HYPERSPECTRAL IMAGE SEGMENTATION AND UNMIXING USING HIDDEN MARKOV TREES

Roni Mittelman and Alfred O. Hero

University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA
{rmittelm,hero}@umich.edu

ABSTRACT

This paper is concerned with joint Bayesian endmember extraction and linear unmixing of hyperspectral images using a spatial prior on the abundance vectors. We hypothesize that hyperspectral images are composed of two types of regions. For the first type, the material proportions of adjacent pixels are similar and can be jointly characterized by a single vector, and in the second, neighboring pixels have very different abundances and are characterized by unique mixing proportions. Using this hypothesis we propose a new unmixing algorithm which simultaneously segments the image into such regions and performs unmixing. The experimental results show that the new algorithm can lead to improved MSE of both the extracted endmembers and the estimated abundances in low SNR cases.

Index Terms— Hyperspectral Imaging, Multiscale Segmentation, Sticky Hierarchical Dirichlet Process.

1. INTRODUCTION

The hyperspectral unmixing problem is concerned with the decomposition of the hyperspectral image into a product form, where the spectrum in each pixel is represented as a collection of material spectra that are referred to as *endmembers*, and the mixing proportions of these materials in each pixel that are known as the *abundances* [1]. One popular approach to this problem is to first estimate the endmembers using endmember extraction algorithms such as N-FINDER [2] or vertex component analysis (VCA) [3], and subsequently estimate the abundances using Bayesian or least squares methods. Another method proposed recently in [5] uses a fully Bayesian approach to estimate the endmembers and abundances simultaneously.

Most existing unmixing algorithms assume that the mixing proportions of each pixel are independent of its neighbors. This may not be justified when the spatial resolution of hyperspectral images is high and the abundances in adjacent pixels are correlated. One may consider cases where an image patch represents a large body of water or a vegetation field, where

adjacent pixels are expected to have similar abundances, thus there may be a benefit to exploit dependencies between such neighboring pixels. The advantage of addressing spatial dependencies in the context of hyperspectral imaging has also been demonstrated to be effective in [6]. In this work we assume that a hyperspectral image can be segmented into regions where the abundance vector can either be shared by all the pixels in the region, or each pixel can have an independent abundance vector. We assume that each of these segments is spatially smooth and propose a hidden Markov tree sticky hierarchical Dirichlet process (HMT-SHDP) to enforce this property while estimating the number of homogeneous abundance regions. The main advantage of using quadtree structured spatial priors, as opposed to other methods such as Markov random fields (MRF) [4], is that hyper-parameter estimation is much simplified compared to MRF [7]. Furthermore, the use of a SHDP enables the number of classes to be learned in an unsupervised fashion. A HMT that is based on hierarchical Dirichlet process was proposed in [9] where it was applied to image de-noising using wavelets. The HMT-SHDP that we present differs from [9] in that: (a) ours is sticky and therefore can capture the persistence-across-scale property; (b) our observations are only available at the leaves rather than in each node. Our experimental results verify that the new approach leads to improved estimation of the endmembers and abundances.

This paper is organized as follows. Section 2 presents background on the linear mixing model and the SHDP, whereas Section 3 presents the new unmixing algorithm. Section 4 presents the experimental results, and Section 5 concludes this paper.

2. BACKGROUND

2.1. The linear mixing model

Let P denote the number of pixels in the image, where each of the pixels is composed of $\mathcal{D}$ spectral bands $[y_{p,1}, \ldots, y_{p,\mathcal{D}}]^T$, then the linear mixing model (LMM) takes the form [5]

$$\mathbf{y}_p = \sum_{r=1}^{R} \mathbf{m}_r a_{p,r} + \mathbf{n}_p, \ p = 1, \ldots, P \qquad (1)$$

where $m_r = [m_{r,1}, \ldots, m_{r,\mathcal{D}}]^T$ denotes the rth endmember, $a_{p,r}$ is the fraction of endmember r in pixel p, R denotes the number of endmembers, and n_p is a zero mean Gaussian noise with covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$. We also denote $a_p = [a_{p,1}, \ldots, a_{p,R}]^T$.

2.2. The sticky HDP

The Dirichlet process (DP) [8] has been often used to perform model order selection in mixture models. Let H denote a distribution on a parameter space Θ. Denote the Dirichlet process by $\mathrm{DP}(\gamma, H)$ where $G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k}$, $\theta_k \sim H$, and where the weights β_k are obtained using a stick breaking construction with parameter γ,

$$\beta_k = \beta'_k \prod_{\ell=1}^{k-1}(1 - \beta'_\ell), \ \beta'_k \sim \mathrm{Beta}(1, \gamma). \qquad (2)$$

The HDP G_j is then obtained as a DP with the base distribution G_0, i.e. $G_j \sim \mathrm{DP}(\alpha, G_0)$. Equivalently we have that

$$G_j = \sum_{k=1}^{\infty} \pi_{jk}\delta_{\theta_k}, \ \boldsymbol{\pi}_j \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}). \qquad (3)$$

The HDP can be used as a prior for the transition probabilities in a hidden Markov model, where the transition probabilities from each state are sampled from the base measure G_0. However since the HDP satisfies the property $\mathbb{E}[G_j|G_0] = G_0$, it is very unlikely that the self transition probabilities for all the states will be sufficiently large. To ensure state persistence in applications such as segmentation where persistence is required, the SHDP was proposed in [8]. The SHDP modifies equation (3) such that $\boldsymbol{\pi}_j$ is sampled from $\boldsymbol{\pi}_j \sim \mathrm{DP}(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha+\kappa})$ instead of $\boldsymbol{\pi}_j \sim \mathrm{DP}(\alpha, \boldsymbol{\beta})$ where κ is a nonnegative constant. This has the effect of increasing the self transition probability for every state and hence the "sticky" HDP nomenclature.

3. THE BAYESIAN MODEL

The graphical model representation of the Bayesian unmixing algorithm is shown in Figure 1. The likelihood of the observation matrix $\mathbf{Y} = [y_1, \ldots, y_P]$ is explicitly obtained from the LMM. The HDP base stick breaking process $\boldsymbol{\beta}$, and the derived transition probabilities $\boldsymbol{\pi}_k^{(\ell)} \ \forall k = 1, \ldots, L-1$, are obtained similarly to those described in the Section 2.2. Next we describe the priors used for each of the remaining random variables, and the inference using the Gibbs sampler.

3.1. Endmember Spectral Prior

We propose a prior for the endmembers similar to [5] where dimensionality reduction and the positivity constraint on spectra are used to define a set over which the endmembers
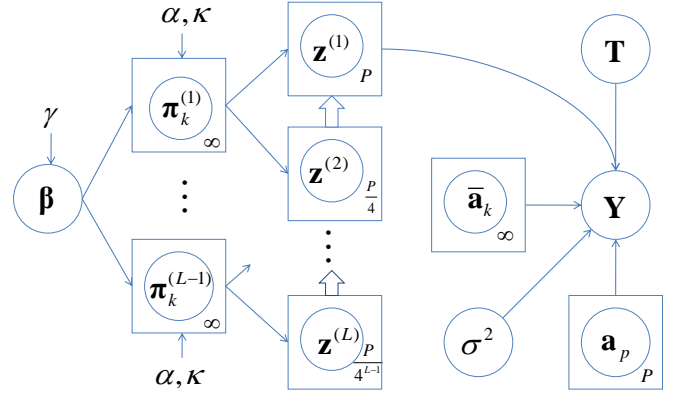

Fig. 1. Graphical model representation for the image segmentation algorithm. The wide arrow notation denotes connections between levels of the quadtree.

are constrained. The prior then takes the form of a multivariate Gaussian distribution truncated over this set. Let $\bar{y}$ denote the mean spectral bands averaged over all pixels and let $\boldsymbol{\Upsilon}$ denote the empirical covariance matrix of the spectral bands computed using all the hyperspectral data. Then by projecting the endmembers onto the subspace spanned by the principal components of $\boldsymbol{\Upsilon}$, the dimensionality of the admissible endmembers and abundances is reduced. Specifically, let D denote a diagonal matrix with the K largest eigenvalues of $\boldsymbol{\Upsilon}$ arranged along the diagonal, and similarly let V denote a matrix with the columns that are the eigenvectors corresponding to the K largest eigenvectors, then the endmember m_r is projected into the lower dimensional space using $t_r = \mathrm{P}(m_r - \bar{y})$, where $\mathrm{P} = \mathrm{D}^{-1/2}\mathrm{V}$. Equivalently we have that $m_r = \mathrm{P}^{\dagger}t_r + \bar{y}$, where $\dagger$ denotes the pseudo inverse. Using the positivity of the endmembers spectra it can be shown that the admissible set for t_r takes the form $\mathscr{T}_r = \{t_r | \bar{y}_d + \sum_{k=1}^{K} u_{d,k}t_{k,r} \geq 0, \ d = 1, \ldots \mathcal{D}\}$. The endmember prior then takes the form of a truncated Gaussian on the set $\mathscr{T}_r$, $\phi_{\mathscr{T}_r}(t_r|e_r, s_r^2\mathrm{I}) \propto \phi(t_r|e_r, s_r^2\mathrm{I})\mathbf{1}_{\mathscr{T}_r}(t_r)$, where $\mathbf{1}_{\mathscr{T}_r}(\mathrm{x})$ is a function that returns one if x is inside the set $\mathscr{T}_r$ and zero otherwise, and $\phi(t_r|e_r, s_r^2\mathrm{I})$ is a multivariate Gaussian with mean e_r and covariance matrix $s_r^2\mathrm{I}$. The mean vectors e_r are chosen using the VCA algorithm [3], and the variances s_r^2 are fixed to some large values, $r = 1, \ldots, R$. We also use the notation $\mathrm{T} = [t_1, \ldots, t_R]$.

3.2. Noise variance prior

The noise variance parameter σ^2 is an inverse-gamma distribution with parameters 1 and $\gamma_0/2$, where γ_0 is also distributed as $f(\gamma_0) \propto \frac{1}{\gamma_0}\mathbf{1}_{\mathbb{R}+}(\gamma_0)$.

3.3. Label Prior

Consider the quadtree lattice, where the nodes are discrete random variables that take their values from the set $\{1, \ldots \mathcal{K}\}$. We denote the labels by $z_p^{(\ell)}$, $\forall p = 1, \ldots, P/4^{\ell-1}$, $\forall \ell = 1, \ldots, \mathcal{L}$, where $\mathcal{L}$ denotes the number of levels in the quadtree. We also denote the vector containing all the labels at the ℓth level by $\mathrm{z}^{(\ell)}$. The labels $\{z_p^{(1)}\}_{p=1}^{P}$ at the bottom most level of the quadtree, are associated with the appropriate pixels of the hyperspectral image. We assume a Markovian relationship between the labels on the quadtree lattice. Specifically, let us define the likelihoods

$$\pi_{ji}^{(\ell)} = p(z_p^{(\ell)} = i | z_{Pa(p)}^{(\ell+1)} = j), \tag{4}$$

where $Pa(p)$ denotes the parent node of p, then the joint probability mass function of all the labels takes the form:

$$p(\{\mathrm{z}^{(\ell)}\}_{\ell=1}^{\mathcal{L}} | \{\boldsymbol{\pi}^{(\ell)}\}_{\ell=1}^{\mathcal{L}-1}) = p(\mathrm{z}^{(\mathcal{L})}) \prod_{\ell=1}^{\mathcal{L}-1} \prod_{p=1}^{\frac{P}{4^{\ell-1}}} \pi_{z_{Pa(p)}^{(\ell+1)} z_p^{(\ell)}}^{(\ell)}. \tag{5}$$

It can readily be observed that by forcing larger self transition probabilities, i.e. $\pi_{ii} > \pi_{ji} \, \forall i \neq j$, one can effectively encourage the formation of spatially smooth label regions at lower levels of the quadtree.

3.4. Abundance prior

Our algorithm employs two types of abundance parameters: (1) a_p, $p = 1, \ldots, P$, which are associated with each pixel independently; (2) $\bar{\mathrm{a}}_k$, $k = 2, \ldots, \mathcal{K}$, which are associated with a group of pixels that have very similar abundances. The abundance vector assigned to each pixel is determined using the labels $\mathrm{z}^{(1)}$. Let $z_p^{(1)} = k$, then if $k = 1$ the abundance vector of the pth indexed pixel is equal to a_p, otherwise it is equal to $\bar{\mathrm{a}}_k$. Similarly to [5], we use abundance priors that are uniform on the validity set which satisfies the positivity and sum to one constraints.

3.5. Gibbs sampler

Estimation of the random parameters is performed using a blocked Gibbs sampler that is outlined in Algorithm 1. The infinite Dirichlet process is approximated using a finite mixture with $\mathcal{K}$ mixtures.

4. EXPERIMENTAL RESULTS

We generated a 50×50 hyperspectral image with 200 spectral bands by generating the abundance vectors randomly from four truncated multivariate Gaussians with given means and covariance matrix $0.001 \times \mathrm{I}$. The simulated endmembers are shown in Figure 3(a)-(c). The segmented image for the 5db case is shown in Figure 3(d), where the darkest gray level
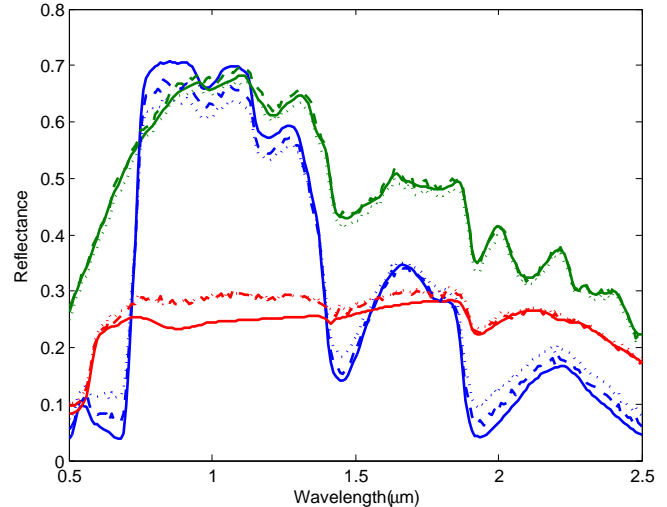

Fig. 2. Experimental results for SNR of 5db. The solid lines are the true endmembers, the dashed lines are the endmembers estimated using the HMT-Bayesian algorithm, and the dotted lines are the endmembers extracted using the Bayesian method in [5].

corresponds to regions where each pixel has a different abundance vector, and non-black gray levels correspond to a group of pixels that share the same mixing proportions. The MSE results in Table 1 show that the new approach can improve the estimation performance of the abundances as compared with [5] for low SNR. Similarly Figure 2 which corresponds to SNR of 5db, shows that the endmembers extracted using the joint segmentation and unmixing method are more accurate compared to the Bayesian algorithm presented in [5].

SNR	Bayesian	HMT-Bayesian
5db	50	23.5
10db	40	22.2
15db	18.7	19.5

Table 1. The MSE of the estimated abundances for different SNRs.

5. CONCLUSIONS

We presented a new hyperspectral unmixing algorithm which uses spatial constraints in order to improve endmember and abundance estimation performance. A HMT-SHDP was presented and used as a spatial prior which enables inference of the number of regions in an unsupervised fashion. The experimental results verify that the new method can improve the performance of abundance and endmember estimation in low SNR scenarios.
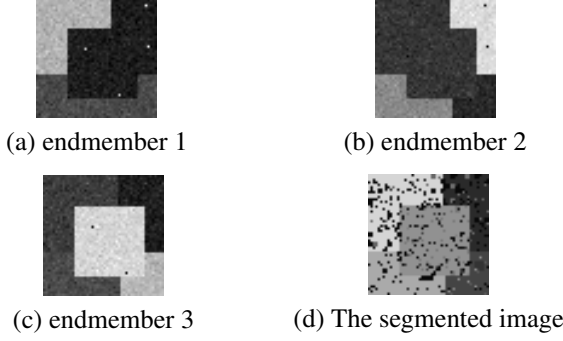
(a) endmember 1   (b) endmember 2

(c) endmember 3   (d) The segmented image

Fig. 3. The simulated abundances and the segmented image.

6. REFERENCES

[1] N. Keshava and J. Mustard, "Spectral unmixing," IEEE Signal Process. Mag., vol. 19, no. 1, pp. 44–57, Jan. 2002.

[2] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in Proc. 13th Int. Conf. on Applied Geologic Remote Sensing, vol. 2, Vancouver, pp. 337–344, Apr. 1999.

[3] J. M. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 4, pp. 898–910, Apr. 2005.

[4] R. Neher and A. Srivastava, "A Bayesian MRF framework for labeling terrain using hyperspectral imaging," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 7, pp. 1543–1551, Jul. 2004.

[5] N. Dobigeon, S. Moussaoui, M. Coulon, J. Y. Tourneret and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," IEEE Trans. Signal Process., vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[6] O. Eches, N. Dobigeon and J. Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," submitted to IEEE Trans. Geosci. Remote Sens.

[7] C. A. Bouman, "A multiscale random field model for Bayesian image segmentation," IEEE Trans. Image Process., vol. 3, no. 2, pp. 162–177, Mar. 1994.

[8] E. B. Fox, E. B. Sudderth, M. I. Jordan and A. S. Willsky, "The sticky HDP-HMM: Bayesian nonparametric hidden Markov models with persistent states," MIT LEADS technical report P-2777, Nov. 2009.

[9] J. J. Kivinen, E. B. Sudderth and M. I. Jordan, "Image denoising with nonparametric hidden Markov trees," IEEE International Conference on Image Processing (ICIP), 2007.

Algorithm 1

---

1. For $\ell = 1, \ldots, \mathcal{L}$ compute:

$$b_{p_\ell}^{(\ell)}(k) = \begin{cases} \mathcal{N}(\mathbf{y}_{p_\ell}; \mathbf{M}\tilde{\mathbf{a}}_p, \sigma^2\mathbf{I}), & \ell = 1 \\ \prod_{j \in c(p_\ell)} \sum_{i=1}^{\mathcal{K}} \pi_{ki}^{(\ell-1)} b_j^{(\ell-1)}(i), & \text{else} \end{cases}$$

   where $c(p_\ell)$ denotes the set composed of the children of node p_ℓ. If $z_p^{(1)} = 1$ then $\tilde{\mathbf{a}}_p = \mathbf{a}_p$, else $\tilde{\mathbf{a}}_p = \bar{\mathbf{a}}_{z_p^{(1)}}$.

2. For $\ell = \mathcal{L}, \ldots, 1$ sample the state assignments $z_{p_\ell}^{(\ell)}$:

$$z_p^{(\ell)} \propto \sum_{k=1}^{\mathcal{K}} \pi_{z_p^{(\ell+1)}k}^{(\ell+1)} b_{p_\ell}^{(\ell)}(k) 1_k(z_p^{(\ell)})$$

3. Compute $n_{jk}^{(\ell)} = \sum_{p_\ell} \sum_{i \in c(p_\ell)} 1(z_{p_\ell}^{(\ell)} = j, z_i^{(\ell-1)} = k)$

4. Sample the auxiliary variables m, w, m̄ as follows

$$p(m_{jk}^{(\ell)} = m | n_{jk}^{(\ell)}, \boldsymbol{\beta}, \alpha, \kappa) \propto s(n_{jk}^{(\ell)}, m)(\alpha\beta_k + \kappa 1(k = j))^m$$

$$p(w_{jt}^{(\ell)} | \boldsymbol{\beta}, \alpha, \kappa) \propto \begin{cases} \frac{\beta_j \alpha}{\alpha + \kappa}, & w_{jt}^{(\ell)} = 0 \\ \frac{\kappa}{\alpha + \kappa}, & w_{jt}^{(\ell)} = 1 \end{cases}, t = 1, \ldots, m_{j,k}^{(\ell)}$$

$$\bar{m}_{jk}^{(\ell)} = \begin{cases} m_{jk}^{(\ell)}, & j \neq k \\ m_{jj}^{(\ell)} - \sum_{t=1}^{m_{jk}^{(\ell)}} w_{jt}^{(\ell)}, & j = k \end{cases}$$

   where $s(n, m)$ are unsigned sterling numbers of the first kind.

5. Sample the new base transition probabilities

$$\boldsymbol{\beta} \sim Dir(\gamma/\mathcal{K} + \sum_{j,\ell} m_{j1}^{(\ell)}, \ldots, \gamma/\mathcal{K} + \sum_{j\ell} m_{j\mathcal{K}}^{(\ell)}).$$

6. Sample the new base DP

$$\boldsymbol{\pi}_k^{(\ell)} \sim Dir(\alpha\beta_1 + n_{k1}^{(\ell)}, \ldots, \alpha\beta_k + \kappa + n_{kk}^{(\ell)}, \ldots, \alpha\beta_\mathcal{K} + n_{k\mathcal{K}}^{(\ell)})$$

7. Sample the new group partial abundance vectors $\bar{\mathbf{c}}_k$, $k = 2, \ldots, \mathcal{K}$ from $\phi(\bar{\mathbf{c}}_k | \boldsymbol{\nu}_k, \boldsymbol{\Sigma}_k) \mathbf{1}_{\mathcal{S}_c}(\bar{\mathbf{c}}_k)$, where,

$$\boldsymbol{\Sigma}_k^{-1} = |Z_k|(\mathbf{M}_{-R} - \mathbf{m}_R \mathbf{1}_{R-1}^T)^T \boldsymbol{\Sigma}_n^{-1}(\mathbf{M}_{-R} - \mathbf{m}_R \mathbf{1}_{R-1}^T)$$

$$\boldsymbol{\Sigma}_k^{-1} \boldsymbol{\nu}_k = (\mathbf{M}_{-R} - \mathbf{m}_R \mathbf{1}_{R-1}^T)^T \boldsymbol{\Sigma}_n^{-1}\left(\frac{1}{|\mathcal{I}_k|} \sum_{p \in \mathcal{I}_k} \mathbf{y}_p - \mathbf{m}_R\right)$$

   where $\mathcal{I}_k = \{p | z_p^{(1)} = k\}$, and set $\bar{\mathbf{a}}_k = [\begin{array}{cc} \bar{\mathbf{c}}_k & 1 - \mathbf{1}_{R-1}^T \bar{\mathbf{c}}_k \end{array}]^T$.

8. Sample the endmembers, abundances $\mathbf{a}_p$, and noise variance σ^2, using the same formulas given in [5].