Self-Estimation of the Number of Endmembers for Hyperspectral Mixture Analysis

Ouiem Bchir
College of Computer and Information Sciences, Computer Science department, King Saud University, Riyadh, KSA
Email: obchir@ksu.edu.sa

***Abstract*—Spectral unmixing consists in finding a set of spectrally pure components (endmembers) and their corresponding fractions coverage for each pixel (abundances) in the hyperspectral data. Most the existing approaches consider the number of endmembers as input to their algorithms. In this paper, we propose a new mixture analysis method which relies on a spectral summarization algorithm that is inspired from convex geometry modelling, and works directly on the whole bandwidth range. It learns the endmembers and the abundances based on the information provided by the spectral summarization of the original scene. Also, it automatically optimizes the number of endmembers required by a particular scene. This optimization is achieved using the Competitive Agglomeration clustering algorithm.**

***Index Terms*—hyper-spectral data, mixture analysis, unmixing**

I. Introduction

The quality of hyperspectral images depends on the performance of the hyperspectral sensors, which emit a signal toward the scene and receive the reflected radiance. This procedure generates hyperspectral information representing both the spectral and the spatial data using a three-dimensional data cube. More specifically, the obtained hyperspectral cubes consist of a stack of 2D images representing a range of electromagnetic spectrum. Each pixel within these images matches the radiance at one specific location. Ideally, each material in a given scene has a unique spectral signature. This assumption relies on a fundamental property that is the spectral reflectance. In other words, the reflectance of most materials has different wavelength because the energy is absorbed differently by distinct materials. However, we should mention that usually information carried by hyperspectral image cannot be processed straightforward. It has to be considered under specific framework such as spectral unmixing analysis [1], supervised classification or segmentation [2]. Therefore several image processing techniques along with machine learning approaches have been adopted to mine useful information from hyperspectral images [3].

Most of the exiting works on hyperspectral mixture rely on the assumption that the number of endmembers is known apriori. A natural solution to find this number relies on the brute force approach, which consists in trying several numbers of endmembers, and evaluates the resulting unmixing accuracy. Then, the number of endmembers which yields the best performance is considered as the optimal number. The main drawback of this exhaustive search is its computational cost. Moreover, considering the unmixing performance as criteria to find the optimal number of endmembers is not straightforward. The Mixture Analysis based on Spectral Summarization (MASS) proposed in [4] is a spectral unmixing approach. It uses all the wavelength of the hyperspectral image and assumes a convex geometry model in order to estimate the endmembers and their corresponding abundances. MASS unmixing technique is based on the information provided by the summarization of the hyperspectral image. The summarization is performed using FCM clustering algorithm [5].

In this work, we propose a novel mixture analysis approach which can be considered as an extension to the MASS algorithm [4]. In addition to the endmembers and their corresponding abundances estimation, the proposed algorithm self-estimates the number of endmembers of a particular hyperspectral scene. More specifically, we adopt the competitive agglomeration clustering approach CA [6] which categorizes the data into homogeneous clusters, and learns the number of clusters. We formulate the hyperspectral abundances using the fuzzy membership functions learned by CA [6]. Similarly, a relation between the centers of the clusters returned by CA [6] and the endmembers is outlined.

II. Mixture Analysis with Self-Estimation of the Number of Endmembers

We propose an unmixing technique named hyperspectral Mixture Analysis with Self-Estimation of the Number of Endmembers (MASENE) which consists of two main steps. The first one is a summarization step, and the second one is mixture step. The summarization step consists in obtaining a quantitative description of the hyperspectral data using unsupervised machine learning technique. Namely, we adopt the competitive agglomeration CA clustering algorithm [6] which partitions the data into homogenous subsets without prior specification of the number of clusters.

Let $X=\{x_1, \ldots x_N\}$ be the set of pixel, C the number of clusters learned by CA, $U=\{\mu_{ij}\}$ the fuzzy membership matrix obtained by CA, $A=\{a_1, \ldots a_N\}$ the center matrix learned by CA [6].

Manuscript received February 2, 2015; revised June 10, 2015.

doi: 10.18178/joig.3.1.56-59

Let us define the Mxd matrix K as follows:

$$K=\begin{bmatrix} a_1 \\ a_2 \\ \dots \\ \dots \\ a_{C-1} \\ a_C \end{bmatrix}\begin{bmatrix} \sum_j^N \mu_{11} & \sum_j^N \mu_{21} \dots\dots\dots\dots \sum_j^N \mu_{C1} \\ \sum_j^N \mu_{22} & \sum_j^N \mu_{22} \dots\dots\dots\dots \sum_j^N \mu_{C2} \\ & \dots \\ & \dots \\ \sum_j^N \mu_{1d} & \sum_j^N \mu_{2d} \dots\dots\dots\dots \sum_j^N \mu_{Cd} \end{bmatrix} \quad (1)$$

The center update equation in [6], can be written as

$$K=(UU^T)(UU^T)^{-1} \quad (2)$$

Since

$$(UU^T)(UU^T)^{-1}=Id \quad (3)$$

where Id is the CxC identity matrix. Equations (2) and (3) yield the following set of equations

$$K=(UU^T)(UU^T)^{-1}K \quad (4)$$

Equation (4) yields

$$X=U^T(UU^T)^{-1}K \quad (5)$$

We intend to exploit the result shown in (5) in order to unmix the hyperspectral data assuming a convex geometry model [7]. This model is defined as follows

$$X=PE \quad (6)$$

Subject to

$$\sum_{i=1}^{M} p_{ij}=1, \text{and } 0<p_{ij}<1, \text{for } \ j\in\{1,\dots,N\}. \quad (7)$$

In (6), $P=\{P_i\}_{i=1\dots n}$ is the abundance matrix, and $E=\{e_i\}_{i=1\dots n}$ is the endmember matrix.

The convex geometry model imposes that all abundances must be positive between 0 and 1, and their sum over all endmembers is equal to one with respect to each considered pixel. These are the same constraints that have to be satisfied by the fuzzy memberships as expressed in [6]. Moreover, the transpose of fuzzy membership matrix U^T, and the abundance matrix have the same size, which is M×N, if we equally set the number of clusters C and the number of endmembers M. Referring to (5) and (6) and taking into consideration the above observations, we have

$$P=U^T \quad (8)$$

and

$$E=(UU^T)^{-1}K \quad (9)$$

where K is defined in (2).

Equations (8) and (9) define the model of new proposed unmixing approach; the hyperspectral Mixture Analysis with Self-Estimation of the Number of Endmembers (MASENE). CA [6] provides the number of clusters automatically, which yields the number of endmembers. MASENE is summarized below.

Algorithm 1: MASENE

Input: The data X, the number of cluster M*max.*
Output: The estimated number of endmembers M**,** E, and P
Begin

- Cluster the pixels $X=\{x_1, \dots x_N\}$ using CA algorithm.
- Compute the matrix of endmembers E using (9).
- Compute the matrix of abundances P using (8).

End

III. Experiments

In order to assess the performance of the proposed approach, we use it for hyperspectral mixture analysis on three standard real hyperspectral datasets. Namely, the real datasets are Indian Pine data [8], Botswana data [9], Pavia [10], and the Salinas data [11]. We compare the performance of MASENE to the most relevant existing approach SPICE [12]. We set the maximum number of clusters to 100.

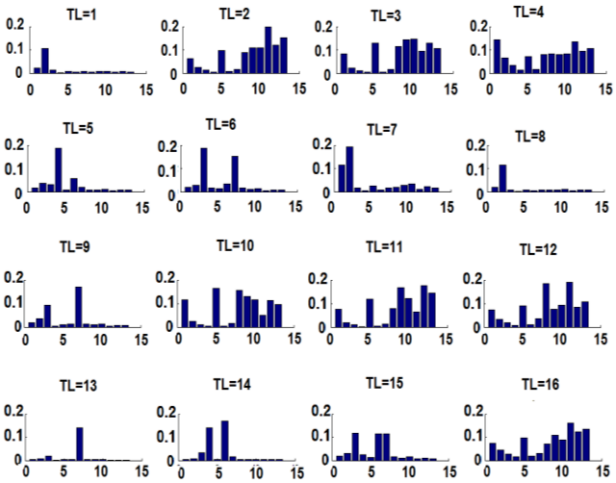

Figure 1. Abundance fraction histograms obtained using MASENE with Indian Pine data [8]

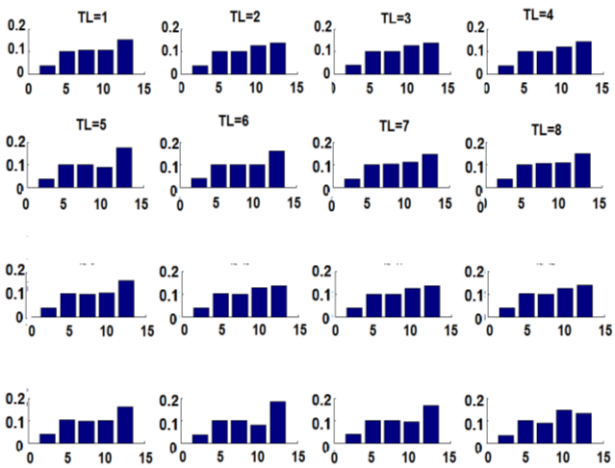

Figure 2. Abundance fraction histograms obtained using SPICE [12] with Indian Pine data [8]

Fig. 1 and Fig. 2 display the obtained abundance fraction histograms for Indian Pine data [8] using MASENE and SPICE [12], respectively. Similarly, Fig. 3,

Fig. 4, and Fig. 5, we show the abundance fraction histograms obtained using MASENE unmixing results for Botswana data [9], Pavia data [10], and Salinas data [11], respectively. Also, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 show the abundance fraction histograms obtained using SPICE for Botswana data [9], and Pavia data [10], respectively. As it can be seen, MASENE gives better unmixing results compared to SPICE [12] with the four considered real data sets. The abundance fractions histograms corresponding to the unmixing results obtained using MASENE are presented mainly by less endmembers than the histograms corresponding to SPICE [12] unmixing results.

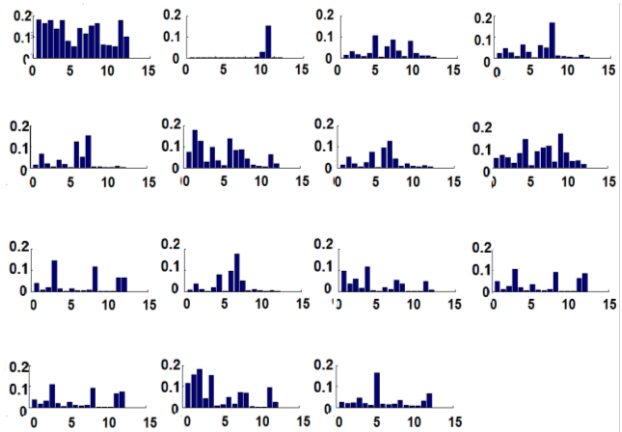

Figure 3. Abundance fraction histograms obtained using MASENE with Botswana data [9]

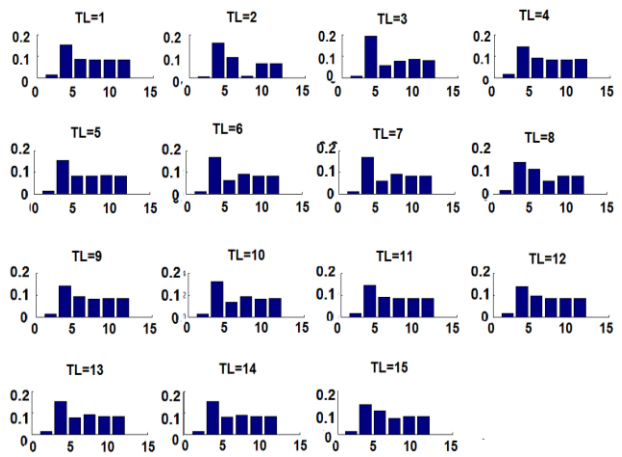

Figure 4. Abundance fraction histograms obtained using SPICE [12] with Botswana data [9]

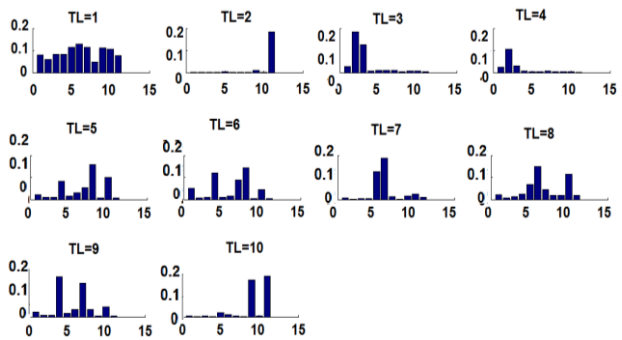

Figure 5. Abundance fraction histograms obtained using MASENE with Pavia data [10]

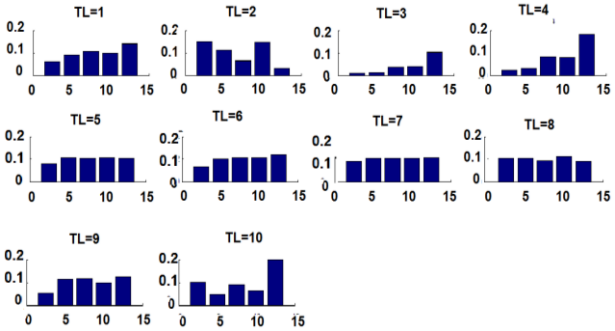

Figure 6. Abundance fraction histograms obtained using SPICE [12] with Pavia data [10]

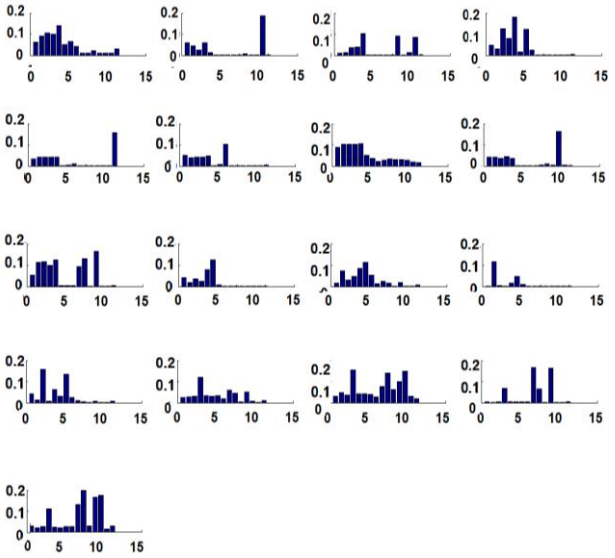

Figure 7. Abundance fraction histograms obtained using MASENE with Salinas data [11]

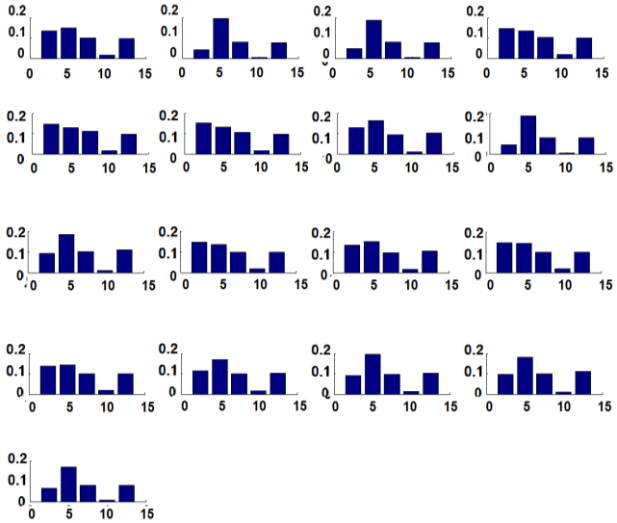

Figure 8. Abundance fraction histograms obtained using SPICE [12] with Salinas data [11]

Also, Fig. 7 and Fig. 8 show the abundance fraction histograms obtained using SPICE for Salinas data [11].

As it can be seen, MASENE gives better unmixing results compared to SPICE [12] with the four considered real data sets. The abundance fractions histograms corresponding to the unmixing results obtained using

MASENE are presented mainly by less endmembers than the histograms corresponding to SPICE [12] unmixing results.

TABLE I. SIMULATION TIMES IN SEC FOR MASENE AND SPICE [12] WITH RESPECT TO THE FOUR CONSIDERED REAL DATA SETS.

	Indian Pine	**Botswana**	**Pavia**	**Salinas**
SPICE [12]	214.37	$5.5\ 10^8$	$1.5\ 10^5$	$2.1\ 10^3$
MASENE	132.73	$1.86\ 10^3$	$2.3\ 10^3$	$1.2\ 10^3$

In Table I, we report the obtained running times in sec for MASENE and SPICE [12] for the four real data sets. In this experiment, we use an HP NIVADA Laptop, core i7, 8 G RAM and MATLAB 10.

IV. CONCLUSION

In this paper, we have proposed a new hyperspectral unmixing approach named hyperspectral Mixture Analysis with Self-Estimation of the Number of Endmembers (MASENE). MASENE discovers the quantitative structure of the hyperspectral data using the competitive agglomeration (CA) algorithm [6]. The fuzzy partition of the data, the centroids, and the number of clusters, learned using CA, are then used to formulate the proposed unmixing approach. MASENE self-estimates the endmembers in an unsupervised manner, and overcomes the challenge of finding the optimal number of endmembers in an unsupervised manner.

REFERENCES

[1] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander," *Journal of Geophysical Research*, vol. 91, pp. 8098-8112, 1986.

[2] X. Jia, J. A. Richards, and D. E. Ricken, *Remote Sensing Digital Image Analysis: An Introduction*, Berlin: Springer-Verlag, 1999.

[3] P. K. Varshney and M. K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer Verlag, 2004.

[4] O. Bchir, M. M. B. Ismail, and H. Frigui, "Mixture analysis based on spectral summarization," in *Proc. WHISPERS*, Gainesville, Florida, USA, 2013.

[5] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, New York: Plenum Press, 1981.

[6] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 450-465, May 1999.

[7] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, pp. 44-57, Jan. 2002.

[8] Hyperspectral images. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

[9] Grupo de Inteligencia Computacional. [Online]. Available: http://www.ehu.es/ccwintco/uploads/7/72/Botswana.mat

[10] University of Pavia. [Online]. Available: http://tlclab.unipv.it/

[11] Grupo de Inteligencia Computacional. [Online]. Available: http://www.ehu.es/ccwintco/uploads/f/f1/Salinas.mat

[12] A. Zare, "Hyperspectral endmember detection and band selection using Bayesian methods," Ph.D. dissertation, Dept. Comp. Sci., University of Florida, 2008.

Dr. Ouiem Bchir is assistant professor at the computer science department, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh, Saudi Arabia. She got her PhD from the University of Louisville, KY, USA. Her research interests are Spectral and kernel clustering, pattern recognition, hyperspectral image analysis, local distance measure learning, and Unsupervised and Semi-supervised machine learning techniques. She received the University of Louisville Dean's Citation, the University of Louisville CSE Doctoral Award, and the Tunisian presidential award for the electrical engineering diploma.