A SPATIAL COMPOSITIONAL MODEL FOR LINEAR UNMIXING

Yuan Zhou, Anand Rangarajan and Paul D. Gader

Dept. of Computer and Information Science and Engineering
University of Florida, Gainesville, FL, USA

ABSTRACT

Hyperspectral image unmixing into endmembers and abundances is important in many applications. Despite the plethora of linear mixing models, there is very little previous work on directly incorporating spatial information in the pixel likelihood while simultaneously modeling the uncertainty in the extracted endmembers. In this paper, we propose a spatial compositional model (SCM) for this purpose. In SCM, the model uncertainty of the endmembers can be estimated, while the spatial information in the abundances is directly incorporated in the likelihood via a random variable transformation. This results in a simple and efficient algorithm for both endmember extraction, abundance and uncertainty estimation. The results compared with current state-of-the-art algorithms on real datasets are promising.

Index Terms— endmember extraction, hyperspectral image analysis, linear unmixing, spatial compositional model

1. INTRODUCTION

Unmixing in hyperspectral imaging has received wide attention in the remote sensing, signal and image processing and geospatial literature [1, 2]. If unmixing can be successfully carried out, many applications can utilize the material information contained in the original image. Despite the plethora of mixing models and inference algorithms available in the literature, very little previous work has focused on linear unmixing using spatial likelihoods while simultaneously modeling endmember uncertainty [3, 4, 5, 6, 7]. Our work is motivated by this observation.

We model the hyperspectral image in the following manner. The linear mixing model assumes that the spectral measurement $g(\mathbf{x}, \lambda)$ at wavelength λ, location $\mathbf{x}$ in the image domain, is a non-negative linear combination of the spectral signature of *endmembers* (pure materials), $f_i(\lambda)$:

$$g(\mathbf{x}, \lambda) = \sum_{i=1}^{M} f_i(\lambda)\, \alpha_i(\mathbf{x}) + n(\mathbf{x}, \lambda), \sum_{i=1}^{M} \alpha_i(\mathbf{x}) = 1 \quad (1)$$

where M is the number of endmembers, $\alpha_i(\mathbf{x})$ is the fractional *abundance map* of the *i*th endmember satisfying simplex constraints and $n(\mathbf{x}, \lambda)$ is a noise field (usually modeled as Gaussian). Note that the abundances directly depend on location in (1).

As a result, the pixels generated by this model form a simplex whose vertices are the endmembers. If we discretize $f_i(\lambda)$ into B bands obtaining $\mathbf{m}_i \in \mathbb{R}^B$ as a discretized value, and further discretize the domain into N locations, (1) reduces to $\mathbf{y}_i = \mathbf{M}^T \boldsymbol{\alpha}_i + \mathbf{n}_i$ for $i = 1, ..., N$, where $\mathbf{M} = [\mathbf{m}_1, ..., \mathbf{m}_M]^T$, $\boldsymbol{\alpha}_i = [\alpha_{i1}, ..., \alpha_{iM}]^T$, $\mathbf{y}_i \in \mathbb{R}^B$, $\mathbf{n}_i \in \mathbb{R}^B$. Writing this out at all locations, we get

$$\mathbf{Y} = \mathbf{A}\mathbf{M} + \mathbf{N} \quad (2)$$

where $\mathbf{Y} \in \mathbb{R}^{N \times B}$, $\mathbf{A} \in \mathbb{R}^{N \times M}$, $\mathbf{M} \in \mathbb{R}^{M \times B}$, $\mathbf{N} \in \mathbb{R}^{N \times B}$. The linear unmixing problem is to find $\mathbf{A}$ and $\mathbf{M}$ given $\mathbf{Y}$.

Among the many linear unmixing models focusing on endmember uncertainty, the normal compositional model (NCM) is a widely used approach [3, 4, 8]. However, as we will show, it does not accurately capture the uncertainties in hyperspectral image formation. That is, given the pixel data, there can be many possible endmember sets (simplexes) that fit the data. In virtually all of the previous research, the probability density $p(\mathbf{Y}) = \prod_{i=1}^{N} p(\mathbf{y}_i)$ is obtained by assuming that the pixels are independent. This assumption robs the model of location dependence and *ignores pixel correlation*.

In this paper, we propose a *spatial compositional model* (SCM) which (i) incorporates spatial information on the abundances, (ii) directly models pixel correlation, and (iii) models endmember uncertainty via physically reasonable priors. The resulting *maximum a posteriori* (MAP) objective function is minimized using a simple and efficient alternating algorithm which simultaneously estimates the abundances, endmembers and associated uncertainty parameters.

2. THE SPATIAL COMPOSITIONAL MODEL

2.1. Modeling the priors

We can model the prior probability of the abundances $\mathbf{A}$ as a Markov random field. That is, by treating the image grid as an undirected graph, the density function of the whole grid can be modeled based on a potential function of the maximal cliques (neighboring nodes). Assuming the hyperspectral image comprises different regions, the prior probability of $\mathbf{A}$ can

This work is partially supported by NSF IIS 1065081

be assumed to be in favor of a smooth assignment of $\boldsymbol{\alpha}_i$ to all the pixels within a region, as such pixels are more likely to be the same mixture of materials.

We use the following prior probability for $\mathbf{A}$:

$$p(\mathbf{A}) \propto \exp\left\{-\frac{\beta}{4}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\|\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_j\|^2\right\} = \exp\left\{-\frac{\beta}{2}\mathrm{Tr}\left(\mathbf{A}^T\mathbf{L}\mathbf{A}\right)\right\}, \quad (3)$$

where w_{ij} controls the spatial intimacy between node i and node j. The summation term $\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\|\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_j\|^2$ in (3) can be written as $2\mathrm{Tr}\left(\mathbf{A}^T\mathbf{L}\mathbf{A}\right)$, where $\mathbf{L}$ is the well known symmetric positive semidefinite *graph Laplacian* matrix [9]. If we have additional image segmentation information, we can set w_{ij} to 1 when node i and node j are neighbors belonging to the same region and 0 otherwise. If we do not have such additional information, we can use $w_{ij} = e^{-\|\mathbf{y}_i-\mathbf{y}_j\|^2/2B\eta^2}$ when node i and node j are neighbors and 0 otherwise. (A similar Potts-Markov model was introduced in [5] with sampling methods used for optimization as opposed to this work wherein deterministic methods are used.)

Assuming that the endmembers are independent and each $\mathbf{m}_i$ is a multivariate Gaussian centered at $\mathbf{r}_i \in \mathbb{R}^B$ with covariance matrix $\boldsymbol{\Sigma}_i$, i.e. $p(\mathbf{m}_i|\mathbf{r}_i,\boldsymbol{\Sigma}_i) = \mathcal{N}(\mathbf{m}_i|\mathbf{r}_i,\boldsymbol{\Sigma}_i)$, the conditional probability density of the entire endmember set becomes the product of the independent components:

$$p(\mathbf{M}|\mathbf{R},\boldsymbol{\Theta}) = \mathcal{N}\left(\mathrm{vec}(\mathbf{M}^T)|\mathrm{vec}(\mathbf{R}^T),\,[\delta_{ij}\boldsymbol{\Sigma}_i]\right), \quad (4)$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\Sigma}_i|i=1,...,M\}$ and $[\delta_{ij}\boldsymbol{\Sigma}_i]$ is a block diagonal matrix. We assume that the Gaussian centroids $\mathbf{r}_i$ can be modeled using a physically reasonable prior. To impose uniqueness, we assume that the endmembers tightly surround the mixed pixels and furthermore smoothly vary with wavelength. To this end, we introduce the following density function for $\mathbf{R} = [\mathbf{r}_1,...,\mathbf{r}_M]^T$:

$$p(\mathbf{R}) \propto \exp\left\{-\frac{\rho_1}{4}\sum_{i=1}^{M}\sum_{j=1}^{M} u_{ij}\|\mathbf{r}_i-\mathbf{r}_j\|^2\right\} \times \exp\left\{-\frac{\rho_2}{4}\sum_{k=1}^{B}\sum_{l=1}^{B} v_{kl}\|\mathbf{r}^k-\mathbf{r}^l\|^2\right\} \quad (5)$$

where $\mathbf{r}^k$ denotes the kth column of $\mathbf{R}$. We set $u_{ij}=1,\ \forall i,j$ making every pair of endmembers close to each other (similar to [10, 11]). We also set v_{kl} to 1 when $|k-l|=1$ and 0 otherwise, obtaining similar values at neighboring wavelengths for each endmember. Akin to (3), we can write $p(\mathbf{R})$ as

$$p(\mathbf{R}) \propto \exp\left\{-\frac{\rho_1}{2}\mathrm{Tr}\left(\mathbf{R}^T\mathbf{H}\mathbf{R}\right)-\frac{\rho_2}{2}\mathrm{Tr}\left(\mathbf{R}\mathbf{G}\mathbf{R}^T\right)\right\}, \quad (6)$$

where $\mathbf{H}\in\mathbb{R}^{M\times M}$ and $\mathbf{G}\in\mathbb{R}^{B\times B}$ are the graph Laplacians constructed from $\{u_{ij}\}$ and $\{v_{kl}\}$ respectively.

2.2. The spatial likelihood and the posterior

Assuming the noise $\mathbf{n}_i$ to be zero mean with a variance of μ^2 at each wavelength and independent at all locations, the spatial likelihood density of $\mathbf{Y}$ can be derived (after some algebra) using the linear transformation in (2) and (4):

$$p(\mathbf{Y}|\mathbf{R},\boldsymbol{\Theta},\mathbf{A},\mu) = \mathcal{N}\left(\mathrm{vec}(\mathbf{Y}^T)|\mathrm{vec}\left((\mathbf{A}\mathbf{R})^T\right),\boldsymbol{\Sigma}_\mathbf{Y}\right) \quad (7)$$

where $\boldsymbol{\Sigma}_\mathbf{Y} = \left[\delta_{ij}\mu^2\mathbf{I}_B + \sum_{k=1}^{M}\alpha_{ik}\alpha_{jk}\boldsymbol{\Sigma}_k\right]$. Note that the covariance matrix in (7) is not a block diagonal matrix, which means the transformed rows in $\mathbf{Y}$ are not independent. Also note that the covariance matrix $\boldsymbol{\Sigma}_\mathbf{Y}$ is a large NB by NB matrix. To improve efficiency, we assume that the covariance matrix $\boldsymbol{\Sigma}_k$ for each endmember is diagonal (with each wavelength being independent) and that the variance of each wavelength is σ_k^2, i.e., $\boldsymbol{\Sigma}_k = \sigma_k^2\mathbf{I}_B$. Then, we obtain $\boldsymbol{\Sigma}_\mathbf{Y} = \mu^2\left(\mathbf{I}_N + \mathbf{A}\mathbf{D}\mathbf{A}^T\right)\otimes\mathbf{I}_B$ where $\mathbf{D}\in\mathbb{R}^{M\times M}$ is a diagonal matrix with diagonal elements $\kappa_i = \sigma_i^2/\mu^2$.

From the priors in (3), (6) and the spatial likelihood in (7), we can write the posterior as

$$p(\mathbf{R},\boldsymbol{\Theta},\mathbf{A},\mu|\mathbf{Y}) \propto p(\mathbf{A})p(\mathbf{R})p(\mathbf{Y}|\mathbf{R},\boldsymbol{\Theta},\mathbf{A},\mu). \quad (8)$$

After straightforward algebra, maximizing $\log\, p(\mathbf{R},\boldsymbol{\Theta},\mathbf{A},\mu|\mathbf{Y})$ can be shown to be equivalent to minimizing

$$\begin{aligned} E(\mathbf{R},\mathbf{D},\mathbf{A},\gamma) =& \gamma\mathrm{Tr}\left\{(\mathbf{Y}-\mathbf{A}\mathbf{R})^T\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{A}\mathbf{R})\right\} \\ &+ B\log|\mathbf{Q}| - NB\log\gamma + \beta\mathrm{Tr}\left(\mathbf{A}^T\mathbf{L}\mathbf{A}\right) \\ &+ \rho_1\mathrm{Tr}\left(\mathbf{R}^T\mathbf{H}\mathbf{R}\right) + \rho_2\mathrm{Tr}\left(\mathbf{R}\mathbf{G}\mathbf{R}^T\right) \quad (9) \end{aligned}$$

subject to

$$\mathbf{A}\geq 0,\ \mathbf{A}\mathbf{1}_M = \mathbf{1}_N,\ \mathbf{R}\geq 0 \quad (10)$$

where $\gamma = \mu^{-2} > 0$ and $\mathbf{Q} = \mathbf{I}_N + \mathbf{A}\mathbf{D}\mathbf{A}^T$ is a symmetric positive definite matrix. *This is the final MAP energy function used in the present work.*

2.3. Optimizing the objective function

There are 4 parameters in the optimization problem: $\mathbf{A},\mathbf{R},\mathbf{D}$ and γ. We first fix $\mathbf{D}$ and γ, optimize (9) with respect to $\mathbf{A}$ and $\mathbf{R}$ and then optimize the remaining pair keeping the first pair fixed. For the sake of simplicity, assume at first that there is no variability for all the endmembers, i.e., $\kappa_i = 0$, $\mathbf{Q} = \mathbf{I}_N$ (with $\mathbf{D}$ taken into account after $\mathbf{A}$ and $\mathbf{R}$ are obtained). Incorporating the linear constraint $\nu\|\mathbf{A}\mathbf{1}_M - \mathbf{1}_N\|^2$ into (9), and taking derivatives with respect to $\mathbf{A}$ and $\mathbf{R}$ and setting them to zero results in

$$\frac{\beta}{\gamma}\mathbf{L}\mathbf{A} + \mathbf{A}\left(\mathbf{R}\mathbf{R}^T + \frac{\nu}{\gamma}\mathbf{1}_M\mathbf{1}_M^T\right) = \mathbf{Y}\mathbf{R}^T + \frac{\nu}{\gamma}\mathbf{1}_N\mathbf{1}_M^T, \quad (11)$$

$$\left(\mathbf{A}^T\mathbf{A} + \frac{\rho_1}{\gamma}\mathbf{H}\right)\mathbf{R} + \frac{\rho_2}{\gamma}\mathbf{R}\mathbf{G} = \mathbf{A}^T\mathbf{Y}, \quad (12)$$

Table 1. Quantitative comparison for Pavia University.

Error	SCM	NCM	PCOMMEND
Asphalt	-	0.0484	**0.0335**
Meadows	0.0314	**0.0187**	-
Trees	**0.0078**	0.0285	0.0419
Painted Metal Sheets	**0.0832**	0.0907	0.1096
Bare Soil	**0.0190**	-	0.0748
Self-Blocking Bricks	**0.0771**	0.0979	0.1885
Shadows	0.0131	0.0372	**0.0043**
Average	**0.0386**	0.0536	0.0754

which are two *Sylvester equations* (as per control theory). We use the classical Krylov subspace method to solve these equations [12]. Given an initial condition for $\mathbf{R}$ (with principal component analysis used to determine the boundary of the pixels as the initial endmembers), we can alternatively update $\mathbf{A}$ and $\mathbf{R}$ based on (11) and (12) respectively (with $\mathbf{A}$ being re-normalized to satisfy the simplex constraint after updated). Assuming $\mathbf{A}$ and $\mathbf{R}$ are held fixed, γ and $\mathbf{D}$ can be similarly obtained by setting their corresponding derivatives to zero.

The choice of free parameters should be invariant w.r.t. N, M and B. From the banded diagonal nature of w_{ij} in (3), the parameter β should have a magnitude $\beta/\gamma = B\beta'$. Similarly, the parameters ν, ρ_1 and ρ_2 should have magnitudes $\nu/\gamma = B\nu'$, $\rho_1/\gamma = \rho_1' N/M^2$ and $\rho_2/\gamma = \rho_2' N/M$.

3. RESULTS

In the experiments, we compared SCM with NCM and PCOMMEND [11] for endmember extraction on two real datasets: Pavia University and Indian Pines. NCM was run using the SCM implementation without spatial information on the abundances and endmembers, i.e., $\beta' = 0$, $\rho_2' = 0$. In all experiments, we use the mean of absolute difference as the endmember error measure, i.e., $\frac{1}{MB}\sum_{i,j}|m_{ij} - m'_{ij}|$.

Pavia University. We ran SCM, NCM and PCOMMEND on the Pavia University dataset, each with 6 endmembers. The parameters for SCM were manually tuned to be $\eta = 0.05$, $\beta' = 0.01$, $\nu' = 0.1$, $\rho_1' = 0.1$, $\rho_2' = 0.1$. The parameters of PCOMMEND followed those in [11]. We compared the abundance maps with the ground truth and permuted the endmembers to calculate the error. Figure 1 shows the abundance maps from SCM. Figure 2 shows the resulting endmember spectra from these algorithms versus the corresponding ground truth endmember. From Figure 2, we see that for the commonly identified materials—trees, painted metal sheets, self-blocking bricks and shadows—SCM matches the first three best while PCOMMEND matches the shadows best. The quantitative errors are reported in Table 1, which shows that SCM performed best overall.

Indian Pines. We also tested all the algorithms on the Indian Pines dataset for $M = 4$. The parameters for SCM

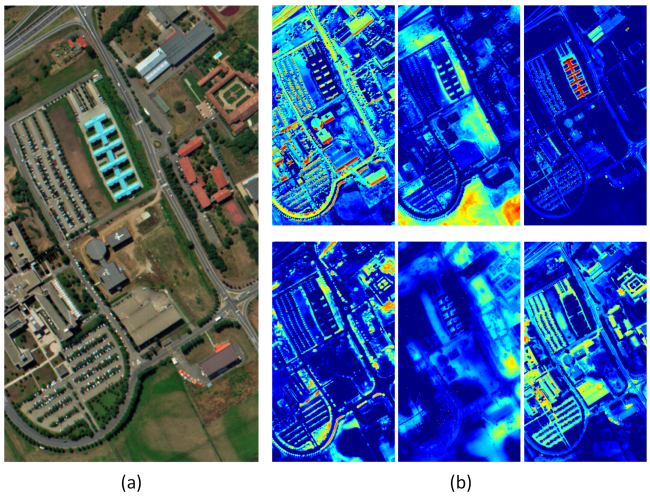

Fig. 1. (a) RGB image and (b) the abundance maps from SCM for Pavia University. The identified materials from top to bottom, left to right are shadows, meadows, painted metal sheets, trees, bare soil and self-blocking bricks respectively.

are $\eta = 0.01$, $\beta' = 0.01$, $\nu' = 0.1$, $\rho_1' = 1$, $\rho_2' = 0$. The parameters for PCOMMEND were set to be 2 clusters with 2 endmembers in each cluster. The other parameters of PCOMMEND were tuned to obtain the best possible results. The abundance maps for SCM, NCM and PCOMMEND are shown in Figure 3. From the results, we see that SCM has a segmented appearance while NCM and PCOMMEND have scattered dots within coherent regions.

4. DISCUSSION

In this paper, we propose a spatial compositional model (SCM) for linear unmixing. The advantages over previous linear unmixing models are the direct incorporation of spatial information in the hyperspectral likelihood and the estimation of endmember uncertainties. We show in the experiments that SCM outperforms closely related methods such as NCM and PCOMMEND. Future work will focus on estimation of the full covariance matrices for uncertainty modeling and optimization improvements.

5. 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. D. 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.

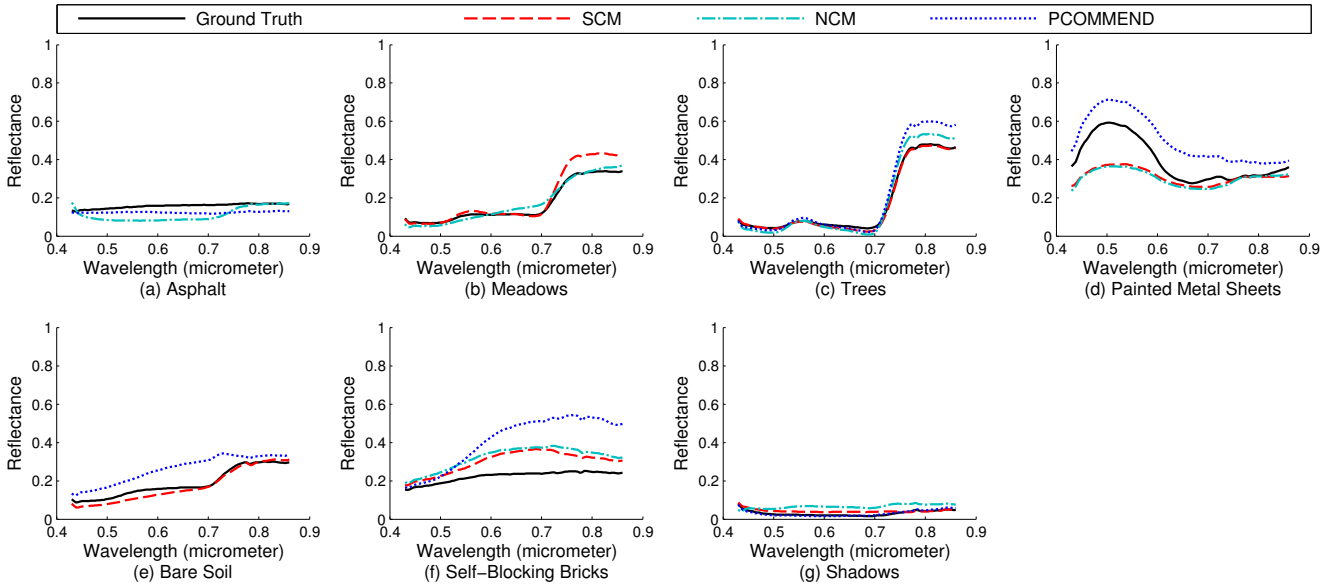

Fig. 2. Qualitative comparison of endmembers on Pavia University for SCM, NCM and PCOMMEND.

[3] 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 Trans. Image Processing*, vol. 19, no. 6, pp. 1403–1413, 2010.

[4] 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 Trans. on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7782–7792, 2014.

[5] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, 2011.

[6] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[7] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.

[8] A. Zare and P. D. Gader, "PCE: Piecewise convex endmember detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2620–2632, 2010.

[9] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[10] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.

[11] A. Zare, P. D. Gader, O. Bchir, and H. Frigui, "Piecewise convex multiple-model endmember detection and spectral unmixing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2853–2862, 2013.

[12] A. El Guennouni, K. Jbilou, and A. Riquet, "Block Krylov subspace methods for solving large Sylvester equations," *Numerical Algorithms*, vol. 29, no. 1-3, pp. 75–96, 2002.

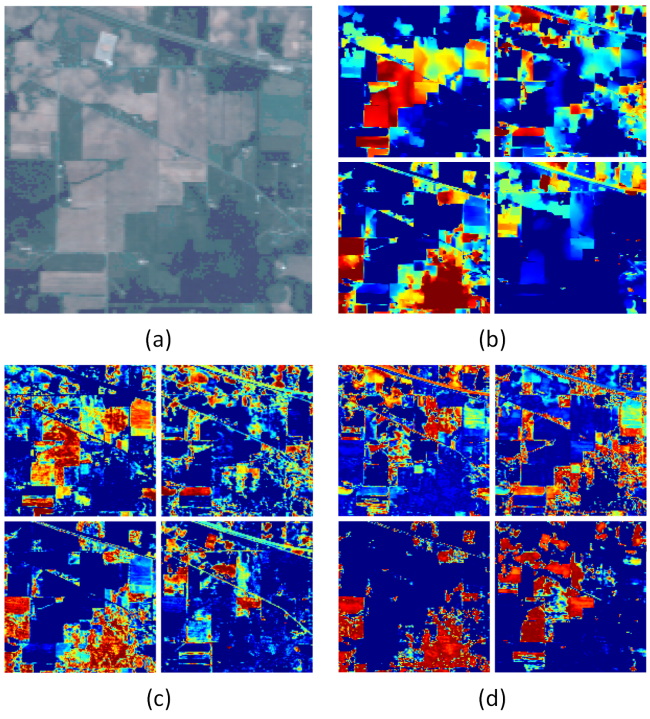

Fig. 3. (a) RGB image and the abundance maps from (b) SCM, (c) NCM and (d) PCOMMEND for Indian Pines.