UNSUPERVISED SOFT CLASSIFICATION OF HYPERSPECTRAL DATA USING CONSTRAINED LEAST SQUARES METHOD

Abolfazl Jamshidzadeh[*,a], Yahya Allahtavakoli[a], Saeid Homayouni[b] and Abdolreza Safari[b]
[a] Graduate student, Dept. of Surveying and Geomatics Eng., College of Engineering, University of Tehran, Tehran, Iran; Tel: +98-21-66493046;
E-mail: {aa_jamshidzadeh, ytavakoli}@ut.ac.ir
[b] Professor, Dept. of Surveying and Geomatics Eng., College of Engineering, University of Tehran, Tehran, Iran; Tel: +98-21-66493046;
E-mail: {homayounis, asafari}@ut.ac.ir

KEY WORDS: Hyperspectral, Soft Classification, Linear Mixture Model, Least Squares.

ABSTRACT: In this paper, we have implemented a three-stage unmixing framework based on linear mixture model (LMM) for sub-pixel land cover classification. Since it doesn't require any a priori knowledge, this method could be considered as an unsupervised approach. In the first step, the number of endmembers will be determined. We have used an eigenvalue thresholding method for determining the number of endmembers. Then, in the two next stages, spectral signatures of endmembers and their corresponding abundance maps are estimated using the Simplex Growing Algorithm (SGA) and Least Squares method, respectively. In order to produce accurate abundance maps of materials using LMM-based estimator, it generally requires that some constraints imposed on least squares method. In this study, we have applied two typical constrained least squares methods (i.e., Fully Constrained and Partially Constrained Least Squares). The focus of our investigation is on the impact of these constraints on the accuracy of soft classification. The proposed algorithms are implemented on a hyperspectral image set. This data is observed by Hyperion of an urban region in Tehran, Iran. For accuracy assessment of soft classification results, a co-registered high spatial resolution land cover map created using QuickBird image of the same area. The overall accuracy and kappa coefficient based on fuzzy confusion matrix have been employed as the measures of accuracy. The evaluations show that P-FCLS method results in more accurate abundance maps, in comparison to FCLS method.

1. INTRODUCTION

Thematic mapping from remotely sensed data is commonly achieved through the application of a conventional hard image classification. In a hard classification, each image pixel is assumed pure and classified to just one class. Often, particularly in coarse spatial resolution images, the pixels may be mixed containing two or more classes. Soft classification allows a pixel to have multiple and partial class memberships and so can accommodate to the effects of mixed pixels. The conventional output of a soft classification is a set of fraction images which indicate the relative coverage of the class in the area represented by the pixel.

From a range of soft classification techniques, we have implemented a three-stage unmixing method based on LMM. These three steps are: number of endmembers determination, endmembers extraction and abundance maps estimation, respectively. In the first step, HFC algorithm has been used. Then, in the second step, SGA algorithm extracts spectral signature of endmembers from the image. At the end, for comparative study we have used two constrained least squares methods to produce abundance maps. One of these methods is FCLS that uses non-negativity and full additivity constraints, simultaneously. The other one that is called P-FCLS, uses non-negativity constraint and partial additivity rather than full additivity constraint (Keshava, 2003). The main aim of this study is to investigate the impact of these constraints on the accuracy of soft classification. The measures of accuracy that is used to evaluate the results are overall

accuracy and kappa coefficient based on fuzzy confusion matrix (Binaghi *et al.*, 1999).
The rest of paper is organized as follows. Section 2 provides a brief overview of the linear mixture model. Section 3 describes the three-stage soft classification approach. Section 4 presents results achieved from experiments. Eventually, Section 5 concludes the paper.

2. LINEAR MIXTURE MODEL

Linear mixture model is a widely used approach for soft classification of remotely sensed hyperspectral imagery. Let $\mathbf{M}$ be an L×P endmember signature matrix denoted by $[\mathbf{m}_1, \mathbf{m}_2 \quad \ldots \quad \mathbf{m}_P]$, where $\mathbf{m}_i$ is an L×1 column vector represented by the signature of the ith material resident in the image scene, and P is the number of materials in the image scene. In linear mixture model, spectral signature of a pixel vector $\mathbf{r} = (r_1, r_2, \ldots, r_L)^T$ can be represented by a linear regression model as follows:

$$\mathbf{r} = \sum_{i=1}^{p} a_i \mathbf{m}_i + \mathbf{n} = \mathbf{Ma} + \mathbf{n} \tag{1}$$

In the equation above, $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$ is a P×1 column abundance vector associated with $\mathbf{r}$, and $\mathbf{n} = (n_1, n_2, \ldots, n_L)^T$ is noise or can be interpreted as a measurement error, where L is the number of spectral band in the image (Heinz and Chang, 2001). A linear unmixing method attempts to unmix the unknown abundance fractions via an inverse of the LMM specified by (1) so as to achieve the tasks of soft classification.

3. SOFT CLASSIFICATION ALGORITHM

From a range of soft classification techniques, we have used a three stage unmixing method. The flowchart of proposed soft classification algorithm is illustrated in Figure 1.

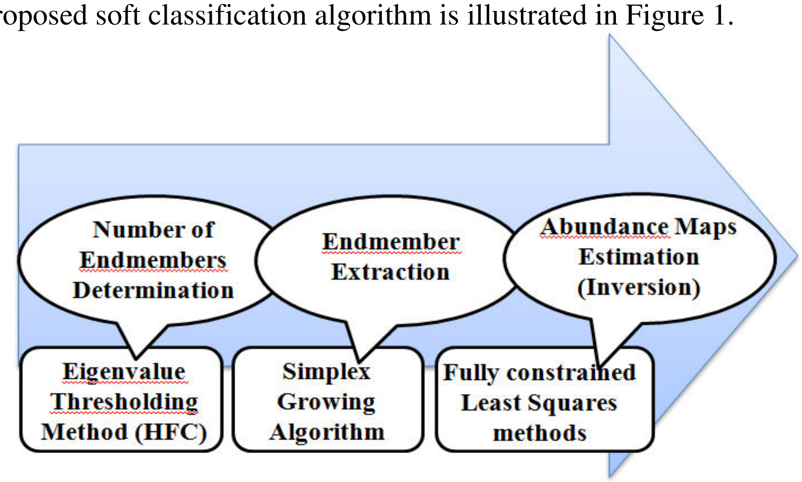


Figure 1. Flowchart for the Soft classification algorithm

As we can see in Figure 1, in the first step, an eigenvalue thresholding method, referred to as the HFC algorithm, is applied to determine the number of endmembers in the image. This method is based on Neyman-Pearson theory-based approach. It doesn't require any a priori knowledge (Chang and Qian, 2004). In the second step, in order to extract the spectral signature of the endmembers, which are the first primary output of unmixing, a geometry-based algorithm is used. This algorithm is called Simplex Growing Algorithm (SGA). It is a sequential algorithm to find a simplex with the maximum volume every time a new vertex is added. The number of simplex

vertices is equal to the number of endmembers (Chang *et al.*, 2006).
The last step is called inversion, which yields the second primary output of unmixing. The objective of these algorithms is to determine the abundance fraction of each endmember in the received pixel spectrum $\mathbf{r}$ or, in terms of the LMM, find the vector $\mathbf{a}$ whose entries weight the columns of $\mathbf{M}$ to yield $\mathbf{r}$. Any meaningful estimate of $\mathbf{a}$, however, must comply with constraints that make it physically realizable. With this in mind, any estimate of $\mathbf{a}$ should obey two constraints for non-negativity and full or partial additivity (Keshava, 2003). According to which of the two additivity constraints is used the fully constrained least squares problem can be stated as follows.
1) Fully constrained least squares problem (FCLS):

$$\hat{\mathbf{a}}_{\mathrm{FCLS}} = \min_{\mathbf{a}}\{(\mathbf{r}-\mathbf{Ma})^{\mathrm{T}}(\mathbf{r}-\mathbf{Ma})\} \quad \text{subject to} \begin{cases} \mathbf{Z}^{\mathrm{T}}\mathbf{a} = 1 \\ \mathbf{a} \geq 0 \end{cases} \tag{2}$$

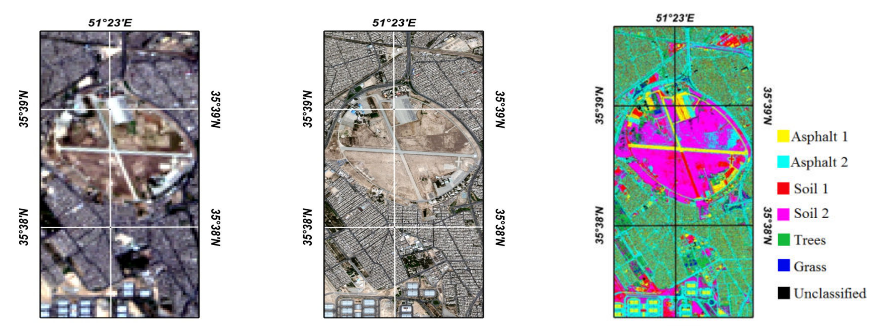
2) Partial- Fully constrained least squares problem (P-FCLS):

$$\hat{\mathbf{a}}_{\mathrm{P-FCLS}} = \min_{\mathbf{a}}\{(\mathbf{r}-\mathbf{Ma})^{\mathrm{T}}(\mathbf{r}-\mathbf{Ma})\} \quad \text{subject to} \begin{cases} \mathbf{Z}^{\mathrm{T}}\mathbf{a} \leq 1 \\ \mathbf{a} \geq 0 \end{cases} \tag{3}$$

In the equations above, $\mathbf{Z}$ is a $\mathrm{P} \times 1$ vector of ones. For comparative study, both the above methods are used in this paper.

4. EXPERIMRNTAL RESULS

In this section a Hyperion image is used to demonstrate the performance of the algorithms. In order to evaluate the estimated abundance maps, a co-registered land cover map achieved from the high spatial resolution QuickBird image is used as a validation data. Figure 2 (a-c) shows co-registered Hyperion image, QuickBird image and the land cover map achieved from QuickBird image, respectively. The number of classes in the land cover map is equal to the number of extracted endmembers from Hyperion image. These classes consist of Asphalt 1 (bright asphalt), Asphalt 2 (dark asphalt), Soil 1 (bright soil), Soil 2 (dark soil), Trees and Grass.

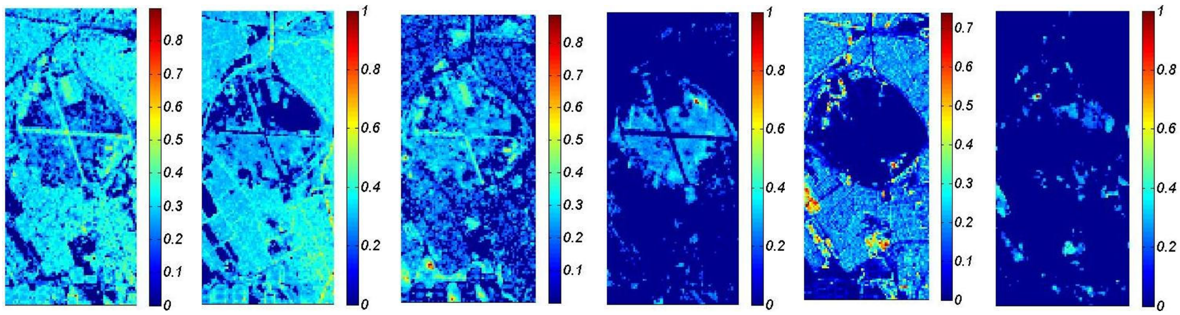


(a) (b) (c)

Figure 2. (a) Hyperion image, (b) QuickBird image, (c) QuickBird land cover map

Spatial resolution of Hyperion and QuickBird images are 30m and 60cm, respectively. So, each pixel in Hyperion image covers 2500 pixels in QuickBird image. A correspondence between the pixels of these two images has to be established. Then, it would be straightforward to determine the percentage of each class in each pixel of Hyperion image from the classified pixels in

Quickbird image. The percentage of high resolution pixels of one class inside one of the low resolution pixels forms the reference abundance for that class. These abundances for each pixel can be joined to compose the ground truth for the Hyperion image. Figure 3 shows these six reference abundance maps that will be used to evaluate estimated abundance maps in soft classification method.



Asphalt 1 Asphalt 2 Soil 1 Soil 2 Trees Grass

Figure 3. Reference abundance maps

According to the section 3, in the first step of soft classification, the number of endmembers is estimated using HFC algorithm. Table 1 tabulates the results of HFC algorithm in the Hyperion image. Based on this table we have adopted six endmembers in the Hyperion image.

False Alarm Probability	0.1	0.01	0.001
Number of Endmembers	7	6	6

Table 1. Estimated number of endmembers

After determining the number of endmembers, in the second step, SGA algorithm is used to extract spectral signature of endmembers. In order to ensure the correct number of endmembers, the algorithms have run with eight endmembers. Spectral angle could be used to match the extracted and reference endmembers, as well as for accuracy assessment of SGA algorithm (Keshava, 2004). But, the cosine of spectral angle is applied to normalize the values between zero and one. Table 2 tabulates the cosine of spectral angle between extracted and reference endmembers.

		Reference Endmembers						MAX
		Asphalt 1	Asphalt 2	Soil 1	Soil 2	Trees	Grass	
Extracted Endmembers	EM 1	1.00	0.99	0.99	0.99	0.98	0.94	1.00
	EM 2	0.98	0.99	0.98	0.99	1.00	0.97	1.00
	EM 3	0.85	0.83	0.87	0.92	0.93	0.98	0.98
	EM 4	0.99	0.98	1.00	0.99	0.98	0.95	1.00
	EM 5	0.98	0.96	0.99	1.00	0.99	0.99	1.00
	EM 6	0.99	1.00	0.98	0.98	0.98	0.94	1.00
	EM 7	0.89	0.88	0.92	0.95	0.96	0.99	0.99
	EM 8	0.92	0.91	0.94	0.97	0.97	1.00	1.00

Table 2. Cosine of spectral angle between extracted and reference endmembers

According to Table 2, extracted endmembers one to eight are matched with Asphalt 1, Trees,

Grass, Soil 1, Soil 2, Asphalt 2, Grass, and Grass, respectively. So, the third, seventh and eighth endmembers are all grass. As a result, the correct number of endmembers is equal to six.
On the other hand, the values in this table for matched endmembers indicate high accuracy of extracted endmembers. Figure 4 depicts spectral signatures of eight extracted endmembers. In this figure, the spectral similarity of the seventh and eighth to the third endmembers is clear.

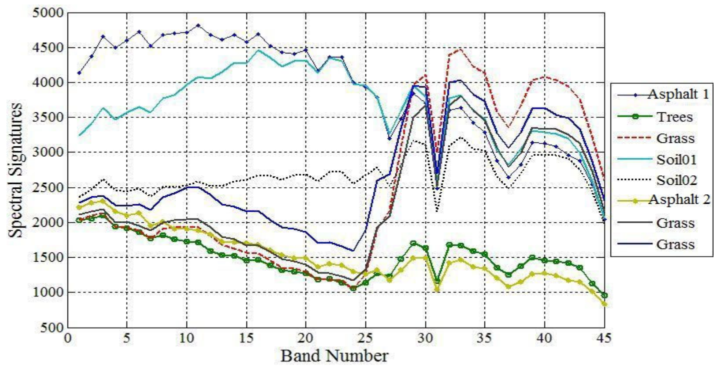


Figure 4. Spectral signatures of eight extracted endmembers

In the last step of soft classification, abundance maps of endmembers have been estimated using FCLS and P-FCLS methods. These abundance maps are shown in Figures Figure 5 and Figure 6. Table 3 shows overall accuracy and kappa coefficient for these abundance maps. This table demonstrates that P-FCLS method performs better than FCLS method.

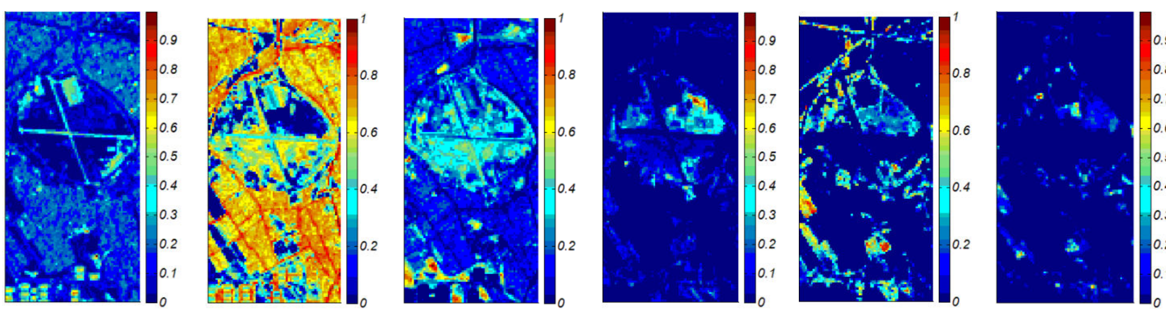

Asphalt 1 Asphalt 2 Soil 1 Soil 2 Trees Grass

Figure 5**.** Estimated abundance maps (FCLS method)

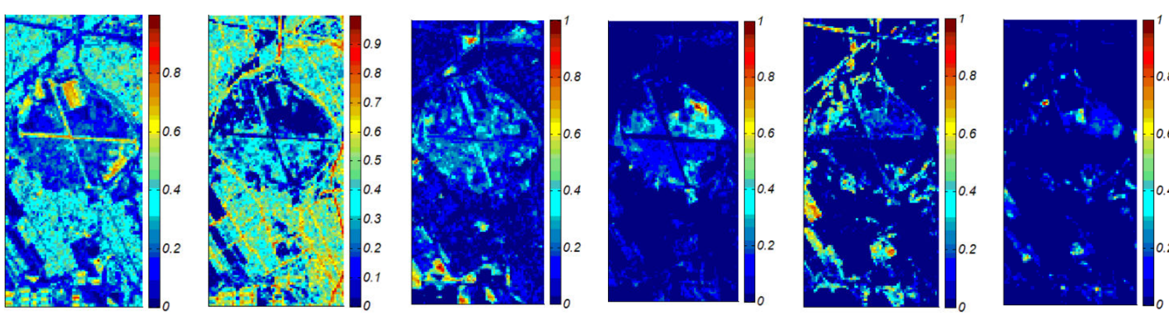

Asphalt 1 Asphalt 2 Soil 1 Soil 2 Trees Grass

Figure 6**.** Estimated abundance maps (P-FCLS method)

It should be noted that estimating signal sources in hyperspectral imagery is a very challenging problem, since many unknown signal sources are usually uncovered on the basis of significantly improved spectral resolution (Chang and Qian, 2004). Due to this fact, the endmember signature matrix **M** usually doesn't comprise all of endmembers existing in the image scene. In this case, sum of abundance fractions in each pixel is less than or equal to one. Accordingly, partial additivity constraint rather than full additivity constraint must be used. Indeed, since full additivity constraint results in overestimation of some abundance fractions, it doesn't work correctly.

		Measure of accuracy	
		Kappa	OA
Method	FCLS	0.63	73.04
	P-FCLS	0.72	78.94

Table 3. Accuracy of estimated abundance maps

5. CONCLUSION

In order to compare the accuracy of abundance maps estimated using two different constrained least squares methods (i.e., FCLS and P-FCLS), we presented a three-stage method based on LMM for soft classification. HFC and SGA algorithms have been used to determine the number of endmembers and to extract the spectral signatures of endmembers, respectively. For abundance maps estimation, we applied FCLS and P-FCLS methods and compared them with each other. We implemented the algorithms on the Hyperion image. A co-registered high resolution land cover map obtained from QuickBird image was applied to evaluate soft classification results. The accuracy of soft classification results was assessed using overall accuracy and kappa coefficient. Experiments have shown that P-FCLS method performs better than FCLS method. This is because the endmember signature matrix **M** doesn't include all of endmembers residing in the image scene. In this situation FCLS method overestimates some of abundance fractions and would not perform well. In contrast, the partial additivity constraint significantly improved the performance of the P-FCLS method. In this study, we focused on a soft classification strategy that exploits spectral information regardless of the spatial information. Integration of spatial and spectral information could improve the accuracy of soft classification. So, developing the algorithms that can use both spectral and spatial information for comparative study with the current results could be seen as a direction of the future work.

6. REFERENCES

Binaghi, E., Brivio, P.A., Ghezzi, P., Rampini, A., 1999. A fuzzy set-based accuracy assessment of soft classification. Pattern Recognition Letters 20, 935-948.

Chang, C.I., Chao-Cheng, W., Wei-min, L., Yen-Chieh, O., 2006. A New Growing Method for Simplex-Based Endmember Extraction Algorithm. IEEE Transactions on Geoscience and Remote Sensing, 44, 2804-2819.

Chang, C.I., Qian, D., 2004. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 42, 608-619.

Heinz, D.C., Chang, C.I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 39, 529-545.

Keshava, N., 2003. A survey of spectral unmixing algorithms. Lincoln Laboratory Journal 14, 55-78.

Keshava, N., 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. Geoscience and Remote Sensing, IEEE Transactions on 42, 1552-1565.