hat{\kappa}_j$. A reasonable approximation to the solution is obtained by following the approach used in [16] from which $\hat{\kappa}_j$ is set to $\frac{\|\bar{x}_j\| d - \|\bar{x}_j\|^3}{1 - \|\bar{x}_j\|^2}$.

**Bayes Decision Rule:** Given a feature vector $x \in \mathbb{S}^{d-1}$, we assign it to class $c_j$ if:

$$P(c_j|x) > P(c_k|x); \quad k = 1, \ldots, J; \; k \neq j. \tag{8}$$

That is, we classify an observation $x$ as belonging to the class that has the highest posterior probability. In the next section, we consider a case where class labels are not available for all sample observations.

## 4 Mixture of von Mises-Fisher Model

When the data sample space is considered to be incomplete due to the absence of class labels, it is not so easy to make an assumption that each sample belongs to a specific model. So a commonly used approach is to consider that the observed samples, $\{x_i\}_{i=1}^N$, were generated from a mixture of $J$ components and each component corresponds to a class which is modeled by a probability distribution that is a member to the assumed family of distributions. We make the assumption that the directional data samples are generated from a mixture of von Mises-Fisher models $f(x_i|\theta_j)$, each with parameter vector $\theta_j = (\mu, \kappa)$ for $1 \leq j \leq J$. A mixture of von Mises-Fisher has a joint density of the form

$$f(x_i|\Theta) = \sum_{j=1}^{J} \alpha_j f_j(x_i|\theta_j) \tag{9}$$

where $\Theta = \{\alpha_1, \ldots, \alpha_J, \theta_1, \ldots, \theta_J\}$ and the $\alpha_j$'s are constrained to $\sum_{j=1}^J \alpha_j = 1$. For a given embedded hyperspectral image we let $\mathcal{X}=\{x_1, \ldots, x_n\}$ be the set of spherical pixel vectors, with each vector sampled according to equation (9). Let $\mathcal{Y}=\{y_1, \ldots, y_n\}$ be the corresponding set of latent variables with each $y_n \in \{1, \ldots, J\}$. For example, $y_i = j$ if $x_i$ is sampled from $f_j(\cdot|\theta_j)$. The log-likelihood of the observed embedded pixel vectors is a random quantity given by

$$\log P(\mathcal{X}, \mathcal{Y}|\Theta) = \sum_{i=1}^{n} \log \alpha_{y_i} f_{y_i}(x_i|\theta_{y_i}) \tag{10}$$

Obtaining the maximum likelihood parameters of the above expression would have been easy if the values of $y_i$ were known just like in the case of supervised classification of section 3.2. Since the label $y_i$ for each coordinate pixel $x_i$ is

unknown, the solution to the derivatives of equation (10) can be found using the expectation-maximization (EM) algorithm [14].

On the $(t+1)^{th}$ iteration of the EM algorithm, the E step is equivalent to replacing the unobserved random quantities in $\mathcal{Y}$ by their current conditional expectations, which are the current conditional probabilities of $\mathcal{Y} = j$ given $\mathcal{X} = x_i$:

$$p_{ij}^{(t)} = \frac{\alpha_j^{(t)} f(x_i; \theta_j^{(t)})}{\sum_{k=1}^{J} \pi_k^{(t)} f(x_i; \theta_k^{(t)})} = p(\mathcal{Y} = j | \mathcal{X} = x_i; \theta) \tag{11}$$

with $1 \leq i \leq n;\ 1 \leq j \leq J$.

The M step requires finding the value of $\Theta$ at the $(t+1)$ iteration. Thus $\Theta^{(t+1)}$ would be the value that globally maximizes the objective function

$$Q(\Theta, \Theta^{(t)}) = \sum_{\mathcal{Y}} p(\mathcal{Y}|\mathcal{X}, \Theta^{(t)}) \ln p(\mathcal{X}, \mathcal{Y}|\Theta)$$

Thus, in the M step, the quantity that is being maximized is the expectation of the complete-data log likelihood. This effectively requires the calculation of the component distribution maximum likelihood estimates. The updated component parameter estimates for the $(t+1)$ iteration, $\theta_j^{(t+1)}$, are obtained by solving the weighted log-likelihood equation

$$\sum_{i=1}^{n} p_{ij}^{(t)} \partial \log f(x_i; \theta_j) / \partial \theta_j = 0. \tag{12}$$

After applying calculus to this equation, we obtain the following required update parameters for each cluster component distribution:

$$\alpha_j = \frac{1}{n} \sum_{i=1}^{n} p(j|x_i, \Theta), \quad \hat{\mu}_j = \frac{\bar{x}_j}{\|\bar{x}_j\|}, \tag{13}$$

$$A(\hat{\kappa}_j) = \frac{I_{d/2}(\hat{\kappa}_j)}{I_{d/2-1}(\hat{\kappa}_j)} \Rightarrow \hat{\kappa}_j = A^{-1}(\|\bar{x}_j\|) \tag{14}$$

$$= \frac{\|\bar{x}_j\| \, d - \|\bar{x}_j\|^3}{1 - \|\bar{x}_j\|^2} \tag{15}$$

where

$$\|\bar{x}_j\| = \frac{\|\sum_{i=1}^{n} x_i p(j|x_i, \Theta)\|}{\sum_{i=1}^{n} p(j|x_i, \Theta)}, \ p(j|x_i, \Theta) = \frac{\alpha_j f_j x_i}{\sum_{k=1}^{K} \alpha_k f_k x_i} \tag{16}$$

The maximum likelihood estimates ensure that the inequality

$$Q(\Theta^{(t+1)}; \Theta^{(t)}) \geq Q(\Theta^{(t)}; \Theta^{(t)})$$

is true for each $\Theta^{(t+1)}$. This is sufficient to ensure that the likelihood is not decreased.

## 5 Experiments

We consider a random unit vector $X$, whose elements are positional coordinates of the intensity values (pixel bands values) of a pixel sample from the corresponding spectral bands of a hyperspectral image. The randomness in the vector is introduced by physical, scattering effects and atmospheric features. As such, it makes sense to consider the physical properties of an area as being characterized more by the distribution of the vector of directional positional intensities than by the magnitude of the vector. We make the assumption that sample directional unit pixel positional vectors were generated by selecting the class $c_j$, with prior probability $\alpha_j$ and then selecting $X$, according to $f(X|\theta_j)$ so that the mixture model derived in (9) can be applied.

### 5.1 Data

**AVIRIS Hyperspectral West Lafayette 1992 Image:-** To establish the effectiveness of the proposed hyperspectral feature transformation onto spherical manifold, and the application of the proposed mixture model, we generate results from the AVIRIS multispectral image. The West Lafayette image was used in the experiments. This data is a multispectral image that was obtained from the Airborne/Infrared Imaging Spectrometer that was built by Jet Propulsion Laboratory and flown by NASA/Ames on June 12, 1992 [12]. 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 $0.663 - 0.673$ $\mu$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 $145 \times 145$ pixels. The pixel resolution is 16 bits, corresponding to 65536 gray levels. 3403 pixels were selected to generate the ground-reference data. For the experiments, each sample pixel is of dimension 81 consisting of the pixel's values from the 9-bands and the 9-bands values for each of its 8 neighbors. In Figure 2, we show the actual land cover usage from the AVIRIS image together a subset of the land cover cosine coordinates for with each pixel embedded onto a spherical manifold. Where we chose the embedding space to be a 2-dimensional sphere for representational purposes.

**AVIRIS Hyperspectral Tippecanoe County Image 1986:-** This is a small segment (169 lines x 169 columns of pixels) of a Thematic Mapper scene of Tippecanoe County, Indiana gathered on July 17, 1986 [12]. The subset consist of 7 bands of a significantly 220 bands. The image has 7 classes (background, corn, soybean, wheat, alfalfa/oats, pasture, and sensor/distortion). Two thousand pixels were selected to generate the ground-reference data. For the experiments, each sample pixel is of dimension 63 consisting of the pixel's values from the 7-bands and the 7-bands values for each of its 8 neighbors. In Figure 3, we show the actual land cover usage from the AVIRIS image and a subset of land cover cosine coordinates for test pixel embedded onto a spherical manifold.

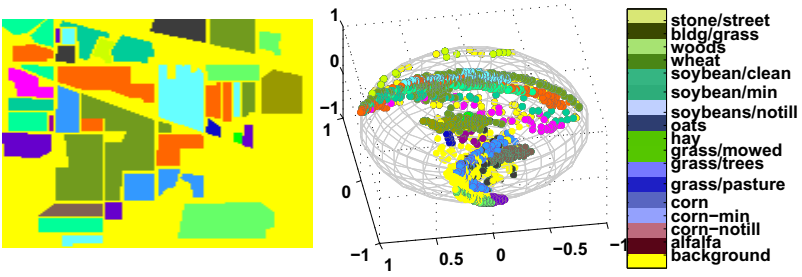


**Fig. 2.** (Left)-AVIRIS 1992 West Lafayette land cover usage, color coded on ground truth. (Right)-corresponding cosine pixel coordinates on a spherical manifold.

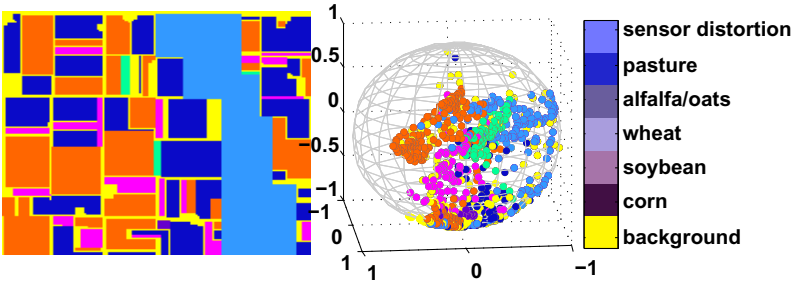


**Fig. 3.** (Left)-AVIRIS 1986 Tippecanoe County land cover usage color coded on ground truth. (Right)-corresponding cosine pixel coordinates on a spherical manifold.

### 5.2 Results

To evaluate the performance of the von Mises-Fisher mixture model on hyperspectral data clustering, we use a metric-accuracy proposed in [17]. The dataset consist of $N$ samples, all with labeled clusters. With each sample's predicted cluster label denoted $t_i$ and the corresponding ground truth labelled $g_i$, the clustering accuracy is defined by

$$accuracy = \frac{\sum_{i=1}^{N} \delta(g_i, map(t_i))}{N} \tag{17}$$

where $\delta(g_i, map(t_i))$ is a delta function equal to 1 if the label $g_i$ is equal to the label $t_i$, otherwise it is 0. The function $map(t_i)$ is the best permutation mapping obtained from the Kuhn-Munkres algorithm [6]. The function maps the predicted cluster labels to the corresponding best permuted representational cluster.

The clustering accuracy of the proposed von Mises-Fisher mixture model is compared to the results obtained using the spherical K-means algorithm [18]. It can be seen from Table 1 that both methods achieve above random guessing accuracy when classes are well separated. This indicates that when a hyperspectral image is embedded onto a spherical manifold, pixel vectors with similar properties tend to have directional properties that are related. The non-linearities

observed in images with water medium results in pixel vectors following a particular directional distribution. For both methods higher accuracy was observed for fewer cluster components. We make a note to compare our results with those obtained in [9], from which the authors used an independent component mixture model to study the same data set but for only four clusters. We applied our proposed method to a small subset image with four clusters and observe the clustering accuracy to be 67%. This value is 7% above the best accuracy value which was reported in [9] for the same ground truth. This indicates that our proposed method has additional capability to carry out better classification over an independent component analysis(ICA) mixture model. In order to give a further quantitative performance evaluation of the proposed model, we collected 2000 cosine pixel coordinates from the spherically mapped Tippecanoe County image. With the selected pixels coordinates, confusion matrix was built based on the relationship between the mapping obtained from the Kuhn-Munkres algorithm [6] and the ground-truth labels shown in Figure 3. The statistical accuracies are shown in Table 2. The mixture model exhibited better accuracy on clustering the pixel coordinates.

The accuracy is however sensitive to an introduction of new sample points from cluster components with overlapping structures. In Figure 4, we show a result of AVIRIS-West Lafayette image clustering accuracy degrading with the introduction of new cluster components. This artifact could be expected from most unsupervised learning methods. The argument being that as more and more overlapping structures are introduced, sample points that are located at the cluster component boundaries are more likely to present more ambiguity as to which cluster they belong to, as a result degrading the performance of the algorithm. However, the results clearly supports a motivation for exploring a new coordi-

**Table 1.** Clustering accuracy(%)- AVIRIS 1992 Indian Pine Site

| number of clusters | spherical-Kmeans | von Mises-Fisher mixture |
|---|---|---|
| 2 | 55.10 ± 0.3 | 68.35 ± 0.1 |
| 3 | 68.72 ± 2.3 | 88.23 ± 3.2 |
| 4 | 58.06 ± 0.1 | 75.33 ± 1.8 |
| 5 | 46.52 ± 0.1 | 68.16 ± 0.6 |
| 6 | 49.77 ± 2.7 | 67.90 ± 1.6 |
| 7 | 51.19 ± 1.4 | 63.16 ± 1.89 |
| 8 | 49.05 ± 0.9 | 64.40± 2.06 |
| 9 | 50.95 ± 0.8 | 63.31 ± 3.8 |
| 10 | 50.57 ± 0.5 | 59.91 ± 2.4 |
| 11 | 48.20 ± 1.6 | 60.9 ± 2.8 |
| 12 | 48.10 ± 1.1 | 54.74 ± 0.8 |
| 13 | 48.96 ± 1.2 | 53.12 ± 0.4 |
| 14 | 47.66 ± 1.16 | 50.53 ± 0.1 |
| 15 | 45.35 ± 0.5 | 48.53 ± 0.8 |
| 16 | 45.37 ± 0.6 | 46.53 ± 0.7 |
| 17 | 43.88 ± 0.4 | 45.07 ± 0.5 |
| Avg accuracy | 50.25 | 61.38 |

**Table 2.** Clustering accuracy(%)- AVIRIS 1986 Tippecanoe County

| number of clusters | spherical-Kmeans | von Mises-Fisher mixture |
|---|---|---|
| 2 | 77.46 ± 0.1 | 75.89 ± 0.3 |
| 3 | 73.77 ± 1.5 | 76.22 ± 1.7 |
| 4 | 54.65 ± 0.03 | 68.95 ± 3.1 |
| 5 | 44.77 ± 0.2 | 66.47 ± 0.7 |
| 6 | 39.33 ± 1.5 | 63.19 ± 2.3 |
| 7 | 38.13 ± 1.1 | 55.81 ± 1.0 |
| Avg accuracy | 46.87 | 58.08 |

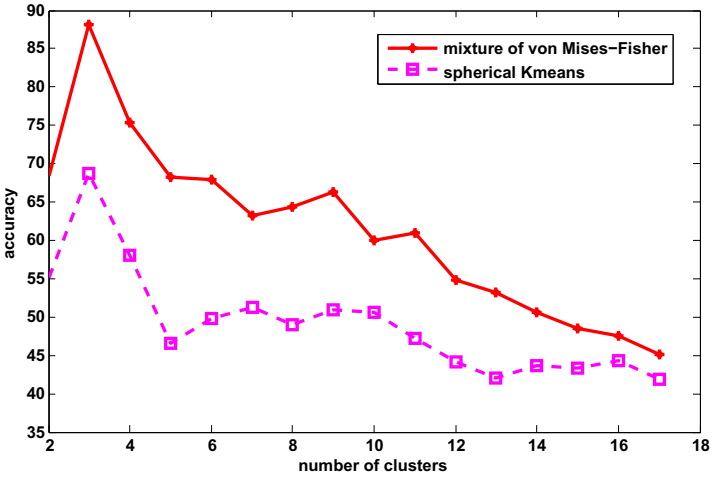


**Fig. 4.** Clustering accuracy on AVIRIS-West Lafayette Image

nate space from which to model hyperspectral images. As we mentioned earlier, a von Mises-Fisher distribution is similar to a constrained covariance Gaussian distribution. As such, all cluster components are constrained to have constant concentric countour shapes. The overall result of modelling hyperspectral image pixels as cosine spherical coordinates using a mixture of von Mises-Fisher model appears to fit the clusters with an oval shape inaccurately. Elliptic or oval shaped cosine coordinate data can be better modelled using Kent distributions [15]. The advantage of the Kent distribution over the von Mises-Fisher distribution on a spherical manifold is that the equal probability contours of the density are not restricted to be circular, they can be elliptical as well. Our future goal is to explore such models and their impact on spherically embedded remote sensing images.

## 6 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 effects 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. Our approach was to seek a constant curvature manifold on which hyperspectral images could be represented by their angle information and proceed to develop an unsupervised algorithm for analysis of the data. The motivation of using cosine coordinates was due to observing the success of the cosine similarity metric in image retrieval systems in Euclidean spaces. We have proposed a novel approach derived from embedding hyperspectral images onto a spherical manifold using the spherical embedding method. The approach models embedded image pixels as random directional quantities generated from a mixture of von Mises-Fisher distributions. The results presented indicate the benefits of seeking spherical coordinates for analysis of hyperspectral images.

**Acknowledgements.** The authors would like to thank the reviewers for helpful comments that improved the final draft of this paper. We also acknowledge inputs and insights from Dr Sergey Kirshner on related topics during the course of the study. Dalton has previously been supported by Fulbright, the National Research Foundation of South Africa and The Oppenheimmer Memorial Trust.

## References

1. Plaza, A., Martínez, P., Perez, R.M., Plaza, J.: A comparative analysis of endmember extraction algorithms using AVIRIS hyeperspectral imagery. Summaries of the 11th JPL Airborne Earth Science Workshop (2002)
2. Clark, R.N., Swayze, G.A., Koch, C., Gallagher, A., Ager, C.: Mapping vegetation types with the multiple spectral feature mapping algorithm in both emission and absorption, vol. 1, pp. 60–62 (1992)
3. Bachmann, C.M., Ainsworth, T.L., Fusina, R.A.: Exploiting manifold geometry in hyperspectral imagery, vol. 43, pp. 11–14 (2005)
4. Sandmeier, S.R., Middleton, E.M., Deering, D.W., Qin, W.: The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization, vol. 104, pp. 9547–9560 (1999)
5. Cox, T.F., Cox, M.A.A.: Multidimensional Scaling. Chapman and Hall, Boca Raton (2001)
6. Lovasz, L., Plummer, M.D.: Matching Theory (1986)
7. Coifman, R., Lafon, S.: Diffusion maps. Applied and Computational Harmonic Analysis: Special issue on Diffusion Maps and Wavelets 21, 5–30 (2006)
8. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
9. Shah, C., Arora, M.K., Robila, S.A., Varshney, P.K.: ICA mixture model based unsupervised classification of hyperspectral imagery. In: Proceedings of the 31st Applied Imagery Pattern Recognition Workshop (2002)
10. Jolliffe, I.T.: Principal Component Analysis. Springer, Heidelberg (1986)
11. Wilson, R.C., Hancock, E.R., Pekalska, E., Duin, R.P.W.: Spherical Embeddings for non-Euclidean Dissimilarities. Comp. Vis. and Patt. Recog. 1903-1910 (2010)

12. Landgrebe, D.A., Biehl, L.: 220 Band Hyperspectral Image: AVIRIS image Indian Pine Test Site 3, Purdue University, West Lafayette, School of Engineering, http://www.dynamo.ecn.purdue.edu/~biehl/MultiSpec/
13. Bao, Q., Guo, P.: Comparative Studies on Similarity Measures for Remote Sensing Image Retrieval. In: IEEE International Conference on Systems, Man and Cybernetics (2004)
14. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1), 1–38 (1977)
15. Mardia, K.V., Jupp, P.: Directional Statistics, 2nd edn. John Wiley, Chichester (2000)
16. Dhillon, I.S., Sra, S.: Modeling Data using Directional Distributions. Technical Report # TR-03-06 (2003)
17. Xu, W., Liu, X., Gong, Y.: Document clustering based on non-negative matrix factorisation. In: SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 267–273 (2003)
18. Dhillon, I.S., Mudha, D.S.: Concept decompositions for large sparse text data using clustering. Machine Learning 42(1), 143–175 (2001)
19. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychom. 1, 211–218 (1936)