1-4-2011

Kent Mixture Model for Hyperspectral Clustering via Cosine Pixel Coordinates on Spherical Manifolds

Dalton Lunga
Electrical and Computer Engineering, Purdue University

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

Kent Mixture Model for Hyperspectral Clustering via Cosine Pixel
Coordinates on Spherical Manifolds

Dalton Lunga

Okan Ersoy

School of Electrical and Computer Engineering

1285 Electrical Engineering Building

Purdue University

West Lafayette, IN 47907-1285

Kent Mixture Model for Hyperspectral Clustering via Cosine Pixel Coordinates on Spherical Manifolds

Dalton Lunga and Okan Ersoy

School of Electrical Engineering

Purdue University

West Lafayette, IN 47907-0501

USA

`{dlunga,ersoy}@purdue.edu`

January 4, 2011

Abstract

Modern hyperspectral imaging sensor technology provides detailed spectral and spatial information that enables precise analysis of land cover usage. From a research point of view, traditional widely used statistical models are often limited in the sense that they do not incorporate some of the useful angle information contained in the feature vectors, and hence alternative modeling methods are required. In the study to be presented, the use of cosine angle information and its embedding onto a spherical manifold is investigated. The transformation of hyperspectral images onto a unit hyperspherical manifold is achieved by using the recently proposed spherical local embeddings approach. Spherical local embeddings is a method that computes high-dimensional local neighborhood preserving coordinates of data on constant curvature manifolds. We further develop a novel Kent mixture model for unsupervised classification of embedded cosine pixel coordinates. A Kent distribution is one of the natural models for multivariate data on a spherical manifold. Parameters for the model are estimated using the Expectation-Maximization procedure. The mixture model is applied to two different Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) datasets that were acquired from the Tippecanoe County in Indiana. The results obtained present insights on cosine pixel coordinates and also serve as a motivation for further development of new models to analyze hyperspectral images in spherical manifolds.

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(Plaza *et al.*, 2002). 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

1

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 indices 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 also 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 sub pixel constituents. Classification of hyperspectral image data that exhibits these non-linearities poses a huge challenge to linear methods. 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 embeddings of high-dimensional data from which the nonlinear properties of observed data can 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(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 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 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. A new approach to embedding similar objects has been recently proposed. The *spherical local embeddings (SLE)* method maps the dissimilarities of feature vectors onto a constant curvature manifold (Wilson *et al.*, 2010). SLE is a method that embeds indefinite data onto a non-Euclidean, but metric space optimizing over the kernel distance matrix of positional vectors.

In this paper we exploit the nonlinear structure of hyperspectral imagery using the SLE method

as a feature transformational tool. The approach seeks a constant curvature coordinate system (Riemannian manifold) that preserves geodesic distances in the high-dimensional hyperspectral feature space. A Riemannian manifold is curved, and the geodesic distances are metric. With data embedded onto a spherical manifold, modeling techniques can now be developed. We first outline the intuition and motivation explaining why a constant curvature manifold is relevant for remote sensing data.

Traditional supervised and unsupervised classification algorithms involve multivariate data drawn from $\mathbb{R}^d$ with most emphasis attached to the magnitude of the feature vectors while the directional element of the feature vectors is usually not sufficiently considered. For some non-linearities observed in remote sensing imagery data, e.g. presence of water, 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 more significant. It has been observed that for most high-dimensional remote sensing feature vectors, the *cosine similarity* measure which is the function of an angle between a pair of vectors, performs better than the *Euclidean distance* metric (Bao & Guo, 2004). Such an observation suggests pursuing a directional model for hyperspectral images. With the above insight, we develop a novel Kent 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 Kent class conditional densities. The Kent distribution can be thought of as a generalization of the Fisher distribution (Mardia & Jupp, 2000),(Kent, 1982). The distribution is comparable to a bivariate normal distribution where the covariance matrices are unconstrained. This allows for distributions of any elliptic shape, size, and orientation on the surface of a sphere. Thus, the Kent distribution is more appealing for clustering of directional data since it provides more flexibility than the von Mises distributions (Mardia & Jupp, 2000),(Kent, 1982).

Our main aim is to introduce constant curvature manifolds to remote sensing data using the SLE method and then 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 (Lovasz & Plummer, n.d.). 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 future work.

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 goemetry that exist in Riemannian manifolds dictates 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 *goedesic*). 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 $u_1, u_2, \ldots, u_d$. This is usually related to an infinetesimal distance element in the space by

$$ds^2 = \sum_{ij} g_{ij} du_i du_j \tag{1}$$

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. This 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

$$x = (r \sin u \sin v, r \cos u \sin v, r \cos v)^T \tag{2}$$

A spherical embedding implies a metric tensor of the form

$$ds^2 \quad = dx^2 + dy^2 + dz^2 \tag{3}$$
$$= r^2 sin^2 v du^2 + r^2 dv^2 \tag{4}$$

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

$$\sum_i x_i^2 = r^2 \tag{5}$$

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 substended by the two points at the center of the hypersphere is θ_{ij}, then the distance between them is given by

$$d_{ij} = r\theta_{ij} \tag{6}$$

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 $< x_i, x_j >= r^2 \cos\theta_{ij}$ we obtain

$$d_{ij} = r \cos^{-1}\left(\frac{< x_i, x_j >}{r^2}\right) \tag{7}$$

4

2.2 Cosine Coordinates on 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 $x_i, \; i = 1, \ldots, n$, then

$$< x_i, x_j > = r^2 \cos \theta_{ij} = r^2 \cos(\frac{d_{ij}}{r}) \tag{8}$$

From (8) 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

$$XX^T = Z \tag{9}$$

where $Z_{ij} = r^2 \cos(\frac{d_{ij}}{r})$ 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 the authors in (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, $r^\star$ is determined by minimizing the magnitude of the smallest eigenvalue as a function of r:

$$r^\star = argmin_r |\lambda_1 \{Z(r)\}| \tag{10}$$

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

$$Z(r^\star) = U \Lambda U^T \tag{11}$$

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

$$X = U \Lambda^{1/2} \tag{12}$$

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.

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

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

Having the singular values in decreasing order $\sigma_1 \geq \sigma_2 \geq \ldots$, 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.

3 Polar Coordinates

Restricting the rank approximation to the first three singular values of the positional matrix from the previous section, we obtain sample points that lie on the surface of a unit sphere. These sample points can be further represented using polar coordinates. The polar coordinates are denoted by (γ_1, γ_2) $(0 \leq \gamma_1 \leq 2\pi, \ 0 \leq \gamma_2 \leq \frac{\pi}{2})$, where γ_1 represents the dip angle and γ_2 is the dip direction (Kent, 1982),(McLachlan & Peel, 2000). The polar coordinates can be obtained from the embedded cosine pixel coordinates by

$$\gamma_1 = \arctan(x_3/x_2), \ \gamma_2 = \arccos(x_1) \tag{15}$$

Following the approach in (McLachlan & Peel, 2000), the polar coordinates are used later for simplifying the parameter estimation of the model.

4 Mixture of Kent Distributions

We now consider a mixture of J Kent distributions that serves a generative model for the cosine pixel coordinates. We also present a convenient method for update equations in estimating the mixture-density parameters from the embedded pixel coordinates using the Expectation Maximation procedure. Kent model based approach has been considered in the context of mining rock fracture studies in (McLachlan & Peel, 2000)(pp 287). In (McLachlan & Peel, 2000), data samples were represented as dip angle and dip direction for different sites. Kent component distributions were then used to identify joint sets in the fractured data. Although the motivation of using directional coordinates may parallel a lot of work in directional statistics, our work is motivated by

the capability of embedding higher dimensional pixel coordinates in a lower dimensional spherical manifold on which the neighborhood relations of pixels are preserved and characterized by the cosine distances.

The mixture model approach assumes each embedded coordinate pixel is a realization of the random $d-$dimensional vector $\mathcal{X}$, which has a probability density given by

$$f(x; \Theta) = \sum_{j=1}^{J} \pi_j f_j(x; \theta_j) \tag{16}$$

where the mixing proportions $\pi_j \geq 0$ and $\sum_j \pi_j = 1$, $\Theta = \{\alpha_1, \ldots, \alpha_J, \theta_1, \ldots, \theta_J\}$ and $f_j(x; \theta_j)$ is denotes the probability density of the jth component which is denoted by

$$f(x; \theta_j) = Z \cdot \exp\left\{ \kappa(x^T \xi_1) + \beta(x^T \xi_2)^2 - \beta(x^T \xi_3)^2 \right\}, \tag{17}$$

where

$$Z \approx \frac{\exp(-\kappa)\sqrt{\kappa^2 - 4\beta^2}}{2\pi}$$

and $\theta = (\kappa, \beta, \xi_1^T, \xi_2^T, \xi_3^T)^T$ is the parameter vector. The parameter ξ_1^T is the directional mean of the component distribution, while β, ξ_2^T and ξ_3^T relates to the shape and orientation of the component distribution, respectively.

For a given cosine coordinate data, we let $\mathcal{X}=\{x_1, \ldots, x_n\}$ be the set of feature vectors sampled according to equation (17). 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 $f_j(\cdot|\theta_j)$.

Since the values in the set $\mathcal{Y}$ are unknown, the log-likelihood of the observed features is a random quantity given by

$$\log P(\mathcal{X}, \mathcal{Y}|\Theta) = \sum_{i=1}^{n} \log \alpha_{y_i} f_{y_i}(x_i|\theta_{y_i}). \tag{18}$$

If the set $\mathcal{Y}$ is known apriori, the maximum likelihood approach to the estimation of Θ can lead to estimates obtained from appropriate roots of

$$\nabla_{\theta_{y_i=j}} \log P(\mathcal{X}, \mathcal{Y}|\Theta) = \sum_{i=1}^{n} \frac{\pi_j \cdot \frac{\partial f(x_i;\theta_j)}{\partial \theta_j}}{\sum_{k=1}^{J} \pi_k f(x_i; \theta_k)} \tag{19}$$

$$= \sum_{i=1}^{n} \frac{\pi_j f(x_i; \theta_j) \cdot \frac{\frac{\partial f(x_i;\theta_j)}{\partial \theta_j}}{f(x_i;\theta_j)}}{\sum_{k=1}^{J} \pi_k f(x_i; \theta_k)} \tag{20}$$

$$= \sum_{i=1}^{n} \frac{\pi_j f(x_i; \theta_j) \cdot \frac{\partial \log f(x_i;\theta_j)}{\partial \theta_j}}{\sum_{k=1}^{J} \pi_k f(x_i; \theta_k)} \tag{21}$$

$$\tag{22}$$

However since the labels y_i for each coordinate pixel x_i is unknown, the solutions of (22) can be found using the expectation-maximization (EM) algorithm (McLachlan & Peel, 2000). 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 $\mathcal{X} = x_i$:

$$p_{ij}^{(t)} = \frac{\pi_j^{(t)} f(x_i; \theta_j^{(t)})}{\sum_{k=1}^J \pi_k^{(t)} f(x_i; \theta_k^{(t)})} = p(\mathcal{Y} = j | \mathcal{X} = x_i; \theta) \qquad (23)$$

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

The M step requires finding the value of Θ 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

$$\sum_{i=1}^n p_{ij}^{(t)} \partial \log f(x_i; \theta_j) / \partial \theta_j = 0. \qquad (24)$$

The solution for (24) does not exist in closed form, and so has to be obtained iteratively. Deriving the corresponding maximum likelihood estimates of (24) poses some challenges due to the curved exponential family structure of the model. A more convenient approach is to use the method of moments for each component distribution as proposed in (Kent, 1982). Moment estimates can be shown to be consistent estimates of the true parameters and hence provide suitable starting values for maximum likelihood iteration. If the data are highly concentrated, the concentration parameters $\hat{\kappa}$ and $\hat{\beta}$ can also be calculated explicitly.

4.0.1 Weighted Moment Estimates

Moment estimates were first proposed in (Kent, 1982) to estimate the parameters of a single Kent distribution from a sample $((\gamma_{11}, \gamma_{21}), \cdots, (\gamma_{1n}, \gamma_{2n}))^T$. We adopt the approach for a mixture of Kent distributions. We let $((x_{11}, x_{21}, x_{31})^T, \cdots, (x_{1n}, x_{2n}, x_{3n})^T)$ denote the respective embedded cosine pixel coordinates of equation (14). The the weighted moment estimates for each cluster $\mathcal{Y} = j$ during iteration t are calculated as follows:

Step 1: Calculate the sample mean direction

$$\bar{\gamma}_{1,j} = \sum_{i=1}^n p_{ij}^{(t)} \gamma_{1i} / n_j, \ \bar{\gamma}_{2,j} = \sum_{i=1}^n p_{ij}^{(t)} \gamma_{2i} / n_j,$$

and

$$R_j^2 = S_{x_1,j}^2 + S_{x_2,j}^2 + S_{x_3,j}^2,$$

where $S_{x_1,j} = \sum_{i=1}^n p_{ij}^{(t)} x_{1i}$, $S_{x_2,j} = \sum_{i=1}^n p_{ij}^{(t)} x_{2i}$, $S_{x_3,j} = \sum_{i=1}^n p_{ij}^{(t)} x_{3i}$ and $n_j = \sum_{i=1}^n p_{ij}^{(t)}$. The mean resultant length is

$$\bar{R}_j = R_j / n_j$$

8

and the matrix $\mathbb{S}_j$ is given by

$$\mathbb{S}_j =$$

$$
\begin{pmatrix}
\sum_{i=1}^n p_{ij}^{(t)} x_{1i}^2 & \sum_{i=1}^n p_{ij}^{(t)} x_{1i} x_{2i} & \sum_{i=1}^n p_{ij}^{(t)} x_{1i} x_{3i} \\
\sum_{i=1}^n p_{ij}^{(t)} x_{1i} x_{2i} & \sum_{i=1}^n p_{ij}^{(t)} x_{2i}^2 & \sum_{i=1}^n p_{ij}^{(t)} x_{2i} x_{3i} \\
\sum_{i=1}^n p_{ij}^{(t)} x_{1i} x_{3i} & \sum_{i=1}^n p_{ij}^{(t)} x_{2i} x_{3i} & \sum_{i=1}^n p_{ij}^{(t)} x_{2i}^2
\end{pmatrix}
$$

Step 2: Compute the matrix

$$\mathbb{H}_j =$$

$$
\begin{pmatrix}
\cos \bar{\gamma}_{2j} & -\sin \bar{\gamma}_{2j} & 0 \\
\sin \bar{\gamma}_{2j} \cos \bar{\gamma}_{1j} & \cos \bar{\gamma}_{2j} \cos \bar{\gamma}_{1j} & -\sin \bar{\gamma}_{1j} \\
\sin \bar{\gamma}_{2j} \sin \bar{\gamma}_{1j} & \cos \bar{\gamma}_{2j} \sin \bar{\gamma}_{1j} & \cos \bar{\gamma}_{1j}
\end{pmatrix}
$$

and then compute matrix $\mathbb{B}_j$ by

$$\mathbb{B}_j = \mathbb{H}_j^T \mathbb{S}_j \mathbb{H}_j.$$

Define $\hat{\alpha}$ to be

$$\hat{\alpha}_j = \frac{1}{2} \arctan\{\frac{2b_{23}}{(b_{22} - b_{33})}\}.$$

Step 3: Compute the matrix

$$\mathbb{K}_j =$$

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \hat{\alpha} & -\sin \hat{\alpha} \\
0 & \sin \hat{\alpha} & \cos \hat{\alpha}
\end{pmatrix}
$$

Let

$$
\begin{align}
\hat{\mathbb{G}}_j &= \mathbb{H}_j \mathbb{K}_j \tag{25} \\
&= (\xi_1, \xi_2, \xi_3), \tag{26}
\end{align}
$$

where ξ_1, ξ_2 and ξ_3 are 3×1 column vectors. Calculate

$$\mathbb{V}_j = \hat{\mathbb{G}}_j^T \mathbb{S}_j \hat{\mathbb{G}}_j$$

and

$$W_j = v_{22} - v_{33},$$

where v_{ij} denotes the element of matrix $\mathbb{V}_j$ in the ith row and jth column.

Step 4: When κ is large, the limiting bivariate normal approximation gives the high-concetration approximations as

$$\hat{\kappa}_j \approx \frac{1}{2(1 - \bar{R}_j) - W_j} + \frac{1}{2(1 - \bar{R}_j) + W_j},$$

$$\hat{\beta}_j \approx \frac{1}{2}\left\{\frac{1}{2(1 - \bar{R}_j) - W_j} - \frac{1}{2(1 - \bar{R}_j) + W_j}\right\}.$$

9

Step 5: Let the mean direction be given by $(x_1, x_2, x_3)^T = \xi_1$. For each EM iteration, we weigh the corresponding data dependent moment estimates by equation (23) to obtain the M step. The moment estimates ensures 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. The spherical Kmeans algorithm(Dhillon & Mudha, 2001) provides the seeding required to initialize the proposed Kent mixture model.

5   Experiments

We consider a random unit vector $\mathcal{X}$, whose elements are positional coordinates of the intensity values of a pixel sample (with its neighborhood) 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 value of the vector. We make the assumption that sample directional unit pixel positional vectors were generated by selecting the class $y_i = j$, with prior probability π_j and then selecting $\mathcal{X}$, according to $f(\mathcal{X}|\theta_j)$ so that the mixture model derived above can now be applied.

5.1   Data

5.1.1   AVIRIS Hyperspectral Image 1992:

As to establish the effectiveness of the proposed hyperspectral feature transformation onto spherical manifold, and the application of the proposed mixture model, we generated results with the AVIRIS multispectral image. The West Lafayette image was used in the experiments. It is a multispectral image obtained with the Airborne/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 $0.663 - 0.673$ μm. 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. 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 1, we show the actual land cover usage from the AVIRIS image together with the land cover cosine coordinates for each pixel embedded onto a spherical manifold.

5.1.2   AVIRIS Hyperspectral Image 1986: Tippecanoe County

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 (Landgrebe & Biehl, 1992). The subset

Figure 1: AVIRIS 1992 West Lafayette land cover usage, color coded on ground truth. The corresponding cosine pixel coordinates on a spherical manifold are shown to the right.

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 2, we show the actual land cover usage from the AVIRIS image together with the land cover cosine coordinates for each pixel embedded onto a spherical manifold. On the final, clustering we exclude the 7th cluster that is due to sensor distortion.

5.2 Results

To evaluate the performance of the mixture model on hyperspectral data clustering, we used a metric-accuracy proposed in (Xu *et al.*, 2003). 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

$$accuracy = \frac{\sum_{i=1}^{N} \delta(g_i, map(t_i))}{N} \tag{27}$$

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

Figure 2: AVIRIS 1986 Tippecanoe County land cover usage color coded on ground truth. The corresponding cosine pixel coordinates on a spherical manifold are shown to the right.

algorithm (Lovasz & Plummer, n.d.). The function maps the predicted cluster labels to the corresponding best permuted representational cluster.

The clustering accuracy of the proposed Kent mixture model is compared to the results obtained by the spherical K-means and the von-Mises Fisher mixture model (**?**) . It can be seen from Table 1 that the accuracies achieved by the proposed algorithm are higher. For all three clustering methods, higher accuracy was observed for fewer cluster components. We also compare our results with those obtained in (Shah *et al.*, 2002), from which the authors used an independent component mixture model to study the same dataset but for only four clusters. We applied our proposed method to a small subset image with four clusters and observed the clustering accuracy to be 78%. This value is 18% above the value which was reported in (Shah *et al.*, 2002) for the same ground truth. This indicates that our proposed method has additional capability to carry out better clustering as compared to an independent component analysis(ICA) mixture model. In order to give a further quantitative performance evaluation of the proposed algorithm, we collected 2000 pixels from the Tippecanoe County image 2. We applied the SLE followed by the proposed Kent mixture model and generated the confusion matrix based on the relationship between the mappings obtained from the Kuhn-Munkres algorithm (Lovasz & Plummer, n.d.) and the ground-truth labels in Figure 2. 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 cluster components. 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 coordinate space from which to model hyperspectral images.

12

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

number of clusters	sphericalKmeans	vonMisesFisher-mixture	Kent-Mixture
2	55.10	63.34	87.48
3	69.86	82.93	79.98
4	55.80	61.31	70.00
5	46.41	51.39	59.33
6	50.49	49.14	57.30
7	51.17	51.11	54.29
8	49.63	48.38	57.94
9	52.07	53.31	61.55
10	50.70	50.71	58.50
11	47.64	50.93	59.11
12	43.90	47.91	57.94
13	43.23	45.92	55.69
14	43.45	46.77	56.67
15	44.66	41.81	51.54
16	42.80	41.46	49.55
17	42.03	40.46	47.19
Avg accuracy	46.41	48.64	56.71

Figure 3: Clustering accuracy on AVIRIS-West Lafayette Image

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

number of clusters	sphericalKmeans	vonMisesFisher-mixture	Kent-Mixture
2	79.58	77.47	75.89
3	56.41	73.77	76.22
4	55.96	52.89	68.66
5	54.11	44.77	65.65
6	52.74	39.33	65.23
Avg accuracy	49.80	48.03	58.61

Figure 4: Clustering accuracy on Tippecanoe County 1986 Hyperspectral Image

14

6 Conclusions

Our primary goal was to seek a constant curvature manifold on which hyperspectral images could be represented by their cosine coordinates and then develop a clustering technique 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 local embedding (SLE) method. The approach models cluster distributions with the Kent densities, resulting in a Kent Mixture Model (KMM). The results presented indicate the benefits of seeking spherical coordinates for analysis of hyperspectral images. The embedding method used introduces a neighborhood preserving constant curvature manifold enabling a potential for higher accuracy in clustering of land cover data.

References

Bachmann, C. M., Ainsworth, T. L., & Fusina, R. A. 2005. Exploiting manifold geometry in hyperspectral imagery. vol. 43.

Bao, Q., & Guo, P. 2004. Comparative studies on similarity measures for remote sensing image retrieval. *Ieee international conference on systems, man and cybernetics*.

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.

Dhillon, I. S., & Mudha, D. S. 2001. Concept decompositions for large sparse text data using clustering. *Machine learning*, 143–175.

Eckart, C., & Young, G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.

Jolliffe, I. T. 1986. *Principal component analysis*. Springer-Verlag.

Kent, J.T. 1982. The fisherbingham distribution on the sphere. *Journal of the royal statistical society*, **44**.

Landgrebe, D.A., & Biehl, L. 1992. 220 band hyperspectral image: Aviris image indian pine test site 3. http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

Lovasz, L., & Plummer, M. D.

Mardia, K.V., & Jupp, P. 2000. *Directional statistics*. John Wiley.

McLachlan, G., & Peel, D. 2000. *Finite mixture models*. Newyork: John Wiley and Sons.

Plaza, A., Martnez, P., Perez, R. M., & Plaza, J. 2002. A comparative analysis of endmember extraction algorithms using aviris hyeperspectral imagery.

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.

Shah, C., Arora, M. K., Robila, S. A., & Varshney, P. K. 2002. Ica mixture model based unsupervised classification of hyperspectral imagery. *In: Applied imagery pattern recognition workshop*.

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*.

Xu, W., Liu, X., & Gong, Y. 2003. Document clustering based on non-negative matrix factorisation. *Pages 267–273 of: Sigir'03: Proceedings of the 26th annual international acm sigir conference on research and development in information retrieval*.