

HYPERSPECTRAL UNMIXING WITH ENDMEMBER VARIABILITY USING SEMI-SUPERVISED PARTIAL MEMBERSHIP LATENT DIRICHLET ALLOCATION

Sheng Zou[†], Hao Sun^{*}, Alina Zare[†]

Department of Electrical and Computer Engineering, University of Florida[†]
Department of Electrical and Computer Engineering, University of Missouri^{*}

ABSTRACT

A semi-supervised Partial Membership Latent Dirichlet Allocation approach is developed for hyperspectral unmixing and endmember estimation while accounting for spectral variability and spatial information. Partial Membership Latent Dirichlet Allocation is an effective approach for spectral unmixing while representing spectral variability and leveraging spatial information. In this work, we extend Partial Membership Latent Dirichlet Allocation to incorporate any available (imprecise) label information to help guide unmixing. Experimental results on two hyperspectral datasets show that the proposed semi-supervised PM-LDA can yield improved hyperspectral unmixing and endmember estimation results.

Index Terms— semi-supervised, partial membership, latent dirichlet allocation, PM-LDA, hyperspectral, unmixing, superpixel, endmember, spectral variability

1. INTRODUCTION

Hyperspectral unmixing aims to decompose a hyperspectral image cube into constituent endmembers and estimate the proportion of each endmember in each pixel [1]. A widely used model in hyperspectral unmixing model is the Linear Mixture Model (LMM) that assumes the spectral signature of each pixel is a convex combination of endmember signatures [2]. In general, the overwhelming majority of unmixing methods represent endmembers as a single spectral signature. However, in practice, endmembers across a scene exhibit spectral variability [3] due to the effects of varying illumination [4], environmental [5, 6] and atmospheric conditions and the intrinsic variability of materials [7]. LMM-based algorithms that represent endmembers as a single spectral signature (and do not account for spectral variability) often have reduced performance in proportion estimation [3]. Thus, a number of methods that perform spectral unmixing while accounting for spectral variability have been developed in the literature [3, 8].

Spectral unmixing methods that account for endmember variability can be roughly partitioned into two major categories: (1) *endmembers as sets* approaches and (2) *endmembers as statistical distributions* approaches. Under “endmembers as sets,” the most prominent approach is the multiple endmember spectral mixture analysis (MESMA) [9, 10, 11], in which each pixel is represented by a collection of endmembers selected from a partitioned spectral library. Each partition corresponds to a set of spectral signatures that represent each endmember and its variations. For “endmembers as statistical distributions,” each endmember is modeled as a random variable according to some statistical distribution. A commonly chosen endmember distribution is the normal distribution which, when combined with the LMM, results in the Normal Compositional Model (NCM) [12] for spectral unmixing. The NCM assumes each pixel signature is a convex combination of K endmember variants drawn from K Normal endmember distributions. NCM-based methods include Echess *et al.*, [13, 14] and Kazianka [15] approaches which propose a Markov Chain Monte Carlo (MCMC) sampler to estimate proportion values and endmember covariances under NCM given endmember mean signatures or a spectral library. Zare *et al.*, [16, 17] presented MCMC sampler approach to estimate endmember spectral means and proportion values given endmember covariances under an NCM model. Zhang *et al.*, [18] introduced a particle swarm optimization expectation maximization (PSO-EM) method to estimate the endmember spectral means, endmembers covariances and proportion values. Sheng *et al.*, [19] introduced the Partial Membership Latent Dirichlet Allocation (PM-LDA) unmixing approach to estimate all endmember distributions and proportion values under the NCM while leveraging spatial information.

In this paper, we propose the semi-supervised PM-LDA (sPM-LDA) “endmember as statistical distributions” approach for spectral unmixing. sPM-LDA is an extension of PM-LDA [19] that provides the ability to leverage any available (even, imprecise) label information to guide unmixing. sPM-LDA also leverages spatial information by encouraging spatial homogeneity in local regions of the resulting proportion maps.

[†]The authors wish to thank the National Geospatial-Intelligence Agency for support of this research under the project entitled “NIP: Functions of Multiple Instances for Hyperspectral Analysis.”

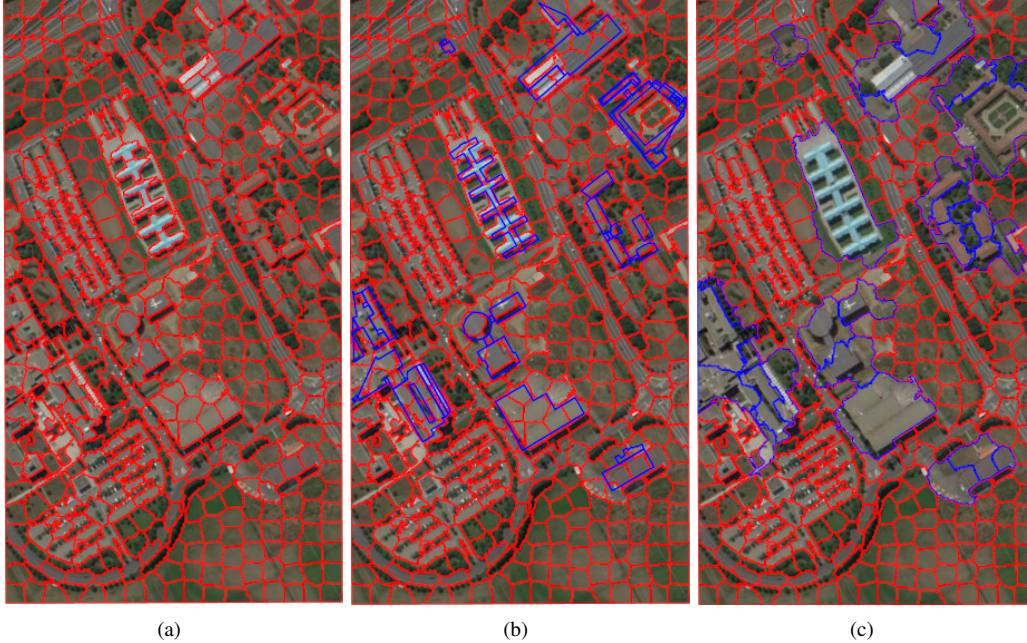


Fig. 1. Superpixel segmentation of the Pavia University hyperspectral datacube: (a) superpixel segmentation obtained using SLIC; (b) red: superpixels from SLIC, blue: open street map information; (c) Superpixel segmentation after merging superpixels that intersect the common building or road outlines found in the OpenStreetMap.org map information.

2. MOTIVATION

In the literature, hyperspectral unmixing methods that are labeled as “supervised unmixing” [13, 14] are generally approaches that rely on prior knowledge of the endmember signatures (*e.g.*, endmembers obtained either from spectral library or an separate endmember extraction algorithm). In these methods, the hyperspectral unmixing performance is highly dependent on the quality of the spectral library or endmember extraction method used. In this paper, as opposed to incorporating supervision through prior identification of endmember signatures, sPM-LDA provides an avenue for supervising the proportion values to be estimated. Since, however, labeling individual pixels with accurate sub-pixel proportion values is, in most cases, infeasible, sPM-LDA leverages *imprecise* proportion labeling. We propose a “lazy” supervision where neither the explicit endmember signatures nor pixel-level proportion values are needed. Instead, sPM-LDA allows for leveraging any available knowledge about which approximate spatial regions share may a common endmember and which do not. More specifically, we segment a hyperspectral image cube into small spatially-contiguous regions using a superpixel segmentation algorithm as illustrated in Fig. 1(a). Then, binary labels, τ , are assigned to a subset of these superpixels where τ is a $K \times C$ binary matrix (*i.e.*, consisting only of ‘1’ and ‘0’ values), K is the number of endmembers, and C is the number of superpixels. The value τ_{ij} is labeled as ‘1’ if

endmember i may be found in anywhere within superpixel j , otherwise, it is labeled as ‘0’. For example, if we want to provide supervision for identifying the ‘blue roof’ endmember in Fig. 1, we simply label any superpixels that may contain blue roof as ‘1’ in endmember label matrix and ‘0’ otherwise. In practice, these superpixel regions can be labeled using an easy manual labeling method (*i.e.* point-and-click on spatial regions that may contain blue roof) or through an automated approach by identifying the superpixels that intersect the outline of the building in a provided map (*e.g.*, from vector map sources such as OpenStreetMap.org [20]).

3. HYPERSEPECTRAL IMAGE SUPERPIXEL SEGMENTATION

Many state-of-the-art image segmentation algorithms have been developed for the gray-scale or RGB imagery. However, relatively little previous work exists for the hyperspectral image superpixel segmentation. The ultrametric contour map (UCM) algorithm is one of the popular superpixel segmentation algorithms for RGB imagery. Brightness, color and texture features are extracted and used to estimate the probability of the existence of a boundary between each pixel in the image. In [21], UCM was extended to hyperspectral imagery by first applying principal component analysis to reduce the image dimensionality to three dimensions then UCM can be directly applied. The normalized cuts algorithm has also been

extended for hyperspectral imagery [22]. This approach combines the spectral and spatial information and uses contour and texture features to partition the given graph recursively while minimizing the cost of the cut at the partition boundaries. In [23], superpixels have been computed with a graph based approach on a grid using the sum of squared differences between neighboring pixels.

In this paper, we consider a hyperspectral extension of Simple Linear Iterative Clustering (SLIC) [24]. The original SLIC algorithm has been widely used for segmenting RGB and gray-scale images. SLIC performs a local clustering of pixels in 5-D space with features as L, a, b value and the coordinates of the pixel. A Lab color space is a color-opponent space with dimensions L for lightness and a and b for the color-opponent dimensions, based on nonlinearly compressed (e.g. CIE XYZ) coordinates [25].

4. PARTIAL MEMBERSHIP LATENT DIRICHLET ALLOCATION

Partial Membership Latent Dirichlet Allocation (PM-LDA) [26, 27] is an extension of Latent Dirichlet Allocation topic modeling [28] that allows *words* to have partial membership across multiple topics. The use of partial memberships allows for topic modeling given data sets in which crisp topic assignments (as done by LDA) is insufficient since data points (or words) may straddle multiple topics simultaneously.

PM-LDA is a hierarchical Bayesian model in which data in a *corpus* is organized at two levels: the *word* and *document* level, as illustrated in Fig. 2. In the PM-LDA model, the

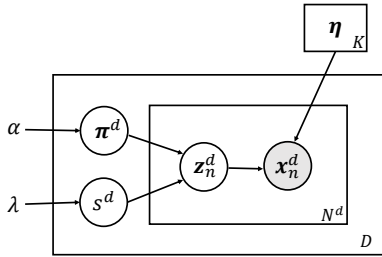


Fig. 2. Graphical model of PM-LDA

random variable associated with a data point, \mathbf{x} , is assumed to be distributed according to multiple topics with a continuous partial membership in each topic, \mathbf{z} . Specifically, the PM-LDA model is

$$p(\pi^d, s^d, \mathbf{z}_n^d, \mathbf{x}_n^d | \alpha, \lambda, \beta) = p(\pi^d | \alpha) p(s^d | \lambda) p(\mathbf{z}_n^d | \pi^d, s^d) \prod_{k=1}^K p_k(\mathbf{x}_n^d | \eta_k)^{z_{nk}^d} \quad (1)$$

where K is the number of topics in a corpus, $\pi^d \sim \text{Dir}(\alpha)$ is the topic proportion vector that provides the proportion of each topic found in document d , and $s^d \sim \exp(\lambda)$ controls the variance around the topic proportion mean. The variable \mathbf{x}_n^d is the n th word in document d and is distributed according to $\prod_{k=1}^K p_k(\mathbf{x}_n^d | \eta_k)^{z_{nk}^d}$ where $p_k(\cdot | \eta_k)$ is the distribution governing the k th topic, η_k are the associated topic distribution parameters, $\mathbf{z}_n^d \sim \text{Dir}(\pi^d s^d)$ is the partial membership vector of \mathbf{x}_n^d , and z_{nk}^d is the k th entry of \mathbf{z}_n^d . If $p_k(\cdot | \eta_k)$ is assumed to be an exponential family distribution, then $\mathbf{x}_n^d \sim \text{Expon}(\sum_k z_{nk}^d \eta_k)$ as shown in [29].

Given the data set (which has been partitioned into documents), $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^D\}$, and hyperparameters $\Psi = \{\alpha, \lambda\}$, the goal of parameter estimation given the PM-LDA model is to estimate the topic proportion of each document, π^d , the topic mixing level in each document, s^d , the partial memberships of each word in each topic, \mathbf{z}_n^d , and the natural parameters defining the probability distribution of each topic, η_k .

PM-LDA has been used previously for NCM-based hyperspectral unmixing and endmember estimation [19]. PM-LDA is applied to unmixing by, first, over-segmenting the hyperspectral scene into spatially-contiguous superpixels, as shown in Fig. 1(a). Each superpixel is assumed to be a *document* in the PM-LDA model. Given the superpixel segmentation and the assumption that the topic (*i.e.*, endmember) distributions are Gaussian (to assume the NCM), then the parameters of the PM-LDA model can be directly related to parameters of interest in the NCM unmixing model. Namely, the K topic distributions governed by parameters $\beta_k = \{\mu_k, \Sigma_k\}$ correspond to the K Gaussian endmember distributions. The cluster parameter $\beta_k = \{\mu_k, \Sigma_k\}$ can be mapped to the natural parameters of the exponential family distribution, $\eta_k = \{\Sigma_k^{-1}, \Sigma_k^{-1} \mu_k\}$. The partial membership vector for data point n in document d , \mathbf{z}_n^d , is the proportion vector associated with the n th data point in the d th superpixel. The topic proportion vectors for a document, π^d , correspond to the average proportion vector for a superpixel with the mixing level s^d corresponding to how much each proportion vector in the document is likely to vary from the average proportion vector. Thus, an entire hyperspectral scene or collection of hyperspectral scenes can be modeled as a corpus in PM-LDA.

Alg 1 summarizes a Metropolis-within-Gibbs sampler to perform parameter estimation for PM-LDA where $\Pi^{(t)}$ are the collection of all document-level topic proportions at time t , $\mathbf{S}^{(t)}$ are the topic mixing levels, $\mathbf{M}^{(t)}$ are the collection of partial membership vectors, and $\beta^{(t)}$ are the parameters defining each topic distribution. A Matlab implementation of PM-LDA parameter estimation can be found on our Github page [30].

An advantage of the use of a superpixel segmentation to define documents during unmixing is that it allows us to leverage the expected similarity of the materials found in neighboring pixels. In other words, spectrally homogeneous neigh-

Algorithm 1 Metropolis-within-Gibbs Sampling Method for Parameter Estimation

Input: A corpus \mathbf{D} , the number of topics K , hyperparameters $\Psi = \{\alpha, \lambda\}$, and the number of iterations T

Output: Collection of all samples: $\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \beta^{(t)}$

```

1: for  $t = 1 : T$  do
2:   for  $d = 1 : D$  do
3:     Sample  $\pi^d$ : Draw candidate:  $\pi^\dagger \sim \text{Dir}(\alpha)$ 
     Accept candidate with probability:
      $a_\pi = \min \left\{ 1, \frac{p(\pi^\dagger, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(\pi^{(t-1)} | \alpha)}{p(\pi^{(t-1)}, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(\pi^\dagger | \alpha)} \right\}$ 
4:     Sample  $s^d$ : Draw candidate:  $s^\dagger \sim \exp(\lambda)$ 
     Accept candidate with probability:
      $a_s = \min \left\{ 1, \frac{p(\pi^{(t)}, s^\dagger, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(s^{(t-1)} | \lambda)}{p(\pi^{(t)}, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(s^\dagger | \lambda)} \right\}$ 
5:     for  $n = 1 : N^d$  do
6:       Sample  $\mathbf{z}_n^d$ : Draw candidate:  $\mathbf{z}_n^\dagger \sim \text{Dir}(\mathbf{1}_K)$ 
       Accept candidate with probability:
        $a_z = \min \left\{ 1, \frac{p(\pi^{(t)}, s^{(t)}, \mathbf{z}_n^\dagger, \mathbf{x}_n | \Psi)}{p(\pi^{(t)}, s^{(t)}, \mathbf{z}_n^{(t-1)}, \mathbf{x}_n | \Psi)} \right\}$ 
7:     end for
8:   end for
9:   for  $k = 1 : K$  do
10:    Sample  $\mu_k$ : Draw proposal:  $\mu_k^\dagger \sim \mathcal{N}(\cdot | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})$ 
     $\mu_{\mathbf{D}}$  and  $\Sigma_{\mathbf{D}}$  are mean and covariance of the data
    Accept candidate with probability:
     $a_k = \min \left\{ 1, \frac{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \mu_k^\dagger) \mathcal{N}(\mu_k^{(t-1)} | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})}{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \mu_k^{(t-1)}) \mathcal{N}(\mu_k^\dagger | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})} \right\}$ 
11:   end for
12:   Sample covariance matrices  $\Sigma = \sigma^2 \mathbf{I}$ :
   Draw candidate from:  $\sigma^2 \sim \text{Unif}(0, u)$ 
   with  $u = \frac{1}{2} \{ \max_{\mathbf{x}_n} d^2(\mathbf{x}_n - \mu_{\mathbf{D}}) - \min_{\mathbf{x}_n} d^2(\mathbf{x}_n - \mu_{\mathbf{D}}) \}$ 
   Accept candidate with probability:
    $a_\Sigma = \min \left\{ 1, \frac{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \Sigma^\dagger)}{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \Sigma^{(t-1)})} \right\}$ 
13: end for

```

borhoods are likely to be grouped within a superpixel. Using PM-LDA, all of the pixels in a superpixel are paired with proportion vectors drawn from the same Dirichlet distribution with a shared average proportion vector (π^d) and the variance around that mean is governed by s^d . Larger s^d values correspond to more spatially and spectrally homogeneous superpixels.

5. PROPOSED METHOD

The proposed semi-supervised partial membership Latent Dirichlet Allocation (sPM-LDA) approach is comprised of three separate steps: 1) Map-Guided Hyperspectral SLIC; 2) Label assignment for each superpixel; and 3) Incorporation of endmember labels into PM-LDA for inference and parameter estimation.

5.1. Step 1: Map-Guided Hyperspectral SLIC

In step one of the proposed approach, an input hyperspectral image is over-segmented into superpixel regions using a map-guided segmentation approach. This superpixel segmentation is used to define spatially-contiguous regions with approximately similar proportion vectors. Furthermore, the map-guidance is used to merge superpixels that are identified to contain the same map objects as well as to autonomously generate map-derived labels for these superpixels.

To perform the superpixel segmentation, we extend the Simple Linear Iterative Clustering (SLIC) to hyperspectral imagery and to leverage map information. Previously in the literature, dimensionality reduction was applied to HSI prior to allow for application of SLIC [31, 32]. In this paper, no dimensionality reduction is applied. Instead, the full spectral information is used in combination with spatial information. The proposed hyperspectral SLIC is iterative: (1) initial cluster centers are obtained with a regular spatial sampling of the image where the cluster centers are vectors in which spectral information is concatenated with spatial coordinates; (2) then, each pixel is assigned to the nearest cluster center using a distance measure that includes a spectral and a spatial term; and (3) cluster centers are then updated to be equal to the average of all pixels assigned to the cluster. This process is iterated until the convergence.

To incorporate the map-guidance, post-processing based on map information (specifically, OpenStreetMap.org (OSM) [33]) is performed. OpenStreetMap is a collaborative project to create a free editable map of the world. From OSM data, we can parse out linear features and polygons that provide roadway locations and building profiles. Each of the OSM building profiles and roadway maps are paired with latitude and longitude information that can be used to align the map with the aerial hyperspectral data cube. However, even with world coordinates for the map information, mis-alignment with the hyperspectral data cube is likely, particularly, if the hyperspectral data cube does not have pixel-level accurate geo-location information. Thus, as opposed to attempting to label or link individual image pixels based on map information, superpixels are merged and labeled instead. The size of each superpixel provides tolerance for some level of mis-alignment between map data and imagery. The full map-guided hyperspectral SLIC is summarized in Algorithm 2.

5.2. Step 2: Superpixel Labeling

In step two, given the superpixel segmentation, supervision is imposed by assigning an endmember label vector τ_j to each superpixel j . The labels are used to identify the superpixels that *do not* contain a specific endmember. In other words, if $\tau_{ij} = 1$, then superpixel j may or may not contain endmember i . In contrast, if $\tau_{ij} = 0$, then superpixel j is constrained to not contain endmember i (*i.e.*, the proportion values for endmember i is constrained to be zero for all pixels

Algorithm 2 Map-Guided Hyperspectral SLIC Superpixel Segmentation

Input: HSI Data, K (number of superpixels), m (scaling factor)

- 1: Initialize cluster centers $\{c_k\}_{k=1}^K$ by sampling pixels at regular grid steps.
 - 2: Perturb $\{c_k\}_{k=1}^K$ in an $n \times n$ neighborhood to the lowest gradient position.
 - 3: **Repeat**
 - 4: **for** each c_k **do**
 - 5: Calculate the spectral distance of pixel x_i and c_k over all bands λ : $d_{spectral} = \sum_{\lambda=1}^B \|x_i(\lambda) - c_k(\lambda)\|_2^2$
 - 6: Calculate the spatial distance of pixel x_i and c_k :
 $d_{spatial} = \sqrt{(a_{x_i} - a_{c_k})^2 + (b_{x_i} - b_{c_k})^2}$ where a, b are pixel coordinates.
 - 7: Assign each pixel to a c_k in a $2S \times 2S$ square neighborhood based on the minimum spectral and spatial distance: $d_{x_i, c_k} = d_{spectral} + \frac{m}{S} d_{spatial}$
 - 8: **end for**
 - 9: Update cluster centers as the mean of all pixels assigned to the cluster.
 - 10: **Until** stopping criterion is reached.
 - 11: Align map data to imagery
 - 12: **for** Each Map Polygon **do**
 - 13: Merge all superpixels that overlap with this polygon
 - 14: **end for**
-

in the superpixel). This type of labeling is similar in concept to labels using with Multiple Instance Learning approaches [34, 35]. Given this labeling framework, an endmember material i whose locations are approximately known and can be supervised by labeling each superpixel j that may contain that endmember with $\tau_{ij} = 1$ and all other superpixel pixels with $\tau_{ij} = 0$, for example, consider the *red roof* material in the Pavia University dataset which can be found in several superpixels. It is relatively easy for a user to label the superpixels containing red roof.

On the contrary, there may be many endmembers whose locations throughout the scene are unknown. These endmembers are treated as unsupervised. For instance, there are a large number of superpixels in the Pavia University data set containing soil, asphalt, shadow, and vegetation materials. It can be time consuming and challenging to label these accurately. Thus, all superpixels are labeled with ‘1’ for those endmembers as they all have the possibility of containing those endmembers. Considering the extreme case where the end-member label matrix $\tau_{ij} = \mathbf{1}_{KC}$, then the proposed semi-supervised PM-LDA algorithm degrades to the standard PM-LDA unmixing algorithm.

5.3. Step 3: Using labels in PM-LDA for inference and parameter estimation

Under the unsupervised PM-LDA algorithm, topic proportions for each superpixel are drawn from a Dirichlet distribution with hyperparameters α . To impose supervision, we propose to draw topic proportions from a Dirichlet distribution whose mean is the point-wise product of original Dirichlet distribution mean and the endmember label vector for that superpixel. By applying the binary endmember labels to the topic proportion means, candidate topic proportion vectors are restricted to have zero entries on the endmembers that are not located in a particular superpixel. To be more specific, the topic proportion sampling step (line 3, Alg 1) is changed to $\pi^{\dagger} \sim \text{Dir}(\tau^d \otimes \alpha)$, where τ^d denotes the binary vectors of endmember labels for d th superpixel and \otimes indicates the point-wise product. Other than this change, sPM-LDA follows the PM-LDA algorithm (Alg. 1) exactly.

6. EXPERIMENTS

Two hyperspectral data sets, *University of Pavia* and *MUUFLL Gulfport*, were unmixed using the proposed method. We compared the proposed sPM-LDA with two other NCM-based algorithms (unsupervised) PM-LDA [19] and NCM-Bayes [13].

6.1. University of Pavia



Fig. 3. RGB-image of the University of Pavia hyperspectral data set

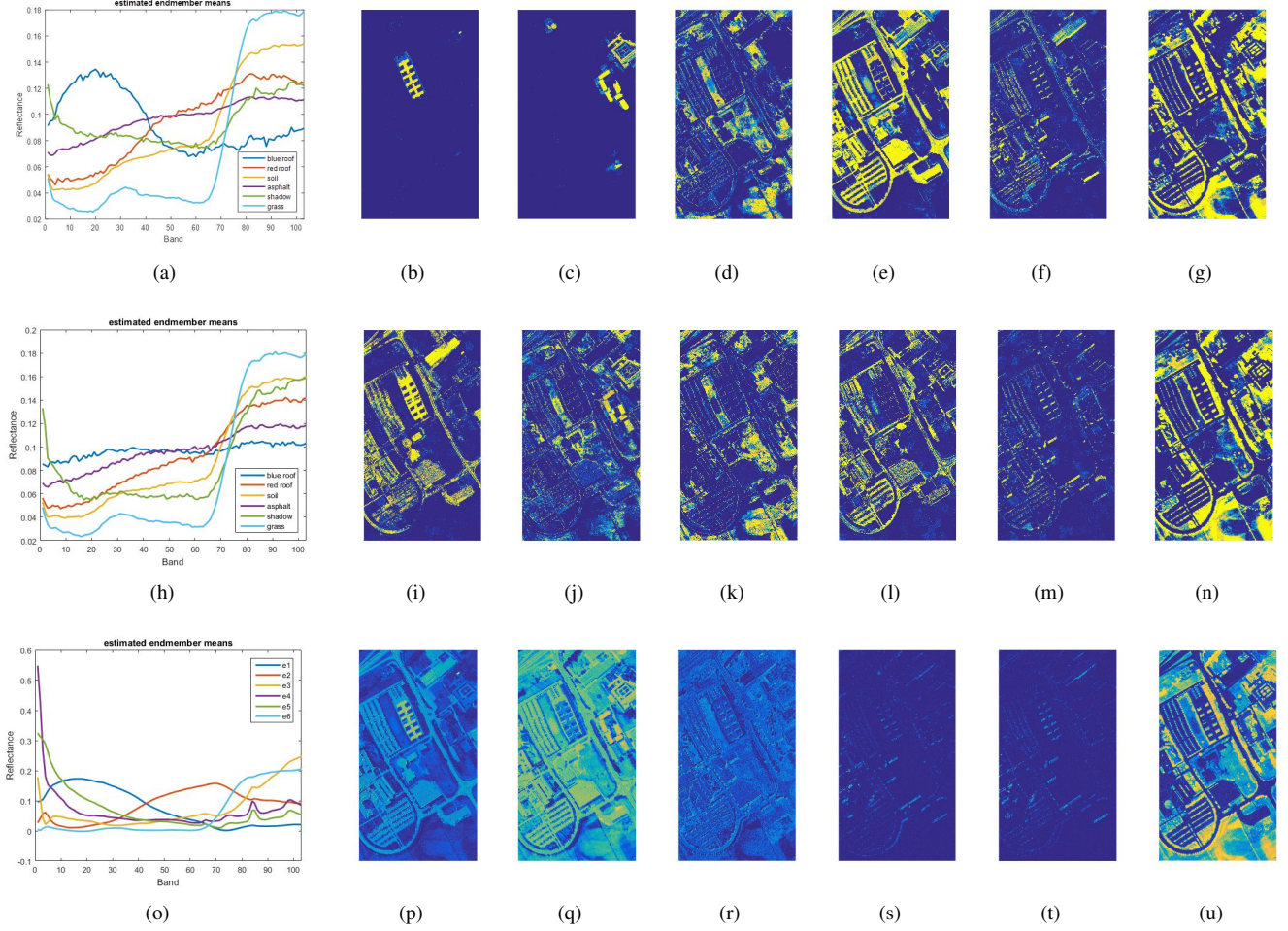


Fig. 4. Endmember means estimated by (a) supervised PM-LDA, (h) PM-LDA, (o) VCA. Estimated proportion maps using (b)-(g) supervised PM-LDA, (i)-(n) PM-LDA, (p)-(u) NCM-Bayes.

sPM-LDA was first applied to University of Pavia data set. This image was acquired in 2002 by Reflective Optics System Imaging Spectrometer (ROSIS) over Pavia, Italy. The scene contains 340×610 pixels, 103 bands with wavelengths ranging from 430 to 860 nm, and a spatial resolution of 1.3 m. An RGB image of the Pavia data set is shown in Fig 3.

The Pavia HSI image was segmented into superpixels using the map-guided hyperspectral SLIC algorithm with the following parameter settings: $k = 500$ and $m = 20$. Fig 1(a) shows the superpixels generated by the hyperspectral SLIC approach. Then, building polygons selected and pulled from OpenStreetMap.org from the region of interest are shown in Fig 1(b). The final superpixels are shown in Fig 1(c) by merging all superpixels overlapping any building polygons.

The number of endmembers was set to 6 for the proposed method and comparison methods. As a pre-processing step, all the pixel signatures in the data set are normalized to have unit length as the input for all methods. Endmembers are ini-

tialized in sPM-LDA and PM-LDA using Vertex Component Analysis (VCA) [36]. VCA was also used as the endmember extraction method applied in NCM-Bayes. Parameter settings for all methods were manually tuned to yield the best performance. For sPM-LDA, $K = 6$, $\lambda = 1$, $\alpha = 0.3$, $\epsilon = 5\%$ and $T = 200$. Blue roof and red roof materials were selected to be supervised in this study. To be more specific, the superpixels overlapping with the 3 blue roof or red roof polygons in Fig 1(b) were labeled as such. For instance, superpixels in blue roof regions are given labels of $\tau = [101111]$ while superpixels in red roof regions are given labels of $\tau = [011111]$, where ‘1’ in the first entry in τ denotes the possible existence of blue roof and ‘1’ in the second entry denotes the possible existence of red roof, while ‘1’ in the other four entries denotes the possible existence of the other four (unsupervised) endmembers. All other pixels in the scene are given $\tau = [001111]$. For NCM-Bayes, the Markov chain length was set to 250, the length of the burn-in period was set to 1000, $\delta = 0.001$, and

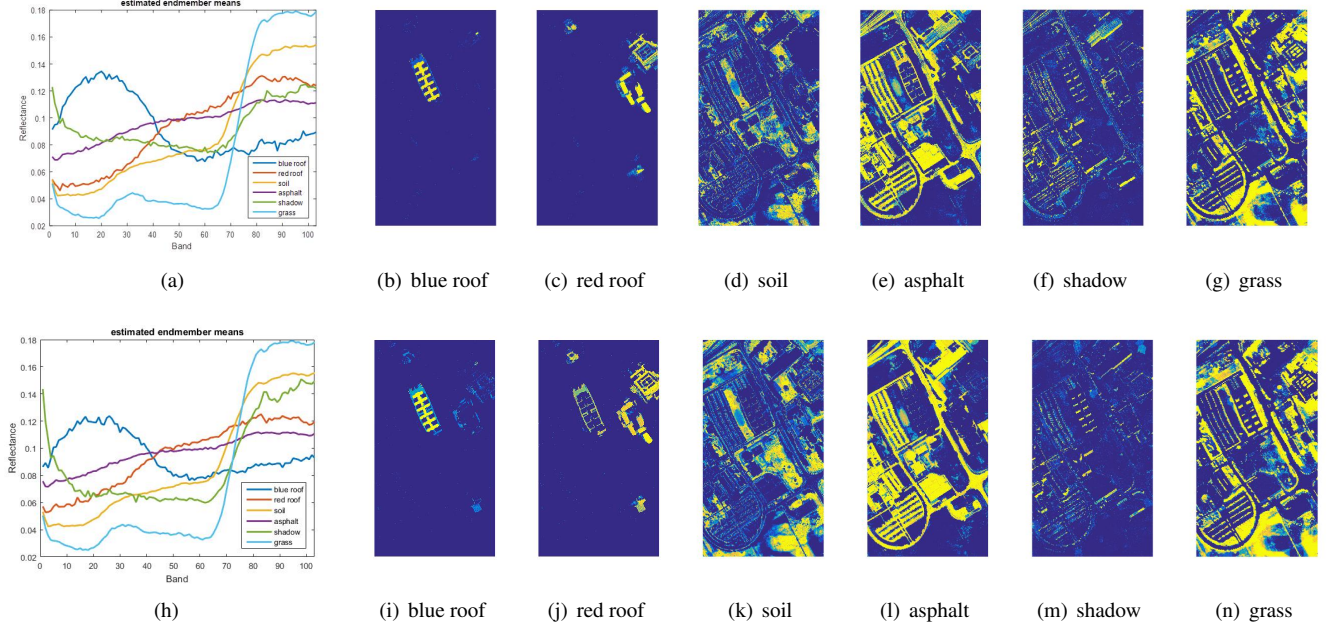


Fig. 5. Endmember means estimated on Pavia by (a) semi-supervised PM-LDA with precise labels, (h) semi-supervised PM-LDA with imprecise labels. Estimated proportion maps on Pavia using (b)-(g) semi-supervised PM-LDA with precise labels, (i)-(n) semi-supervised PM-LDA with imprecise labels.

the initial endmember variance was set to 0.001. For PM-LDA, $K = 6$, $\lambda = 1$, $\alpha = 0.3$ and $T = 200$.

For experimental results, all six endmember distribution parameters including endmember means and covariance matrices are estimated by sPM-LDA. The estimated endmember signatures and associated proportion maps, we can see that the signatures estimated by sPM-LDA are qualitatively more accurate on the Pavia data set. For example, the blue roof endmember clearly aligns only with blue roof pixels in the scene. The estimated proportion maps in Fig 4 illustrate the performance of the different methods. All proportion maps are permuted and aligned vertically according to the same major materials on each map so that each vertical map could be the same and compared. The six desired endmembers are blue roof, red roof, soil, asphalt, shadow and grass. Proportion maps estimated by supervised PM-LDA in the first row of Fig 4 are found to be smooth and the estimated proportion values are high for corresponding pixels dominated by single desired material and low for other materials. In contrast, (unsupervised) PM-LDA is unable to separate blue roof from some other buildings and roads consisting of different materials across the scene.

For quantitative evaluation, two evaluation metrics were computed: (1) entropy of proportion maps; and (2) data likelihood given estimation results. The proportion entropy is defined in (2).

$$H(\mathbf{P}) = - \sum_{n=1}^N \sum_{k=1}^K p_{nk} \ln p_{nk} \quad (2)$$

where p_{nk} is the proportion value for the n^{th} pixel and k^{th} endmember, N is the number of pixels and K is the number of endmembers. The motivation for applying proportion entropy is that in a hyperspectral image, there are usually a small number of endmembers that are present in each pixel. Therefore, accurate proportion values for each pixels yield low proportion entropy indicating that only a few endmembers are present in each pixel. The second evaluation metric is that NCM log-likelihood over all pixels in the dataset. The NCM-likelihood provides a measure of the overall fit between the hyperspectral data points and endmember distributions under the NCM model. The metric is indicated in (3).

$$f(\mathbf{X}|\mathbf{E}, \mathbf{P}, \mathbf{\Sigma}) = \sum_{n=1}^N \ln \mathcal{N} \left(\mathbf{x}_n \left| \sum_{k=1}^K p_{nk} \mathbf{e}_k, \sum_{k=1}^K p_{nk}^2 \mathbf{\Sigma}_k \right. \right) \quad (3)$$

The quantitative evaluation using two metrics are shown in Table 1 and Tabel 2, respectively.

6.1.1. Imprecise vs Precise Endmember Labels

In the sPM-LDA results above, relatively precise labels of $\tau = [101111]$ are given to indicate no red roof endmember

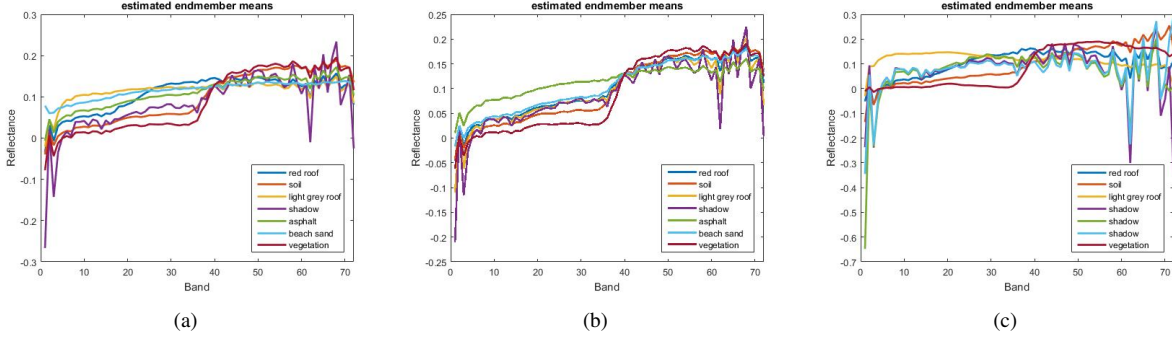


Fig. 6. Endmember means estimated by (a) supervised PM-LDA, (b) PM-LDA, (c) VCA.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	2.23e5	8.81e4	8.39e4

Table 1. Overall proportion map entropy for three methods on Pavia

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	8.97e6	6.68e7	6.53e7

Table 2. Overall log-likelihood for three methods on Pavia

exists for the superpixels in blue roof region by setting the second entry of τ to '0'. Similarly, relatively precise labels for superpixels in red roof regions were applied in the previous experiment. In this experiment, these relatively “precise” labels were compared to relatively imprecise labels. For imprecise labels, the superpixels in blue roof and red roof regions are all given the same label vector, $\tau = [111111]$, which relaxes the restriction for candidate endmembers in blue and red roof regions. All other pixels in non-blue-roof, non-red-roof regions are still labeled as $\tau = [001111]$. In other words, blue and red roof materials were not distinguished from each other in the label vectors. All other parameters are kept the same as in the original experiment.

Overall, the estimated endmember means and proportion maps by sPM-LDA with imprecise labels were found to have slightly more error when compared with those of sPM-LDA with precise labels. When examining the proportion maps in Fig 5, even with imprecise labels, the two different blue roof and red roof pixels are effectively separated. However, sidewalk pixels around the blue roof building are incorrectly partially represented by blue roof and red roof proportion values. This is the result of the fact that the sidewalk pixels are better represented by a linear combination of the blue and red roof endmembers than other endmembers (e.g., the asphalt signature). Therefore, as can be seen, sPM-LDA with imprecise labels can obtain good results but, of course, more precise labels can improve performance. As expected, the quantita-

tive results, both the entropy and likelihood values in Table 3, show that precise labels yield better unmixing performance than imprecise labels.

Labels	Precise	Imprecise
Entropy	8.39e4	8.81e4
Likelihood	6.53e7	6.46e7

Table 3. Overall proportion map entropy and log-likelihood for sPM-LDA with precise and imprecise labels on Pavia.

6.2. MUUFL Gulfport

sPM-LDA was also applied to the MUUFL Gulfport data set [37]. This image is acquired by the CASI-1500 hyperspectral imager flying over the University of Southern Mississippi-Gulfport in Long Beach, Mississippi in November 2010. The scene contains 325×337 pixels with 72 bands, of which wavelengths range from 375 to 1050 nm. The spatial resolution is 1 m. The RGB image of MUUFL Gulfport data set is shown in Fig 8 (a). This dataset can be found on our Github page [38].

The MUUFL Gulfport data were segmented into superpixels as is shown in Fig 4 (h). The parameters for superpixel segmentation was set to $k = 500$ and $m = 20$. The number of endmembers was set as seven for the proposed and comparison methods. The parameter settings for all methods are selected to yield the best performance. For sPM-LDA, endmembers for red roof, light grey roof and beach sand were selected to be the semi-supervised endmembers in this study. Parameters used in sPM-LDA were set to be: $K = 7$, $\lambda = 1$, $\alpha = 0.3$, $\epsilon = 10\%$ and $T = 200$. For NCM-Bayes, the Markov chain length was set to 250, the length of the burn-in period was set to 1000, $\delta = 0.001$, and the initial endmember variance was set to 0.001. For PM-LDA, $K = 7$, $\lambda = 1$, $\alpha = 0.3$ and $T = 200$.

Similar to experiments on Pavia, seven endmember distribution parameters including endmember means and covari-

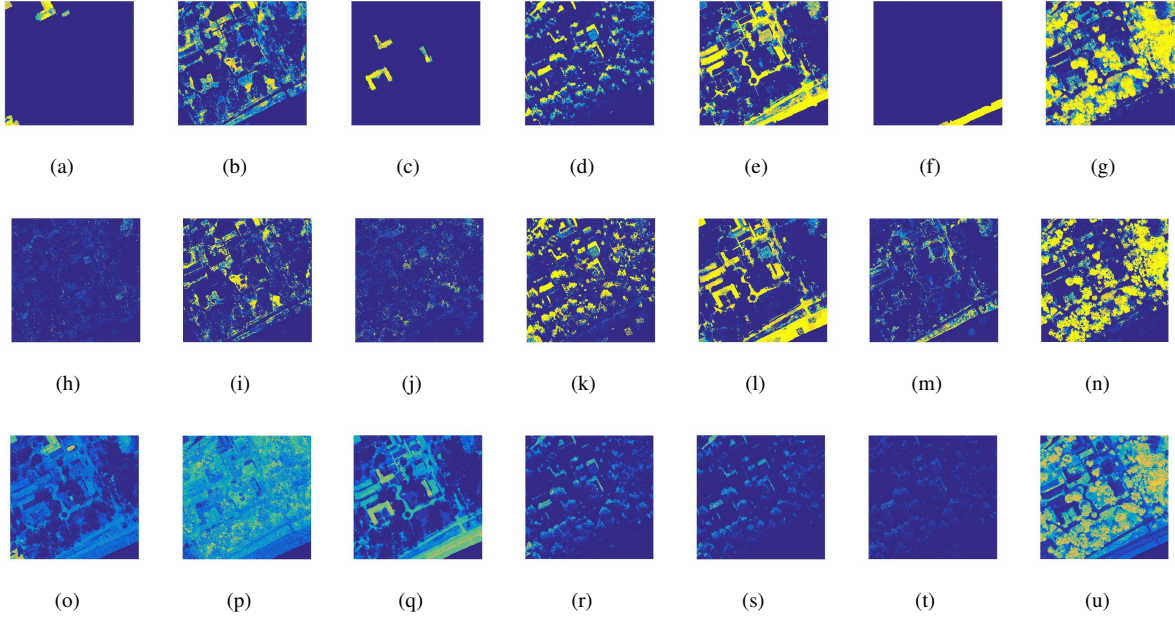


Fig. 7. Estimated proportion maps using (a)-(g) supervised PM-LDA, (h)-(n) PM-LDA, (o)-(u) NCM-Bayes.

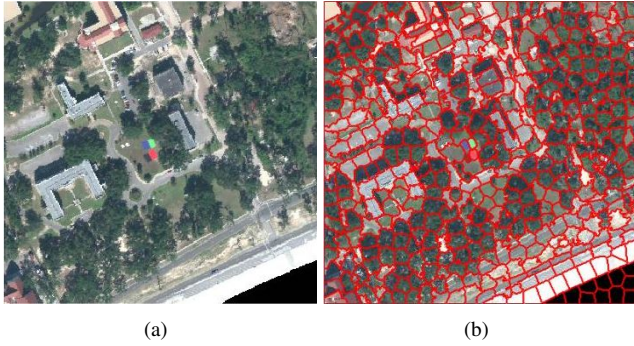


Fig. 8. MUUFL Gulfport dataset: (a) RGB image; (b) superpixels by hyperspectral SLIC.

ance matrices were estimated. The seven endmembers corresponded to red roof, soil, light grey roof, shadow, asphalt, beach sand and vegetation materials. The estimated endmember means are shown in Fig 6 and proportion maps are shown in Fig 7. All desired materials are found in proportion maps of sPM-LDA. For PM-LDA, the red roof and light grey roof are not accurately estimated. For NCM-Bayes, soil is mixed with most materials in the dataset as is shown in (p) and shadow is mapped to three endmembers from (r) to (t).

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	1.15e5	4.18e4	3.96e4

Table 4. Overall proportion map entropy for three methods on Gulfport.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	3.28e6	2.09e7	1.97e7

Table 5. Overall log-likelihood for three methods on Gulfport

7. SUMMARY

In this paper, a semi-supervised hyperspectral unmixing method modeled by PM-LDA is presented. This approach (sPM-LDA) is capable of incorporating imprecise labels in a semi-supervised hyperspectral unmixing process to obtain more sparse and interpretable endmember distribution and proportion estimations.

8. REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected*

Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 2, pp. 354–379, 2012.

- [3] A. Zare and K. C. Ho, “Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing,” *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.
- [4] A. Bateson and B. Curtiss, “A method for manual endmember selection and spectral unmixing,” *Remote Sensing of Environment*, vol. 55, no. 3, pp. 229–243, 1996.
- [5] G. P. Asner, C. A. Wessman, C. A. Bateson, and J. L. Privette, “Impact of tissue, canopy, and landscape factors on the hyperspectral reflectance variability of arid ecosystems,” *Remote Sensing of Environment*, vol. 74, no. 1, pp. 69–84, 2000.
- [6] G. S. Okin, D. A. Roberts, B. Murray, and W. J. Okin, “Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments,” *Remote Sensing of Environment*, vol. 77, no. 2, pp. 212–225, 2001.
- [7] J. Zhang, B. Rivard, A. Sánchez-Azofeifa, and K. Castro-Esau, “Intra-and inter-class spectral variability of tropical tree species at la selva, costa rica: Implications for species identification using hydice imagery,” *Remote Sensing of Environment*, vol. 105, no. 2, pp. 129–141, 2006.
- [8] B. Somers, G. P. Asner, L. Tits, and P. Coppin, “Endmember variability in spectral mixture analysis: A review,” *Remote Sensing of Environment*, vol. 115, no. 7, pp. 1603–1616, 2011.
- [9] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. O. Green, “Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models,” *Remote Sensing of Environment*, vol. 65, no. 3, pp. 267–279, 1998.
- [10] R. L. Powell, D. A. Roberts, P. E. Dennison, and L. L. Hess, “Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, brazil,” *Remote Sensing of Environment*, vol. 106, no. 2, pp. 253–267, 2007.
- [11] J. Franke, D. A. Roberts, K. Halligan, and G. Menz, “Hierarchical multiple endmember spectral mixture analysis (mesma) of hyperspectral imagery for urban environments,” *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1712–1723, 2009.
- [12] D. Stein, “Application of the normal compositional model to the analysis of hyperspectral imagery,” in *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*. IEEE, 2003, pp. 44–51.
- [13] O. Eches, N. Dobigeon, C. Mailhes, and J. Y. Tourneret, “Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, 2010.
- [14] O. Eches, N. Dobigeon, and J. Y. Tourneret, “Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical bayesian algorithm,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 582–591, 2010.
- [15] H. Kazianka, “Objective bayesian analysis for the normal compositional model,” *Computational Statistics & Data Analysis*, vol. 56, no. 6, pp. 1528–1544, 2012.
- [16] A. Zare and P. Gader, “PCE: Piecewise convex endmember detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2620–2632, 2010.
- [17] A. Zare, P. Gader, and G. Casella, “Sampling piecewise convex unmixing and endmember extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1655–1665, 2013.
- [18] B. Zhang, L. Zhuang, L. Gao, W. Luo, Q. Ran, and Q. Du, “PSO-EM: a hyperspectral unmixing algorithm based on normal compositional model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7782–7792, 2014.
- [19] S. Zou and A. Zare, “Hyperspectral unmixing with endmember variability using partial membership latent dirichlet allocation,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2017.
- [20] “OpenStreetMap,” <https://OpenStreetMap.org/>.
- [21] M. Sethi, Y. Yan, A. Rangarajan, R. R. Vatsavai, and S. Ranka, “Scalable machine learning approaches for neighborhood classification using very high resolution remote sensing imagery,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015, pp. 2069–2078.
- [22] D. B. Gillis and J. H. Bowles, “Hyperspectral image segmentation using spatial-spectral graphs,” in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83901Q–83901Q.
- [23] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castaño, “Superpixel endmember detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4023–4033, 2010.

- [24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [25] C. Connolly and T. Fleiss, "A study of efficiency and accuracy in the transformation from rgb to cielab color space," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 1046–1048, 1997.
- [26] C. Chen, *Partial Membership Latent Dirichlet Allocation*, Ph.D. thesis, University of Missouri-Columbia, 2016.
- [27] C. Chen, A. Zare, and J. T. Cobb, "Partial membership latent dirichlet allocation for image segmentation," in *IEEE International Conference on Pattern Recognition Workshops*. IEEE, 2016.
- [28] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [29] K. A. Heller, S. Williamson, and Z. Ghahramani, "Statistical models for partial membership," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 392–399.
- [30] "PM-LDA Matlab implementation," <https://github.com/GatorSense/PMLDA>.
- [31] X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "Slic superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2015, pp. 947209–947209.
- [32] H. Sun and A. Zare, "Map-guided hyperspectral image superpixel segmentation using proportion maps," in *IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2017.
- [33] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.
- [34] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," *Advances in neural information processing systems*, pp. 570–576, 1998.
- [35] C. Jiao and A. Zare, "Functions of multiple instances for learning target signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4670–4686, 2015.
- [36] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.
- [37] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL Gulfport hyperspectral and lidar airborne data set," Tech. Rep. REP-2013-570, University of Florida, Gainesville, FL, Oct 2013.
- [38] "MUUFL Gulfport data set," <https://github.com/GatorSense/MUUFLGulfport>.