Scientific Research

Area-Correlated Spectral Unmixing Based on Bayesian Nonnegative Matrix Factorization

Xiawei Chen, Jing Yu, Weidong Sun

Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: hawaiix@gmail.com, yujing@tsinghua.edu.cn, wdsun@tsinghua.edu.cn

ABSTRACT

To solve the problem of the spatial correlation for adjacent areas in traditional spectral unmixing methods, we propose an area-correlated spectral unmixing method based on Bayesian nonnegative matrix factorization. In the proposed method, the spatial correlation property between two adjacent areas is expressed by a priori probability density function, and the endmembers extracted from one of the adjacent areas are used to estimate the priori probability density functions of the endmembers in the current area, which works as a type of constraint in the iterative spectral unmixing process. Experimental results demonstrate the effectivity and efficiency of the proposed method both for synthetic and real hyperspectral images, and it can provide a useful tool for spatial correlation and comparation analysis between adjacent or similar areas.

Keywords: Hyperspectral Image; Spectral Unmixing; Area-Correlation; Bayesian Nonnegative Matrix Factorization

1. Introduction

Spectral unmixing is a technique to extract a collection of pure spectra (called *endmembers*), by decomposing mixed pixels of a hyperspectral image and estimating the corresponding fractions (called as *abundances*) of these endmembers [1]. Most of the existing spectral unmixing methods focus on the hyperspectral images themselves and no extra information is needed in the spectral unmixing procedure. Vertex Component Analysis (VCA) [2], Pixel Purity Index (PPI) [3] and N-FINDR [4] are widely‒used typical methods, whose basic idea is to extract the endmembers by identifying the vertexes of the simplex formed by the endmembers. VCA and PPI project the hyperspectral data into a low-dimensional space based on the fact that vertexes of a simplex are still vertexes even if the simplex is projected into a low-dimensional space, while N-FINDR tries to find a fixed number of pixels to minimize the volume of the simplex formed by these pixels. Iterative Error Analysis (IEA) [5] updates every endmember by selecting the pixel with the maximum error in the reconstructed image using the latest endmember set in an iterative process. A new trend is to apply nonnegative matrix factorization to the spectral unmixing, since all of elements in the endmember matrix and the abundance matrix are nonnegative [6, 7]. These methods mentioned above are characterized by no requirement for extra information, and therefore only a few constraints need be considered when they are implemented.

But in the case of comprehensive analysis, we need also consider some other information besides the hyperspectral image itself, such as spectra from the spectral library and hyperspectral images observed at different time or in different area. If we could bring extra information into the current framework of spectral unmixing, the spatial correlation and comparation analysis between different images can be carried out. Plazza *et al.* put forward an Automatic Morphological Endmember Extraction method (AMEE) using the correlation of spatial distribution for different substances [8]. Zortea *et al.* propose a pre-processing method on a pixel neighborhood scale to enhance the accuracy of spectral unmixing [9]. Zhang focuses on spectral unmixing supported by digital geomorphology model and geographical information system [10]. Dobigeon *et al.* present a semi-supervised spectral unmixing based on a hierarchical Bayesian model [11], where spectra from a spectral library are used in the spectral unmixing procedure. However, endmembers of hyperspectral images observed at different time or in different area may hold not only strong time or spatial correlation but also certain difference at the same time, the problem is how to use these kinds of correlations in the spectral unmixing procedure.

In this paper, we only concentrate on the spatial correlation between hyperspectral images observed in adjacent or similar areas. Our method is based on Bayesian nonnegative matrix factorization [12, 13] and the spatial correlation between hyperspectral images

observed in adjacent or similar areas is used in the form of a priori probability density function as a constraint to the spectral unmixing procedure. The remainder of this paper is organized as follows: Section 2 gives the detailed description of the proposed method; Section 3 discusses the experimental results both on synthetic and real hyperspectral data; Section 4 is the conclusions.

2. Spatial Correlation Method

2.1. Area-correlated Spectral Unmixing

A hyperspectral image X can be written in this form: $X = SA + E$, where $S = (s_1; s_2; \cdots; s_P) \in R^+_{L \times P}$ is the endmember matrix with the kth endmember represented by s_k, $A = (a_1; a_2; \cdots; a_P)^T \in R_{P \times N}$ is the abundance matrix in which a_k corresponds to the kth endmember and $E = (E_{i;j})_{L \times N} \in R_{L \times N}$ is the noise matrix with the number of endmembers P and the number of mixed pixels N. The noise $E_{i;j}$ is assumed to an i.i.d. white Gaussian noise with zero mean and variance σ^2. The objective of spectral unmixing is to find the proper S and A that minimize $\|X - SA\|_F^2$ with the Frobenius norm $\|\cdot\|_F$ (or other kinds of norm).

The basic idea of area-correlated spectral unmixing is to use the common endmembers in adjacent or similar areas as an aid in the spectral unmixing procedure. On one hand, they can be used as initial values after proper pre-processing with the current hyperspectral image; on the other hand, we bring them into the iterative spectral unmixing procedure as a priori knowledge.

The whole procedure of our area-correlated spectral unmixing method consists of three steps: pre-processing, Bayesian nonnegative matrix factorization and post-processing.

In pre-processing, for each endmember extracted from the adjacent area, we search for pixels similar to it in the current hyperspectral image. If the number of these similar pixels is too small, then we abandon this endmember, otherwise, we will calculate the average spectrum of these pixels and put it into the initial endmember matrix. We fill the rest of the initial endmember matrix with random pixels selected from the current hyperspectral image, and the abundance matrix will be also initialized randomly.

The similarity of two spectra is measured by the Spectral Correlation-coefficient Distance (SCD). The SCD of spectrum x and y is defined by Equation (1).

$$SCD(x;y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (1)$$

where $\bar{x} = \frac{1}{L}\sum_j x_j$, $\bar{y} = \frac{1}{L}\sum_j y_j$ and L is the number of bands.

Besides, we need also estimate the number of endmembers. If the estimated number is less than the real one, then it's evident that some endmembers cannot be extracted. To avoid this situation, our strategy is to increase the estimated number of endmembers.

After pre-processing, Bayesian nonnegative matrix factorization with the spatial correlation constraint, as described later in Section 2.2, will be applied to the current hyperspectral data to estimate the endmembers and their abundances. When the spectral unmixing procedure is finished, the extracted endmembers are clustered and the abundances are merged within every cluster in the post-processing stage.

2.2. Bayesian Nonnegative Matrix Factorization

Assuming that S, A and σ^2 are statistically independent, the Bayesian rule gives rise to

$$
\begin{aligned}
p(S; A; \sigma^2 | X) &= \frac{p(X | S; A; \sigma^2) p(S; A; \sigma^2)}{p(X)} \\
&= \frac{p(X | S; A; \sigma^2) p(S) p(A) p(\sigma^2)}{p(X)} \\
&\propto p(X | S; A; \sigma^2) p(S) p(A) p(\sigma^2)
\end{aligned} \quad (2)
$$

One of the possible ways is to find the S and A by maximizing Equation (2). But, due to the complexity of Equation (2), the optimization process would be computatively expensive. Instead, in Bayesian nonnegative matrix factorization, Gibbs sampling [14], a sampling method based on Markov Chain Monte Carlo (MCMC) is applied to estimate S, A and σ^2.

Gibbs sampling estimates the model parameters by sampling from the posterior probability density functions of these parameters. These samples will converge to the samples taken from the joint posterior probability density functions of these model parameters. Reference [14] gives the proof of the convergence of Gibbs sampling. The procedure of Gibbs sampling can be described as: in every round of sampling, we sample each parameter in turn; when we sample one of the parameters, we fix the rest and sample from the posterior probability density function of the current parameter; the sample is then used to update the posterior probability density function of the next parameter. Therefore, the posterior probability functions of S, A and σ^2 are required.

$E_{i;j}$ is assumed to an i.i.d white Gaussian noise, that is,

$$p(X | S; A; \sigma^2) = \prod_{i;j} N\left(X_{i;j}; \hat{X}_{i;j}; \sigma^2\right), \quad (3)$$

where $N(x; \mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ and $\hat{X} = SA$. An inverse Gamma distribution is chosen for the noise variance σ^2 and the posterior probability density function of the noise variance is given by

$$p\left(\sigma^2 \mid X; S; A\right) = G^{-1}\left(\sigma^2; k; \mu\right)$$

$$= \frac{\mu^k}{\Gamma(k)} \sigma^{-2(k+1)} e^{-\frac{\mu}{\sigma^2}}, \quad (4)$$

where $k = \frac{LN}{2} + k_0$ and $\mu = \frac{1}{2}\sum_{i,j} E_{i,j}^2 + \mu_0$.

According to the Bayesian rule, the posterior probability density function of s_k is

$$p\left(s_k \mid X; S_{\overline{k}}; A; \sigma^2\right) \propto N(s_k; \eta_{s_k}; \Sigma_{s_k}) p(s_k), \quad (5)$$

where $S_{\overline{k}}$ is the remaining matrix after deleting the kth column of the endmembers matrix S,

$N(x; \eta; \Sigma) = \frac{1}{\sqrt{(2\pi)^L \det(\Sigma)}} e^{-\frac{1}{2}(x-\eta)^T \Sigma^{-1}(x-\eta)}$ is the

L-dimensional Gaussian distribution function, and η_{s_k} and Σ_{s_k} is the mean vector and the covariance matrix of s_k and these parameters are estimated as

$$\eta_{s_k} = \left(\eta_{S_{1;k}}; \eta_{S_{2;k}}; \cdots; \eta_{S_{L;k}}\right),$$

$$\eta_{S_{i;k}} = \frac{\sum_j \left(X_{i,j} - \sum_{m \neq k} S_{i;m} A_{m;j}\right) A_{m;j}}{\sum_j A_{k;j}^2}$$

$$\Sigma_{s_k} = \text{diag}\left\{\sigma_{S_{1;k}}^2; \sigma_{S_{2;k}}^2; \cdots; \sigma_{S_{L;k}}^2\right\}, \quad (6)$$

$$\sigma_{S_{i;k}}^2 = \frac{\sigma^2}{\sum_j A_{k;j}^2}$$

$$i = 1; 2; 3; \cdots; L.$$

The posterior probability density functions of the endmembers are actually multiple-dimensional Gaussian distribution functions weighted by their priori probability density functions. The next step is to choose proper priori probability functions for the endmembers. We adopt the following distribution as the priori probability density function of the endmember s_k:

$$p(s_k) \propto \begin{cases} 1 - \gamma e^{-\beta v_k^T s_k} & 0 \leq s_k \leq 1 \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where the vector v_k is orthogonal to all the column vectors of $S_{\overline{k}}$, 0 and 1 represent all-zero and all-one L-dimensional vectors respectively and γ is the parameter for adjusting. One of the reasons we form the priori probability density function of the endmembers in this way is that given a linear combination c of the column vectors of $S_{\overline{k}}$, we have $v_k^T c = 0$, and therefore the corresponding density value will be zero, which reduces the possibility of exacting a spectrum mixed by the existing endmembers. Another reason is that when projected to v_k, endmembers tend to have larger values of $v_k^T s_k$ which leads to larger probability density values. By substituting the priori probability density function in Equation (7), the posterior probability density function of the endmember s_k is written by

$$p\left(s_k \mid X; S_{\overline{k}}; A; \sigma^2\right)$$

$$\propto \begin{cases} N(s_k; \eta_{s_k}; \Sigma_{s_k})\left(1 - \gamma e^{-\beta v_k^T s_k}\right) & 0 \leq s_k \leq 1. \\ 0 & \text{otherwise} \end{cases} (8)$$

As to the abundances, similarly, the posterior probability density function of the abundance a_k is

$$p\left(a_k \mid X; S; A_{\overline{k}}; \sigma^2\right) \propto N(a_k; \eta_{a_k}; \Sigma_{a_k}) p(a_k) \quad (9)$$

where $A_{\overline{k}}$ represent the remaining matrix after deleting the kth row of A with the estimation of η_{a_k} and Σ_{a_k}

$$\eta_{a_k} = \left(\eta_{A_{k;1}}; \eta_{A_{k;2}}; \cdots; \eta_{A_{k;N}}\right),$$

$$\eta_{A_{k;j}} = \frac{\sum_i S_{i;k}\left(X_{i,j} - \sum_{n \neq k} S_{i;n} A_{n;j}\right)}{\sum_i S_{i;k}^2}$$

$$\Sigma_{a_k} = \text{diag}\left\{\sigma_{A_{k;1}}^2; \sigma_{A_{k;2}}^2; \cdots; \sigma_{A_{k;N}}^2\right\}, \quad (10)$$

$$\sigma_{A_{k;j}}^2 = \frac{\sigma^2}{\sum_i S_{i;k}^2}$$

$$j = 1; 2; 3; \cdots; N.$$

Since the kth endmember is assumed to distribute randomly in the observed area, the priori probability density function of the abundances is modeled as the uniform distribution. Thus, the posterior probability density function of the abundance a_k becomes

$$p\left(a_k \mid X; S; A_{\overline{k}}; \sigma^2\right) \propto \begin{cases} N(a_k; \eta_k; \Sigma_k) & 0 \leq a_k \leq 1 \\ 0 & \text{otherwise} \end{cases}. \quad (11)$$

When estimating the model parameters mentioned above, the method of Slice Sampling [15] is used to sample from the posterior probability density functions of these parameters.

3. Experiments and Discussions

3.1. Experiments on Synthetic Data

Ten spectra are selected from the spectral library of the United States Geological Survey (USGS) for generating synthetic hyperspectral images. For each pixel in the synthetic hyperspectral image, we first pick several spectra from the ten spectra, and then generate a set of abundances which sum to one, and finally mix the spectra linearly by their corresponding abundances.

Two synthetic hyperspectral images are generated to simulate the hyperspectral images of two adjacent areas. Six spectra are used to generate the first image while the second image has eight, and four spectra are shared between them. The white Gaussian noise is added to each image with the SNR up to 30dB approximately. Firstly, the first image is pre-processed using the Vertex Component Analysis (VCA) method, to extract the endmembers from it as the initial.

Figure 1 shows the result, where the red curves represent the real spectra and the endmembers extracted by our method are represented by the blue curves. **Figure 1(a)** is the final result of our method, in which the first row corresponds to the four initial endmembers from pre-processing and the second row is the other extracted new endmembers obtained from the spectral

unmixing procedure. **Table 1** lists the Spectral Correlation-coefficient Distances and the Spectral Angle Distances (SAD) between the real spectra and the extracted endmembers. **Figure 1(b)** plots the error between the original image and the reconstructed image, where $error = \|X - SA\|_F^2$ with the Frobenius norm $\|\phi\|_F$. As we can see from **Figure 1(a)** and **Table 1**, the exacted endmembers fit the real spectra well. The error between the original image and the reconstructed image decreases sharply with the sampling process approaching, which indicates that our method is able to estimate the endmember matrix and the abundance matrix correctly.

Arngren *et al.* give another form of priori probability density function with a volume constraint [16]. The volume of the simplex formed by the endmembers is used as a constraint in the priori probability density function of the endmember matrix. **Figure 2(a)** shows the results using the two different priori probability density functions, where the red curves are still the real spectra, the green curves are results under the volume constraint and the blue curves are results of our method. The results of our method fits the real spectra better. **Figure 2(b)** shows the error comparison between the two methods. At the beginning, the errors of both methods decrease almost at the same pace. With the sampling process going on, the error of our method is dropping faster than the method with the volume constraint.

3.2. Experiments on Real Data

Two adjacent square areas of size 100×100 are cropped from the AVIRIS data. The VD method [17] and the Hysime method [18] are used to estimate the number of endmembers of both images. **Table 2** gives the estimation of both methods showing that the Hysime method gives a larger estimated number.

Table 1. SCD and SAD between the real spectra and the extracted endmembers shared by both images.

Endmember	SCD	SAD
1	0.950335	7.303504
2	0.973790	8.976964
3	0.977021	7.848447
4	0.997171	5.941172

Table 2. Estimation of numbers of endmembers in both images using VD and Hysime method.

	VD method	Hysime method
Image 1	3	16
Image 2	4	18

According to [17], the VD method gives a relatively better estimation than other methods when applied to the AVIRIS data. **Figure 3(a)** illustrates the extracted endmembers and the corresponding abundances. Five spectra remain after clustering, which consistent with the result of the VD method. **Figure 3(b)** plots the error between the original image and the reconstructed image, showing that the error declines rapidly in the first few rounds of iteration and approaches zero in the end.

To avoid missing some endmembers, we increase the estimated number of endmembers. But, there are two disadvantages of this strategy. One is that some endmembers may be extracted repeatedly and the other is that mixed spectrum may be extracted. The solution to the first problem is cluster. As to the second problem, one possible solution is considering the abundances while clustering. If a mixed spectrum was extracted, it is most likely that its abundance will be small, which can be used as a clue to discard the poorly extracted endmembers.

Furthermore, our method is based on the fact that there exist some common endmembers between adjacent areas, which means that it should be applied in some certain occasions. For instance, a large hyperspectral image cannot be processed immediately. The large image can be split into several smaller pieces and then our method is applied to the small hyperspectral images. Another application is for areas that are not adjacent geographically yet sharing some common ground surface types.

4. Conclusions

In this paper, we propose an area-correlated method of spectral unmixing based on Bayesian nonnegative matrix factorization. Common endmembers within adjacent areas are used to assist the spectral unmixing procedure. Experiment results show that these endmembers can serve as a kind of priori knowledge in the procedure to improve the results. It is thought that the proposed method can provide a useful tool for spatial correlation and comparation analysis between adjacent or similar areas. Future researches will focus on investigating other correlations between the hyperspectral images of adjacent areas, optimizing the sampling process and how to use the correlations more efficiently.
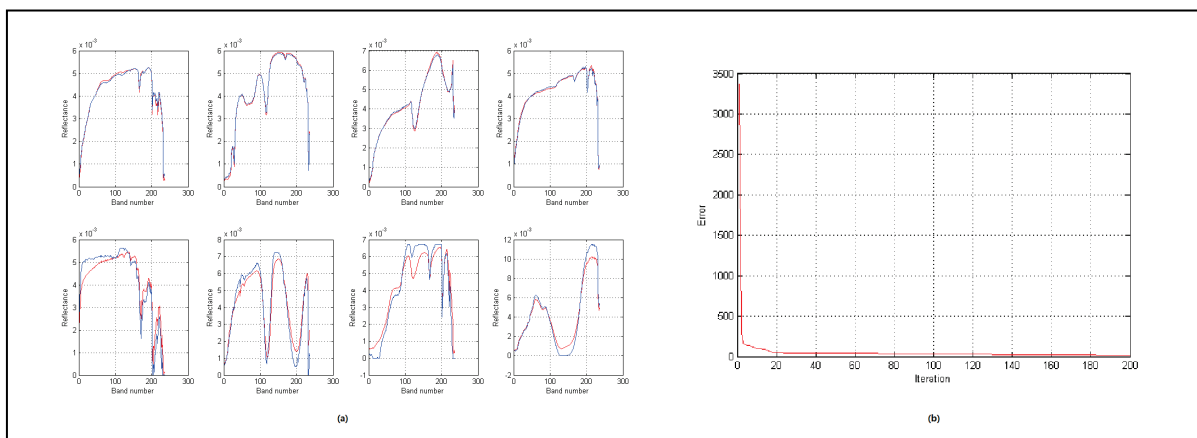
X. CHEN *ET AL*.

Figure 1. Experimental results on the synthetic data, (a)Comparison of the real spectra and the extracted endmembers, (b)Error between the original image and the reconstructed image.
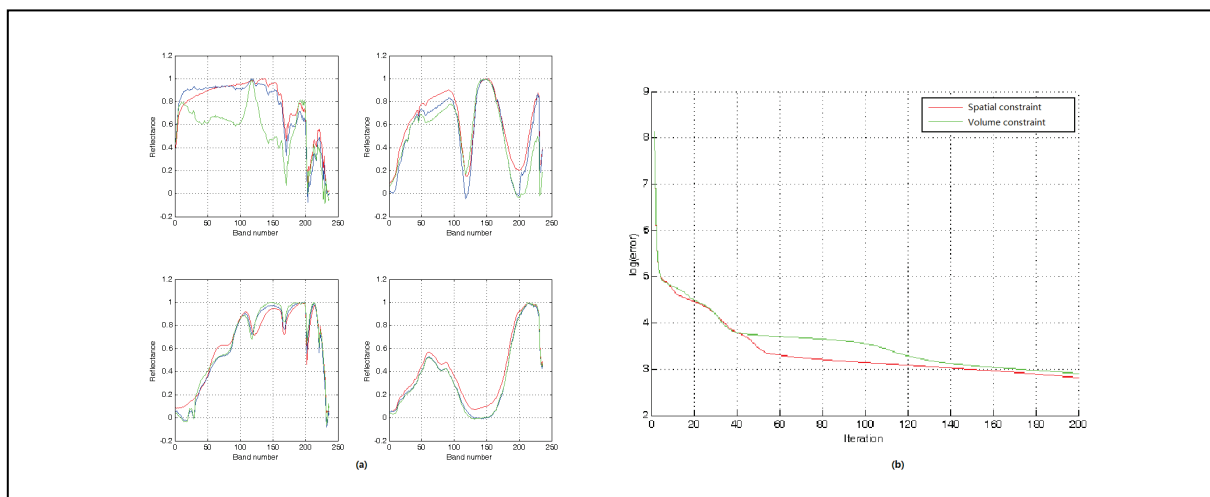

Figure 2. Comparison results of two different priori probability density functions, (a)Extracted endmembers of the two methods, (b)Error of the two methods.
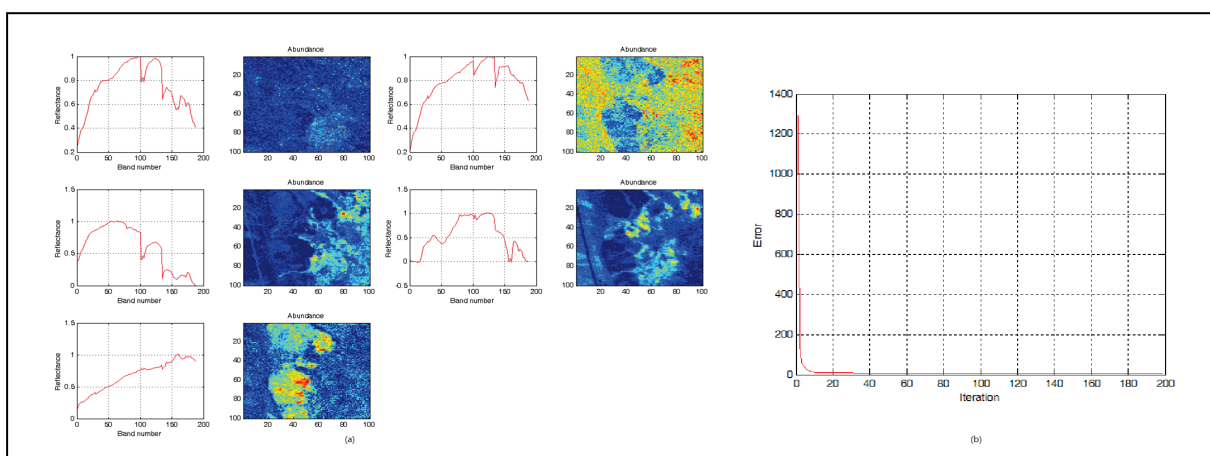

Figure 3. Experimental results on the real data, (a)Endmembers extracted from AVIRIS data with their abundances, (b)Error between the original image and the reconstructed image.

5. Acknowledgements

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral Unmixing," *Signal Processing Magazine, IEEE*, 2002. 19(1): p. 44-57. doi:10.1109/79.974727

[2] J. Nascimento and J. Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, 2005. 43(4): p. 898-910. doi:10.1109/TGRS.2005.844293

[3] J. W. Boardman, F. A. Kruse and R. O. Green, "Mapping Target Signatures via Partial Unmixing of AVIRIS Data," 1995: Pasadena, CA.

[4] M. E. Winter, "N-FINDR: An Algorithm for Fast Autonomous Spectral End-member Determination in Hyperspectral Data," *Proceedings of SPIE*, 1999. p. 266-275. doi:10.1117/12.366289

[5] L. Sun, Y. Zhang and B. Guindon, "Improved Iterative Error Analysis for Endmember Extraction from Hyperspectral Imagery," *Proceedings of SPIE*, 2008. doi:10.1117/12.799232

[6] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," *Advances in Neural Information Processing Systems*, 2001: p. 556-562.

[7] V. P. Pauca, J. Piper and R. J. Plemmons, "Nonnegative Matrix Factorization for Spectral Data Analysis," *Linear Algebra and Its Applications*, 2006. 416(1SI): p. 29-47. doi:10.1016/j.laa.2005.06.025

[8] A. Plaza, P. Martinez, R. Perez and J. Plaza, "Spatial/spectral Endmember Extraction by Multidimensional morphological Operations," *IEEE Transactions on Geoscience and Remote Sensing*, 2002. 40(9): p. 2025-2041. doi:10.1109/TGRS.2002.802494

[9] M. Zortea and A. Plaza, "Spatial Preprocessing for Endmember Extraction," *IEEE Transactions on Geoscience and Remote Sensing*, 2009. 47(8): p. 2679-2693. doi:10.1109/TGRS.2009.2014945

[10] B. Zhang, "Hyperspectral Data Mining Supported by Temporal and Spatial Information," PH.D. Thesis, Institute of Remote Sensing Applications, Chinese Academy of Sciences, 2002.

[11] N. Dobigeon, J. Y. Tourneret and C. Chein-I, "Semi-supervised Linear Spectral Unmixing Using a Hierarchical Bayesian Model for Hyperspectral Imagery," *IEEE Transactions on Signal Processing*, 2008. 56(7): p. 2684-2695. doi:10.1109/TSP.2008.917851

[12] M. N. Schmidt, "Linearly Constrained Bayesian Matrix Factorization for Blind Source Separation," *Advances in Neural Information Processing Systems*, 2009.22: p. 1624-1632.

[13] M. N. Schmidt, O. Winther and L. K. Hansen, "Bayesian Non-negative Matrix Factorization," *Independent Component Analysis and Signal Separation*, 2009. p. 540-547.

[14] G. Casella and E. I. George, "Explaining The Gibbs Sampler," *American Statistician*, 1992. 46(3): p. 167-174. doi:10.1080/00031305.1992.10475878

[15] R. M. Neal, "Slice Sampling," *Annals of Statistics*, 2003. 31(3): p. 705-767. doi:10.1214/aos/1056562461

[16] M. Arngren, M. N. Schmidt and J. Larsen, "Bayesian Nonnegative Matrix Factorization with Volume Prior for Unmixing of Hyperspectral Images," *Proceedings of the 2009 IEEE International Workshop on Machine Learning for Signal Processing* (MLSP 2009), 2009: p. 6 pp.-6 pp. doi:10.1109/MLSP.2009.5306262

[17] C. I. Chang and Q. Du, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2004. 42(3): p. 608-619. doi:10.1109/TGRS.2003.819189

[18] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral Subspace Identification," *IEEE Transactions on Geoscience and Remote Sensing*, 2008. 46(8): p. 2435-2445. doi:10.1109/TGRS.2008.918089