I J C T A, 9(3), 2016, pp. 1443-1451

Geometrical Endmember Extraction and Linear Spectral Unmixing of Multispectral Image

K. Niranjani*, and K. Vani**

Abstract: Accurate mapping is prepared using Linear unmixing of satellite images. Endmember extraction contributes the unmixing accuracy. In this paper, Endmembers are extracted using different Geometrical algorithms like Pixel Purity Index (PPI), Nearest Finder (N-FINDR) and Sequential Maximum Angle Convex Cone (SMACC) algorithms. Extracted Endmembers are given as input for unmixing and it is attempted using Linear mixing model. Here, Landsat 4-5 Thematic Mapper dataset is tested. Experimental results are compared and it is inferred that SMACC algorithm performs better compared to the PPI and NFINDR algorithm. PPI shows less performance in determining the endmember that are less contributed in a pixel.

Keywords: Multispectral image, Linear Spectral Unmixing, Endmember Extraction

1. INTRODUCTION

Satellite Image classification is attempted by two approaches namely hard classification and soft classification. In hard classification, the decision is ‘Winner take all’, with only one label permitted at each pixel, that is, each pixel belongs to the class it most closely resembles. In soft classification, the decision is multi-valued, with the possibility of more than one label per pixel, that is, each pixel belongs to more than one class and has membership grades for each class. There are two types of sub-pixel classification, namely Spectral unmixing and fuzzy classification.

Satellite images are used for preparing maps with high accuracy and updated details. High resolution images provide more accurate maps. Every pixel does not belongs to a single class. Each pixel consists of many classes like soil, water, vegetation, urban area etc. Those pixels are called the mixed pixels. For creation of accurate maps, unmixing of mixed pixels should be done. Spectral unmixing is used for determining the fractions of materials in the scene. Endmembers are the pure signature for a class. Pure Endmember identification contributes the classification accuracy. The Extracted Endmembers must be spectrally pure for improving the classification accuracy. Spectral unmixing is accompanied by Linear and Non-linear mixing models. Linear and Non-linear mixing model assumes single and multiple reflectance respectively. In Linear mixing model, each incident photon interacts with one earth surface component only and that reflected spectra do not mix before entering the sensor. Linear mixing model assumes single reflectance. The sunlight from the sun reaches the earth surface in which it reacts with only one earth surface component and this do not mix before entering the sensor. Let E_f be the fraction of endmember and S be the Spectrum of the Image which is given by $N \times n$ where N is the number of bands and n be the total number of pixels in the Image, e is the error (or) the noise in the Image.

$$X = E_f S + e \tag{1}$$

* Research Scholar, Department of IST, Anna university CEG, *Email: nirairam@auist.net*

** Associate Professor, Department of IST, Anna university CEG, *Email: vanirrk@gmail.com*

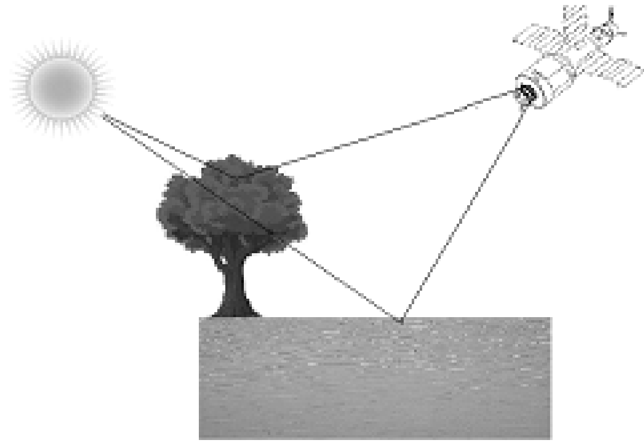


Figure 1: Linear mixing model

Spectral unmixing decomposes a mixed pixel into a set of spectra called endmembers as well as the corresponding abundances. In this way, the endmembers spectral information within the mixed pixel can be revealed[1]. A simplex geometric theory Minimum Volume Simplex Analysis(MVSA) is used as an Unmixing approach by fitting a minimum volume to the data. Major problem in MVSA is to solve the quadratic problem. The inequality can be solved by using several optimization criteria. Multi-GPU implementation is introduced as an optimization criteria[2]. Spectral unmixing is a source separation problem. when using linear mixing model, the sources are the fractional abundances and the endmember spectral signatures are the columns of the mixing matrix. Independent component analysis (ICA) framework is used to unmix spectral data[3].

Vertex Component Analysis (VCA) is an unsupervised endmember extraction method for Linear mixtures. VCA assumes the presence of pure pixels in the image. This algorithm iteratively projects data onto orthogonal direction to the endmember already determined. The algorithm iterates until all the endmembers are estimated[4]. Linear Spectral Mixture Analysis is a widely used technique in remote sensing for finding the spectral signatures. It imposes two constraints namely Sum to one constraint and Abundance Non-negativity constraint. Fully constraint Least square method is developed which uses Least square error and Orthogonal Subspace Projection(OSP) method for eliminating the drawback of LSMA method which requires prior knowledge of material signature[5].

Geometrical algorithm is used for the geometrical selection of endmembers, geometrical algorithms are PPI, SGA, N-FINDR, VCA etc. Geometrical endmember extraction in non linear unmixing is less active, successful approach to use non linear unmixing is Non linear kernel functions. Statistical algorithm is used to process a mixed pixel. Statistical algorithms are AMEE, SSEE method[6]. A hierarchial Bayesian model is a semi-supervised technique. The abundance parameter satisfy the positivity and additivity constraint. The endmembers are normally expressed in Bayesian context by using the abundance prior distribution. The unknown parameters are then derived. That parameters are estimated using Gibbs sampler. The number of endmembers present in the algorithm can be estimated by using Markov Chain Monte Carlo(MCMC) algorithm[7].

Pixel purity Index is a widely used technique for Endmember Extraction. A fast iteration algorithm is used for implementing the pixel purity index which improves the results in many aspects. As a preprocessing step, MNF Transform is used as a dimensionality reduction technique[8]. NFINDR algorithm automatically determines the Endmembers without using any prior knowledge[9]. Spectral Information Divergence(SID) views each pixel spectrum as a random variable and then measures the discrepancy of probabilistic behaviors between two spectra[10]. Endmember Dissimilarity Constraint was proposed on NMF[11].

All the survey papers deals with Spectral unmixing of Synthetic dataset. In this paper an attempt is made with a real time satellite image for spectral unmixing of multispectral Image. Endmembers are extracted using geometrical algorithms like PPI, NFINDR, SMACC and Statistical algorithm like NMF and unmixing is attempted with Linear mixing model. For quantification Root Mean Square Error (RMSE) and Spectral Angle Divergence(SAD) are used.

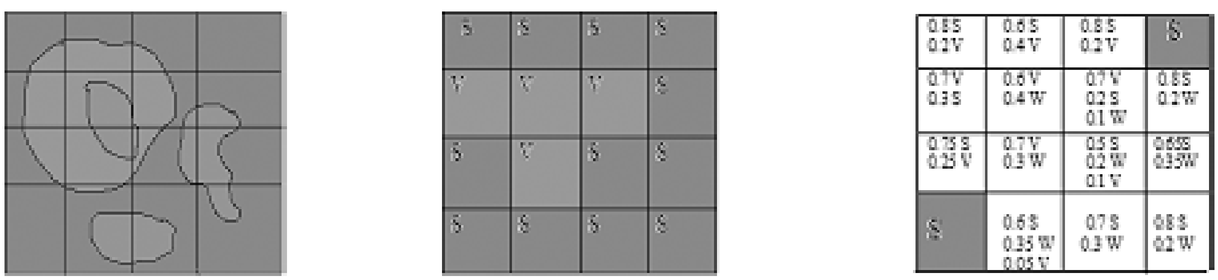
2. METHODOLOGY

The methodology adopted for this study is given in the Fig. 3. Multispectral image is given as input. Endmembers for different classes are extracted using three algorithms namely PPI (Pixel Purity Index), N-FINDR and SMACC(Sequential Maximum Angle Convex Cone). Linear unmixing is attempted on those Endmembers which are extracted from those algorithms.The abundance images or fraction images are produced. These Fraction Images are compared and validated using Root Mean Square Error (RMSE), Reconstruction error (RE) and Spectral Angle Divergence (SAD).

2.1. Pixel purity index (PPI) Algorithm

Pixel Purity Index (PPI) is one of the geometrical algorithm for Endmember Extraction. It selects extremely pure pixels in the image. The pixels are projected onto a vector and the extreme pixels will get a score. The pixels with a highest score is said to be spectrally pure.

Step 1) Principal Component Analysis is used as the preprocessing step for reducing upto *n* bands.



(a) (b) (c)

Figure 2: Toy example for concept of sub-pixel classification(a) Actual ground surface showing soil, Vegetation and Water, (b) Per pixel Classification, (c) Sub pixel classification

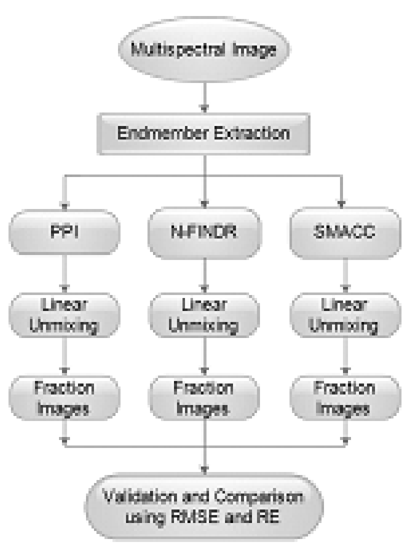


Figure 3: Flowchart depicting the methodology used in this study

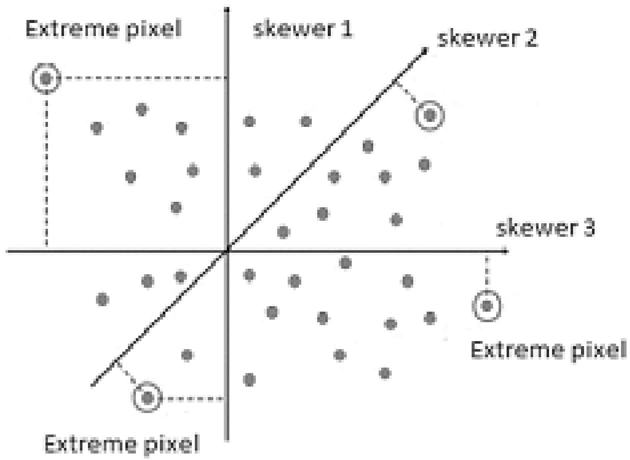


Figure 4: Pixel Purity Index

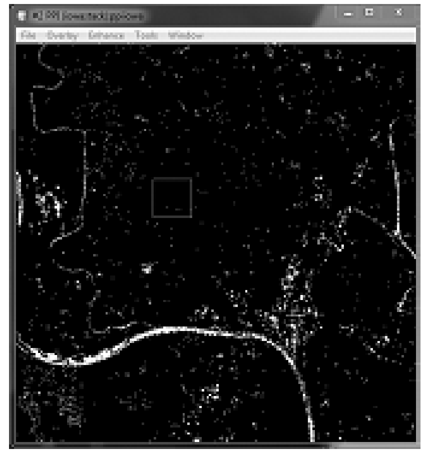


Figure 5: PPI output obtained from ENVI tool

Step 2) Assume a set of m unit vectors called skewers. The data points are projected onto the skewers. The data points that correspond to the extreme values in the direction of the skewers are selected and placed on the list which is denoted by $S_{extrema}$.

Step 3) The above step is repeated with many skewers and the extreme pixels are included in the list. PPI score is assigned to the each extreme sample vector denoted by S_{PPI}.

Step 4) Let t be the threshold set for the PPI score. Extract all the sample vectors greater than the threshold value. The pixel with highest score is selected as the Endmember.

The above algorithm can be implemented using ENVI tool which allows the manual selection of the Final endmembers.

2.2. NFINDR Algorithm

N-FINDR algorithm is the selection algorithm. It is based on the simplex volume Analysis. The potential Endmember is replaced and the pixels with the maximum volume is selected as the next Endmember. It automatically extracts the Endmember based on the maximum volume simplex.

Step 1) Determine the number of the Endmembers using Scatter plot.

Step 2) Reduce the dimensionality of the data from n to l-1 using the dimensionality reduction technique PCA transform.

Step 3) Randomly select one pixel as the first endmember vector $E = \{e_0, e_1, \ldots. e_{l-1}\}$ where l is the number of Endmembers.

Step 4) Determine the Volume of the simplex using the selected Endmember.

$$volume = \frac{\begin{vmatrix} 1 & 1 & \ldots 1 \\ e_0 & e_1 & \ldots e_{l-1} \end{vmatrix}}{(1-l)!} \tag{2}$$

Step 5) Replace the Endmember vector with all the other pixels in the Image and recalculate the volume.

Step 6) All the Endmembers are compared and the endmembers with the highest volume is selected as the new Endmember.

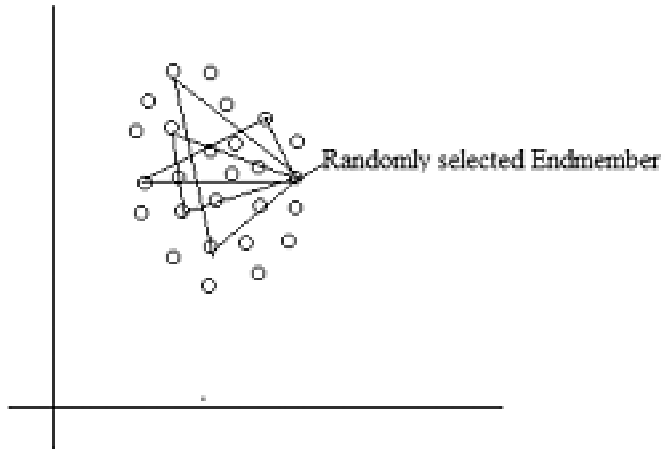


Figure 6: Illustration of NFINDR algorithm

Sequential Maximum Angle Convex Cone(SMACC) Algorithm

SMACC is a sequential algorithm for Endmember Extraction. It is based on the convex cone model. The pixel vectors present inside the cone region is considered as the mixed spectra.

Step 1) In the scatterplot, one extreme pixel is selected as the first endmember.

Step 2) A convex cone is developed by selecting the extreme points along with the existing endmember.

Step 3) An oblique projection is applied to the existing cone. The cone is increased to derive the next endmember.

Step 4) A new Endmember is determined based on the angle it makes with the existing convex cone. The step 3 and 4 are repeated until all the endmembers are found.

It is faster and more automated method for obtaining the pure pixels in the image. Endmember must satisfy two constraints namely 1) Positivity Constraint and 2) Abundance sum to one constraint. The extreme points are connected to determine the convex cone. An oblique projection is applied to the existing cone for deriving the next Endmember. This process is repeated and the cone is increased in size until the specified number of the Endmembers are found.

3. MULTISPECTRAL IMAGE EXPERIMENT

The dataset covers the area of Sioux city in western Iowa which was collected by Landsat 4-5 thematic mapper and it has 7 spectral channels. The size of the original dataset is 512*512.. Multispectral Image has a spatial resolution of 30 meters and swath width of 185 km. The number of endmembers is estimated using Scatterplot. It is determined by the Principal Component Transform (PC 1 vs PC 2). The Scatterplot obtained is triangular in shape, since there are three Endmembers present namely soil, Vegetation and water. For extracting the pure pixels using PPI, the pixel purity index parameters are mentioned, the number of iterations is set to 10,000 and threshold factor is given as random. PPI is extracted and the output is produced as given in Figure 5. The pure pixels are selected using PPI output based on the visual interpretation. The disadvantage of this technique is that there is no method for the computation of threshold factor

In the Preprocessing step of the N-FINDR algorithm, Principal Component Transform is used for reducing the number of PC bands from N to N-1, where N is the number of Endmembers. Here the PC bands are reduced from three to two since there are three endmembers. A random pixel for N-FINDR is selected from the PPI output and then the initial volume of the simplex is calculated. The endmembers are replaced with all the other pixel vectors and the volume of the simplex is calculated. The maximum volume obtained is 302 sq.units. The fraction images of different endmembers are given by

Figure 7: False Color Composite image created using ENVI

$$X_{soil} = E_{soil} * S + e_{soil} \tag{3}$$

$$X_{veg} = E_{veg} * S + e_{veg} \tag{4}$$

$$X_{water} = E_{water} * S + e_{water} \tag{5}$$

Where X_{soil}, X_{veg} and X_{water} are the fraction images of soil, vegetation and water respectively, E_{soil}, E_{veg} and E_{water} are the endmember vectors of soil, vegetation and water respectively and e represents error for the same. The endmember vectors obtained using the different geometrical algorithms are defined in the Linear equation and the fraction images are determined.

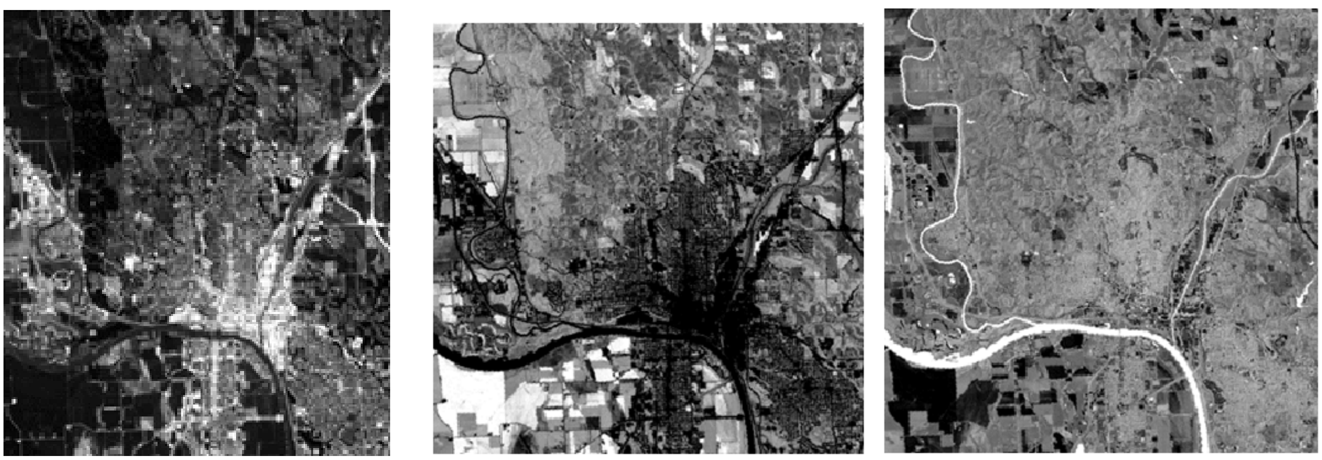
The unmixing results are validated by using three metrics namely RMSE which was defined in[11] by Zhang.et al., and it is defined as

Abundance of soil

Abundance of vegetation

Abundance of water

Abundance of soil **Abundance of vegetation** **Abundance of water**



Abundance of soil **Abundance of vegetation** **Abundance of Water**

Figure 8: (Top) Abundance maps estimated by the PPI algorithm, (Middle) Abundance maps estimated by the NFINDR algorithm, (bottom) Abundance maps estimated by the SMACC algorithm

$$RMSE = \sqrt{\frac{1}{M*N}\sum_{i=1}^{N}\sum_{m=1}^{M}(A_{im} - a_{im})^2} \tag{6}$$

Where A_{im} is the original Multispectral data, a_{im} is the generated Multispectral data with extracted Endmembers, M and N are the number of pixels and number of bands respectively. For evaluating the unmixing accuracy, another criteria called Reconstruction Error is used. It is defined as

$$RE = \frac{1}{L}\sum_{k=1}^{L}(X_{kj} - x_{kj})^2 \tag{7}$$

Where, X_{kj} is the original Multispectral data, x_{kj} is the generated multispectral data with the Extracted Endmembers, L is the number of bands.

Root mean square Error and Reconstruction error values in Table.1 shows the error is very less in SMACC algorithm comparatively.

Spectral Angle distance (SAD) measures the spectral angle between the reference spectrum and the obtained spectrum. SAD is defined as

Table 1
Root Mean Square Error and Reconstruction Error

Methodology	RMSE	RE
PPI	0.4334	0.2210
N-FINDR	0.1286	0.0656
SMACC	0.0574	0.0243

$$SAD = cos^{-1} \frac{a^T.A}{|a^T|.|A|} \tag{8}$$

Where a is an estimated endmember signature and A is the endmember reference signature taken from Spectral libraries provided with Environment for Visualizing Images (ENVI) software.

Table 2
Spectral Angle Distance

Class	PPI	N-FINDR	SMACC
soil	0.3196	0.2971	0.1873
water	0.8863	0.7928	0.2735
Vegetation	0.2039	0.2003	0.1539

There are reference spectrums for various classes in ENVI tool. If the SAD values are higher, then it indicates that the deviation between the reference and the obtained spectrum is more. SAD values have to be lesser for lesser Angle deviation.SAD values in in Table 2 shows that SMACC algorithm has very less angle deviation for all the classes compared to the other two algorithms. Endmembers are extracted using three techniques and fraction images are determined. The soil Fraction images of PPI is not clear, while it is comparatively better in N-FINDR algorithm but the Missouri river is also seen as grey color. Vegetation fraction image is almost same for both PPI and N-FINDR while it is improved in SMACC algorithm.

In water fraction images of both PPI and N-FINDR some vegetation pixels are also classified as water but that error is rectified in SMACC algorithm. The Unmixing results are validated using Root mean square error and Reconstruction Error. The RMSE error is almost 43.34% for PPI, while it is 12.86 % for N-FINDR and a meagre 5.74% for SMACC. Similarly, Reconstruction Error value is 22.1% for PPI, while it is 6.56% for N-FINDR and a meagre 2.43% for SMACC.

4. CONCLUSION

Satellite images are usually used for preparing latest maps. Per pixel classification alone is not sufficient to prepare accurate maps. Hence unmixing of mixed pixels have to be done. Endmember extraction contributes the unmixing accuracy. In this paper, Endmembers are determined using three geometrical algorithms such as Pixel Purity Index (PPI), N-FINDR and SMACC algorithms. This paper has attempted unmixing using Linear mixing model and the fraction images are obtained. The results are validated using Reconstruction error and Root mean square error. The fraction images obtained with the input image derived from SMACC algorithm shows better results comparing to the PPI and NFINDR algorithms.Future work includes the use of Hyperspectral Images and Non-linear unmixing.

References

[1] Keshava. N. and Mustard.J.F., "Spectral Unmixing", IEEE Signal Processing Magazine, Vol. 19, No. 1, pp: 44–57, January 2002.

[2] Li.J. and Bioucas-Dias.J., "Minimum Volume Simplex Analysis: A Fast Algorithm to Unmix Hyperspectral Data", Proceedings of IGARSS, Boston, MA, Vol. 3, pp: 250–253, July 2008.

[3] Bioucas-Dias J. M, Plaza A., Dobigeon N., Parente M., Qian. B., Gader. P and Channus. J., "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse regression-based Approaches", IEEE Journal of Selected Topics in Application of Earth Observation Remote Sensing, Vol. 5, No. 2, pp: 354–379,2012.

[4] Nascimento J. and Bioucas-Dias J., "Vertex Component Analysis: A Fast Algorithm to unmix Hyperspectral Data", IEEE Transactions on Geoscience and Remote Sensing, Vol. 43, No. 4, pp: 898–910, April 2002.

[5] Heinz D.C. and Chein I.C., "Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery", IEEE Transactions on Geoscience and Remote Sensing, Vol. 39, pp: 529–545, 2001.

[6] Parente M. and Plaza A., "Survey of Geometric and Statistical Unmixing Algorithms for Hyperspectral Images", Proceedings of IEEE GRSS WHISPERS, Iceland, 2010.

[7] Dobigeon N., Tourneret J.Y. and Chang C.I., "Semi-Supervised Linear Spectral Unmixing using a Hierarchial Bayesian Model for Hyperspectral Imagery," IEEE Transactions on Signal Processing, Vol. 56, No. 7, July 2008.

[8] Chein I. chang and Antonio plaza, "A Fast Iterative Algorithm for Implementation of Pixel Purity Index", IEEE Geoscience and Remote sensing Letters, vol. 3, Issue 1,2006.

[9] Chein I.C., "Spectral Information Divergence for Hyperspectral Image Analysis", InProceedings of IGARSS '99, vol. 1, pp: 509–511,1999.

[10] Wei Xiong, Chein-I Chang, Fellow, Chao-Cheng Wu, Konstantinos Kalpakis, and Hsian Min Chen, "Fast Algorithms to Implement N-FINDR for Hyperspectral Endmember Extraction", IEEE Journal of selected topics in applied earth observations and remote sensing, Vol. 4, no. 3, September 2011.

[11] Bing Z., Lina Z., Lianru G., Wenfei L., Qiong R. and Qion D., "PSO-EM : A Hyperspectral Unmixing Algorithm based on Normal Compositional Model", IEEE Transactions on Geoscience and Remote sensing, volume 52, Issue 12, 2014..