

# Unsupervised Classification of Remote Sensing Imagery With Non-negative Matrix Factorization

Cheng-Yuan Liou<sup>†</sup> and Kuang-De Ou Yang

Department of Computer Science and Information Engineering, National Taiwan University

**Abstract**—An unsupervised classification method provides the interpretation, feature extraction and endmember estimation for the remote sensing image data without any prior knowledge of the ground truth. We explore such method and construct an algorithm based on the non-negative matrix factorization (NMF). The use of the NMF is to match the non-negative property in sensing spectrum data. The data dimensionality is estimated by using the partitioned noise-adjusted principal component analysis (PNAPCA). The initial matrix used to start the NMF is obtained by using the fuzzy c-mean (FCM). This algorithm is capable to produce a region- or part-based representation of objects in images. Both simulated and real sensing data are used to test the algorithm.

## I. INTRODUCTION

In recent years, remote sensing images are widely used in many fields, including agriculture, geology, military intelligence etc. With the development of hyper-spectral sensors, hundreds bands of high resolution data of the same area can be acquired at the same time. Therefore hyper-spectral imagery analysis has become one of the most powerful and fastest growing technologies in remote sensing. In remote sensing, multi-spectral and hyper-spectral sensors acquire huge amount of data without knowing ground quality in advance. People need unsupervised methods to analyze the imagery. Principal components analysis (PCA) is the one applied in many analysis. PCA calculates the first and second order statistics from the data and removes the correlation redundancy by finding a rotated orthogonal coordinate system. Thus PCA finds a set of orthonormal bases and projects images arranged in order of decreasing variance. Independent component analysis (ICA) [2] tries to find a set of hidden variables as statistical independent as possible in the sense of maximizing a certain function that measures independency. The motivation to use NMF is because the remote sensing model is positively defined and NMF is based on positive restrictions. That is, in stead of assuming statistical independence like PCA and ICA, NMF [8] [9] assumes that the hidden variables and the features are non-negative. Therefore the resultant matrixes will be more intuitive and interpretable.

With different initial conditions, NMF may lead to different local optima. In order to obtain a satisfactory result, initialization is crucial and needs to setup properly. The parameters needed to be determined in NMF algorithm include : 1) The

intrinsic dimensionality. In remote sensing, that's the number of endmembers appear in the imagery. 2) The signature matrix. Each column of the matrix represents the normalized spectral reflectance of the material. 3) The membership matrix. Each column represents the proportion of each endmember within the pixel.

To determine the number of endmembers in remote sensing data, observing the eigenvalues derived from PCA is an option. The drawback is that it needs a subjective judgement for the threshold. If the noise covariance can be estimated in advance, the minimum noise fraction (MNF) [7] can effectively solve the inherent dimensionality problem. Lee et al. further interpreted this transform as noise-adjusted principal component analysis (NAPCA) [4]. However, inaccuracy of the noise estimation will degrade NAPCA's ability to estimate the intrinsic dimensionality. Tu et al. proposed partitioned noise-adjusted principal component analysis (PNAPCA) [11] to resolve inaccurate estimation of noise in NAPCA and effectively resolve the dimensionality problem.

To initial the feature matrix, Stefan [10] proposed to use Spherical K-Means to estimate the initial the cluster centroids. However Spherical K-Means can't describe the fuzzy boundry between clusters very well because it assigns each data vector to a certain cluster with the membership equal to 1. Thus it can't provide a good estimation of the membership matrix because of the multiplicative update characteristics of NMF. Instead, Fuzzy C-Means(FCM) relax the constraint and allows each data vector belongs to a certain cluster to some degree. Therefore FCM is capable to provide more accurate initial cluster centroids and the membership matrix simultaneously. The drawback is it takes longer computation time compared with Spherical K-Means.

The rest of the paper is organized as follows. Section II formulates the images as a linear spectral mixture model. Section III reviews NMF algorithm and explore the relationship between the image model. In section IV, the initialization and evaluation approaches are briefly introduced. Section V presents two experiments with simulated and real remote sensing data respectively, and the results of FCM and NMF with random or FCM initialization will also be compared. Concluding remarks are finally made in Section VI.

## II. PROBLEM STATEMENT AND SIGNAL MODEL

Linear Spectral Unmixing assumes that the reflectance spectrum of pixels is the result of linear combinations of the

This work has been partially supported by National Science Council under project NSC 93-2213-E-002-008. <sup>†</sup>Correspondent, Dept. CSIE, NTU, Taipei, Taiwan 106, ROC, email: cylieou@csie.ntu.edu.tw

endmembers' spectra within the pixel, and the weight is the percentage of each endmember in the area covered by the pixel. A linear mixed model with noise for a hyper-spectral image pixel can be described by Eq. (1):

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad (1)$$

where  $\mathbf{x}$  is a  $m \times 1$  column vector represents the reflectance spectrum and  $m$  is the number of bands. Reflectance matrix  $\mathbf{A}$  is a  $m \times r$  matrix denoted by  $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r)$  where  $\mathbf{a}_i$  is a  $m \times 1$  column vector represents the reflectance spectrum of the  $i_{th}$  endmember and  $m$  represents the number of endmembers. Abundance vector  $\mathbf{s}$  is a  $r \times 1$  vector and denotes the fraction of the  $r$  endmembers present in  $x$ .  $\mathbf{n}$  is an  $m$ -dimensional random vector that describes additive noise. The correlation matrix of  $\mathbf{x}$  becomes

$$\begin{aligned} \mathbf{R}_\mathbf{x} &= E\{\mathbf{x}\mathbf{x}^T\} = E\{(\mathbf{A}\mathbf{s} + \mathbf{n})(\mathbf{A}\mathbf{s} + \mathbf{n})^T\} \\ &= \mathbf{A}E\{\mathbf{s}\mathbf{s}^T\}\mathbf{A}^T + \mathbf{A}E\{\mathbf{s}\mathbf{n}^T\} + E\{\mathbf{n}\mathbf{s}^T\}\mathbf{A}^T + E\{\mathbf{n}\mathbf{n}^T\} \\ &= \mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{A}^T + \mathbf{A}\mathbf{R}_{\mathbf{s}\mathbf{n}} + \mathbf{R}_{\mathbf{n}\mathbf{s}}\mathbf{A}^T + \mathbf{R}_\mathbf{n} \end{aligned}$$

Noise vector  $\mathbf{n}$  is assumed to have zero mean and uncorrelated with the vector  $\mathbf{s}$ . Then the cross correlation term  $\mathbf{R}_{\mathbf{n}\mathbf{s}}$  and  $\mathbf{R}_{\mathbf{s}\mathbf{n}}$  both equal to zero. The correlation matrix  $\mathbf{R}_\mathbf{x}$  becomes

$$\mathbf{R}_\mathbf{x} = \mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{A}^T + \mathbf{R}_\mathbf{n} = \mathbf{R}_\alpha + \mathbf{R}_\mathbf{n} \quad (2)$$

where  $\mathbf{R}_\alpha$  is the signal correlation matrix and its rank equals to  $r$ .

### III. NONNEGATIVE MATRIX FACTORIZATION

NMF, proposed by Lee and Seung [8], is a matrix factorization algorithm under the non-negativity constraints. Given a  $m \times n$  non-negative matrix  $\mathbf{X}$ , NMF tries to find a non-negative  $m \times r$  matrix  $\mathbf{W}$  and a non-negative  $r \times n$  matrix  $\mathbf{H}$  such that  $\mathbf{X} \approx \mathbf{W}\mathbf{H}$ . The parameter  $r$  is the desired rank of matrix  $\mathbf{W}$  and usually chosen to be smaller than  $n$  and  $m$ . The size will be reduced from  $m \times n$  to  $r \times (m + n)$ . Therefore the product  $\mathbf{W}\mathbf{H}$  can be regarded as a compressed form of the data in  $\mathbf{X}$ .

NMF does not allow negative elements in the matrix factors  $\mathbf{W}$  and  $\mathbf{H}$ . To represent the data vector  $\mathbf{x}_j$ , NMF combines non-negative multiple basis  $\mathbf{w}_i$  such that

$$\mathbf{x}_j = \sum_{i=1}^r \mathbf{w}_i * h_{ij}, 1 \leq j \leq n \quad (3)$$

, but no subtractions can occur because  $h_{ij}$  is also non-negative. In short, the entries of  $\mathbf{H}$  combine these basis  $\mathbf{w}_i$  to generate a whole. Thus  $\mathbf{W}$  will be part-based representation of  $\mathbf{X}$ . That is how non-negativity constraint learns a parts-based representation and combines parts to form a whole. Also, compare Eq. (2) and (3), we find that NMF can be used to model the generation of image data, i.e.

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \approx \mathbf{W}\mathbf{h}. \quad (4)$$

Thus we may apply NMF to the unmixing problem of remote sensing imagery.

NMF is very easy to use and implement. In [9], two different multiplicative update algorithms are analyzed and proved to monotonically decrease the objective function.

## IV. INITIALIZATION AND EVALUATION APPROACHES

### A. Initialization Approaches

For hyper-spectral imagery, it is reasonable to assume that the number of the endmembers is much less than the number of image bands. Therefore the number of endmembers should be estimated first then dimension reduction can be carried out to release the burden of computation. The risk is that to reduce the dimension too much, some components will be excluded from the reduced data.

Tu et al. [11] proposed partitioned noise-adjusted principal components analysis (PNAPCA) to solve the intrinsic dimensionality of remote sensing data. PNAPCA is a partitioned version of NAPCA. It partitions the original data space into two distinct subspaces by a simultaneous transform. In addition, it applies a simple union-intersection margin test (UIMT) which is able to estimate the endmember accurately.

After the number of endmembers is estimated, Fuzzy C-Means is suggested to initialize the signature and member matrix of NMF. Fuzzy C-Means was developed by Dunn [5] in 1973 and improved by Bezdek [6] in 1981. Assuming that we want to partition data vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  into  $k$  disjoint clusters  $\pi_1, \pi_2, \dots, \pi_k$ , the objective function of Fuzzy C-Means can be defined as:

$$J = \sum_{i=1}^k J_i = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m d_{ji}, \quad (5)$$

represents the distance from any given data point to a cluster center weighted by the data point's membership degree. The fuzzy parameter  $f \in [1, \infty)$  is used to define membership fuzziness. Higher value of  $f$  will make the clustering result fuzzier. Notably, if  $f$  is set to 1, Fuzzy C-Means becomes exactly K-means algorithm.

Fuzzy C-Means algorithm based on cosine similarity has been named Hyper-spherical Fuzzy C-Means (H-FCM). It's because it works on normalized high dimensional data vectors that lie in hyper-sphere of unit radius. Thus we normalize the data vector and define the dissimilarity function  $d_{ji}$  in Eq. (5) as  $d_{ji} = 1 - s_{ji}$  where  $s_{ji} = \mathbf{x}_j^T \mathbf{c}_i$  is the cosine similarity between the data vector  $\mathbf{x}_j$  and the cluster centroid  $\mathbf{c}_i$ .

By minimizing the constrained objective function (5), we can obtain the update equations (6) for cluster centroid  $\mathbf{c}_i$  and (7) for membership  $u_{ij}$  respectively:

$$\mathbf{c}_i = \sum_{j=1}^n u_{ij}^f \mathbf{x}_j \left[ \sum_{\alpha=1}^k \left( \sum_{j=1}^n u_{ij}^f x_{j\alpha} \right)^2 \right]^{(-1/2)} \quad (6)$$

$$u_{ij} = \sum_{\beta=1}^k \left( \frac{d_{j\beta}}{d_{\beta i}} \right)^{-\frac{1}{(f-1)}} = \sum_{\beta=1}^k \left( \frac{1 - \mathbf{x}_j^T \mathbf{c}_i}{1 - \mathbf{x}_j^T \mathbf{c}_\beta} \right)^{-\frac{1}{(f-1)}} \quad (7)$$

To sum up, the steps we propose to initialize NMF include to estimate the cluster number with PNAPCA first, then initialize

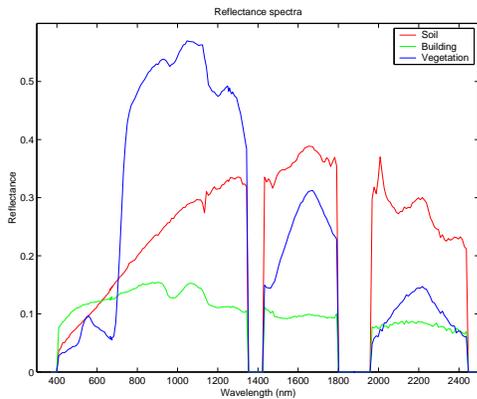


Fig. 1. Signatures of soil, building, and vegetation extracted from Aviris data are used to generate the simulated data.

the signature and membership matrix with the clustering result of H-FCM.

### B. Evaluate Clustering Results With Xie-Beni Index

To evaluate the clustering result, Xie [1] proposed the compactness and separation index  $S$  defined as:

$$S = \frac{\kappa}{s}. \quad (8)$$

The compactness  $\kappa$  is defined as the ratio of the total variation  $\sigma$  to the size  $n$  of the dataset, that is,  $\kappa = \frac{\sigma}{n}$ . Smaller  $\kappa$  indicates the clusters are more compacted. The separation  $s$  is defined as the minimum distance between the cluster centroids and formulated as  $s = d_{min}^2$ . Larger  $s$  indicates the clusters are more separated. According to Eq. (8), a smaller  $S$  represents a better partitioning because the clusters are compacted and separated to each other.

## V. EXPERIMENTAL ANALYSIS

### A. Evaluation with Simulated Data

Fig. 1 shows the reflectance spectra of soil, building and vegetation extracted from a Aviris image scene of the Moffett Field [3]. The bands corresponding to the water absorption regions and with negative values are removed before processing and kept the other 187 bands in this study. These three endmembers are used to generate a  $60 \times 60$ -pixel hyperspectral image cube. Each image is formed by twelve stripes of five-pixels wide. Each strip is the linear mixture of the reflectance spectra of the three endmembers. The corresponding membership assignment is listed in table I. Fig. 2 shows the membership plots for each endmember. White Gaussian Noise is added to each image to generate SNR of 30:1. In the end, we create a  $60 \times 60 \times 187$  image scenes. Fig. 3 shows the first three images of the simulated data.

Table II lists the initial condition of H-FCM, NMF with random and with H-FCM initialization. Note that the fuzzy parameter  $f$  of H-FCM is setup empirically. Table VII lists the clustering results compared with Xie-Beni Index. We find that H-FCM finds the most compacted clusters with less separated signatures. On the contrary NMF finds the most separated

TABLE I

MEMBERSHIP ASSIGNMENT FOR STRIPES IN SIMULATED DATA

Stripe	1	2	3	4	5	6	7	8	9	10	11	12
Soil	1	0	0	0.8	0	0.2	0.6	0.1	0.3	0.5	0.2	0.3
Building	0	1	0	0.2	0.8	0	0.3	0.6	0.1	0.3	0.5	0.2
Vegetation	0	0	1	0	0.2	0.8	0.1	0.3	0.6	0.2	0.3	0.5

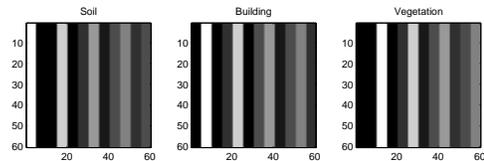


Fig. 2. The real membership plots of soil, building and vegetation in the simulation.

signatures with less compacted clusters. However, NMF with H-FCM initialization provides better clustering result of the three techniques according to the index.

Since we already know the real membership matrix listed in table I, matrix approximation can also be evaluated by summing up the Frobenius error between the resultant membership matrix and the real matrix. The result is listed in table IV and the better one is presented in bold face. We find that NMF with H-FCM initialization provides better estimation of the membership matrix  $\mathbf{H}$ . Fig. 4 shows the comparison of the membership plots. The cosine similarity values between the real and the estimated signatures are listed in table V. The highest similarity of each signature is presented using bold face. Notably, all the signatures acquired by NMF with H-FCM initialization are most similar to the real signatures. Fig. 5 shows the comparison between the real signatures and the signatures acquired with different techniques.

In this section, three different techniques are implemented on the simulated data and the clustering results are evaluated. We observe that H-FCM did provide better initialization so that NMF can find more accurate signatures and membership plots. In the next section, the proposed algorithm will be implemented on real data.

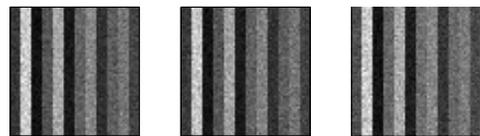


Fig. 3. The first three bands of the simulated data.

TABLE II

INITIAL CONDITION SETUP

	HFCM	NMF(Random)	NMF(H-FCM)
C	3	3	3
f	1.6	N/A	N/A
W	random	random	H-FCM
H	random	random	H-FCM

TABLE III  
EVALUATE CLUSTERING RESULT WITH XIE-BENI INDEX

	H-FCM	NMF(Random)	NMF(H-FCM)
Compactness $\kappa$	<b>0.0074</b>	0.0722	0.0128
Separation $s$	0.0733	<b>0.3038</b>	0.1869
$S = \frac{\kappa}{s}$	0.1004	0.2375	<b>0.0683</b>

TABLE IV  
MEMBERSHIP MATRIX APPROXIMATION OF DIFFERENT CLUSTERING APPROACHES

	H-FCM	NMF(Random)	NMF(H-FCM)
Frobenius error	26.31	22.43	<b>15.14</b>

### B. Evaluation with AVIRIS Data

AVIRIS remote sensing data is now used as the test imagery, and the test area is Cuprite, NV [3]. For computation consideration, we subset a 200 by 200 pixel square area from the Cuprite imagery. Fig. 6 is the subset area image composed by band 27, band 17, and band 7. Some bands effected seriously by water absorption or with low SNR are removed. In the end, we have 192\*40000 band by pixel data matrix  $X$ . The data is divided by 10000 to retrieve the original reflectance value and normalized to unit length. Since the resolution of Aviris sensor is 20 meter, our goal is to improve the resolution by sub-pixel extraction. That is, to extract the spectral signatures from the imagery and determine the proportion of each endmembers within each pixel.

At first, PNAPCA is applied to estimate the number of the endmembers appeared within the AVIRIS dataset. Table VI shows the values of PNAPCA sorted in descending order. For clarity, only the first ten significant values are listed. By counting the number of value larger than 1, PNAPCA can clearly estimate the number of endmembers appear in the scene should be eight.

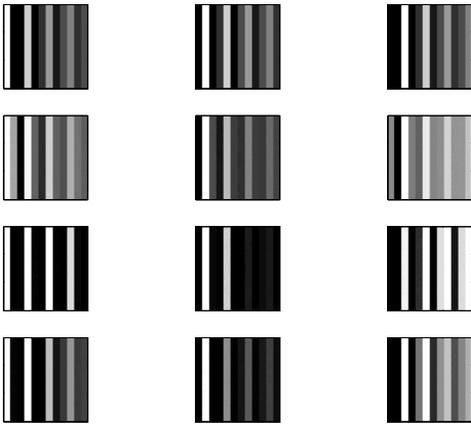


Fig. 4. Comparison of the membership plots with different clustering approaches. The first row is the real membership plots, the second row is acquired by NMF, the third row is acquire with H-FCM, and the last row is acquired by NMF with H-FCM initialization.

TABLE V  
THE COSINE SIMILARITY BETWEEN THE REAL SIGNATURES AND THE ESTIMATED SIGNATURES

	H-FCM	NMF(Random)	NMF(H-FCM)
Soil	0.9964	0.9095	<b>0.9998</b>
Building	0.9957	0.9282	<b>0.9999</b>
Vegetation	0.9883	0.9946	<b>0.9988</b>

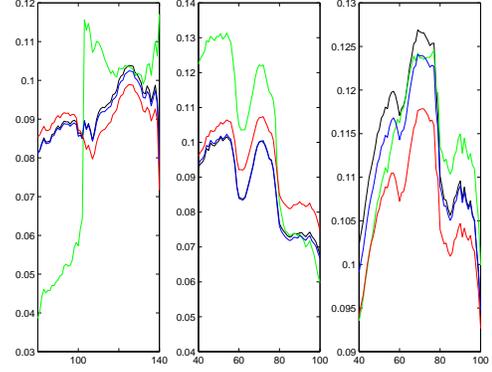


Fig. 5. Comparison of the resultant NMF signatures with different initialization. In each plot, the real signature is drawn in black solid line. The signatures of NMF are in green. The signatures of H-FCM are in red. Signatures of NMF with H-FCM initialization are drawn in blue. Only fifty bands are selected for clarity reason.

After the number of endmembers is estimate, we may now proceed to cluster the data. In the initial condition, we set the cluster number as eight, randomly choose eight samples in the data set as the initial centroids of each cluster, and randomly initialize the membership matrix for H-FCM. Owing to a variety of unknown noises appeared in the real remote sensing data, we need to setup the fuzzy parameter  $f$  properly. Xie-Beni index is now used to estimate suitable fuzzy parameter  $f$ . Fig. 7 shows Xie-Beni index of different fuzzy parameter values. We observe that 1.3 should be a better estimation for the fuzzy parameter  $f$ .

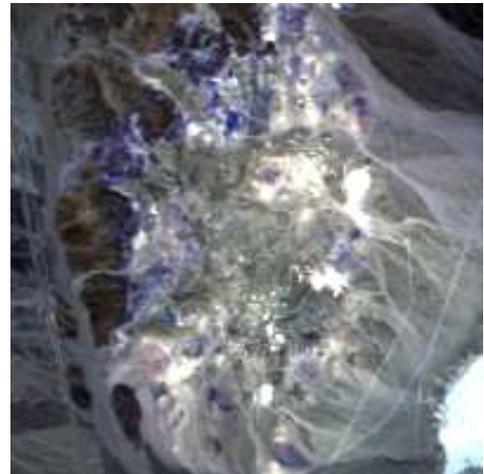


Fig. 6. 200\*200 color image of the Cuprite, NV region [3].

TABLE VI  
THE FIRST TEN SIGNIFICANT VALUES OF PNAPCA.

PNAPCA	8.97e+7	6.29e+4	4.48e+4	3.24e+2	2.19e+2
	1.24e+1	5.34	1.53	3.68e-1	2.83e-1

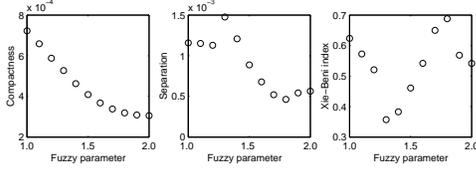


Fig. 7. Compactness (left), Separation (middle) and The Xie-Beni Index (right) of different fuzzy parameter from 1.0 to 2.0

Next, NMF is used to improve the cluster result. That is, to initialize the matrix  $\mathbf{W}$  and  $\mathbf{H}$  with the resultant matrix of H-FCM. NMF stops when the change of the objective function is less than a threshold  $\epsilon$ . Fig. 8 and 9 show the signatures and the membership matrix of each class obtained by NMF. Note that, in each membership plot, the brightness of the pixels represents the proportion of the endmembers within the pixel.

To compare the consequence with random initialization, Fig. 10 and 11 show the signatures and membership plots of NMF with random initialization. From Fig. 10, we can observe that the signatures with random initialized are discontinuous. Instead, signatures with H-FCM initialization shown in Fig. 8 all have good and smooth continuity. In Fig. 11, all of the membership plots look too noisy and hardly tell the feature appearance. However, with H-FCM initialization, the appearance of features shown in Fig. 9 are very clear. Therefore, we may conclude that NMF with H-FCM initialization provides better classification result than random initialization.

Table VII lists Xie-Beni index of the clustering results of H-FCM, and NMF with random or H-FCM initialization. Although the membership of NMF with H-FCM initialization is less compacted than H-FCM and the signatures are less separated than random initialization, Xie-Beni index shows that NMF initialized by H-FCM provides better clustering result of the three.

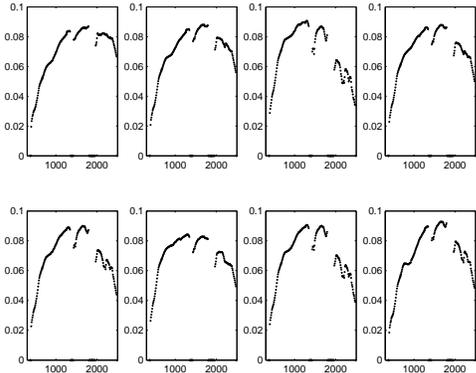


Fig. 8. NMF centroids of clusters with H-FCM initialization.

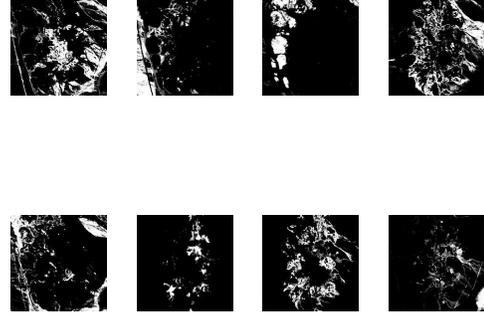


Fig. 9. NMF membership plots of clusters with H-FCM initialization.

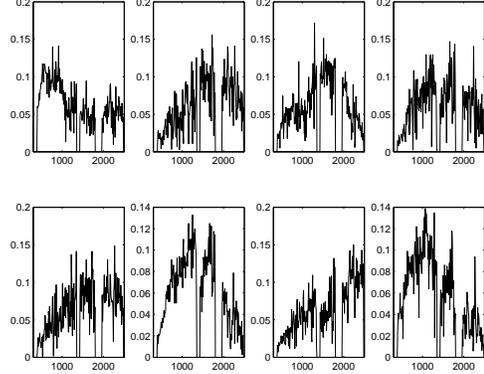


Fig. 10. NMF centroids of clusters with random initialization.

## VI. CONCLUSION

In this work, we have presented the result of spectral unmixing of remote sensing data with NMF technique initialized with PNAPCA and H-FCM. In the initial condition, we have to determine the data intrinsic dimensionality  $r$ , the signature matrix  $\mathbf{W}$  and the membership matrix  $\mathbf{H}$ . To determine  $r$ , most techniques like NAPCA, noise needs to be estimate accurately. However, the variety of unknown noise appeared in remote sensing data makes it hard to estimate correctly. PNAPCA overcomes the noise estimation error by a simultaneous transform of the partitioned data and applied a simple hypothesis test UIMT to accurately estimate the number of the endmembers. Since NMF is an iterative process and promised to converged to a local optimun, the initial



Fig. 11. NMF membership plots of clusters with random initialization.

TABLE VII

XIE-BENI INDEX OF DIFFERENT CLUSTERING RESULTS.

	H-FCM	NMF	NMF(H-FCM)
Compactness $\kappa$	<b>5.27e - 4</b>	2.16e-2	5.38e-4
Separation $s$	1.47e-3	<b>5.32e - 2</b>	1.73e-3
$S = \frac{\kappa}{s}$	0.3571	0.4067	<b>0.3094</b>

conditions of  $\mathbf{W}$  and  $\mathbf{H}$  will directly affect the consequence. H-FCM, an iterative process like NMF, is suggested to initialize  $\mathbf{W}$  and  $\mathbf{H}$ . H-FCM allows each data vector belongs to a cluster to some degree. Though it is less efficient, it is more stable and able to initialize both the signature matrix and the membership matrix simultaneously. For real imagery data, the fuzzy parameter of H-FCM is crucial and needs to be setup properly according to the data characteristic. Xie-Beni index provides a good indication according to the H-FCM clustering result. In the experiments with simulated and real imagery data, the proposed technique both provides satisfactory results. We also found that NMF doesn't take too many iterations with good estimations of the initial conditions.

## REFERENCES

- [1] X. Xie and G. Beni (1991) A Validity for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Learning*, Vol 13, No. 8, August 1991, 841-847.
- [2] A. J. Bell and T. T. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7, 1129-1159 (1995).
- [3] California Institute of Technology. AVIRIS (Airborne Visible/Infrared Imaging Spectrometer)homepage. [Online]. <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>
- [4] J. B. Lee, A. S. Woodyatt, and M. Berman (1990) Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform. *IEEE Trans. Geos. Remote Sensing*, 28,295-304.
- [5] J. C. Dunn (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* ,3 ,32-57
- [6] J. C. Bezdek (1981) Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York
- [7] A. A. Green, M. Berman, P. Switzer, and M. Craig (1988)A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geos. Remote Sensing*, 26(1), 65-74.
- [8] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788-791, 1999.
- [9] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. NIPS, 2000.
- [10] S. M. Wild, Seeding Non-Negative Matrix Factorizations with the Spherical K-Means Clustering, Thesis for the Department of Applied Mathematics, University of Colorado (April 2003).
- [11] T. M. Tu, H. C. SHYU., Y. S. SUN, and C. H. LEE. Determination of data dimensionality in hyperspectral imagery-PNAPCA. *Multidimensional Systems and Signal Processing*, 10,255-274 (1999).