Joint Linear/Nonlinear Spectral Unmixing of Hyperspectral Image Data

Javier Plaza, Antonio Plaza, Rosa Pérez and Pablo Martínez

Department of Technology of Computers and Communications

University of Extremadura, Avda. de la Universidad s/n, E-10071 Cáceres, Spain

E-mail: {jplaza, aplaza, rosapere, pablomar}@unex.es

Abstract—Many available techniques for spectral mixture analysis involve the separation of mixed pixel spectra collected by imaging spectrometers into pure component (endmember) spectra, and the estimation of abundance values for each endmember. Although linear mixing models generally provide a good abstraction of the mixing process, several naturally occurring situations exist where nonlinear models may provide the most accurate assessment of endmember abundance. In this paper, we propose a combined linear/nonlinear mixture model which makes use of linear mixture analysis to provide an initial model estimation, which is then thoroughly refined using a multi-layer neural network coupled with intelligent algorithms for automatic selection of training samples. Three different algorithms for automatic selection of training samples, such as border training algorithm (BTA), mixed signature algorithm (MSA) and mophological erosion algorithm (MEA) are developed for this purpose. The proposed model is evaluated in the context of a real application which involves the use of hyperspectral data sets, collected by the Digital Airborne (DAIS 7915) and Reflective Optics System (ROSIS) imaging spectrometers of DLR, operating simultaneously at multiple spatial resolutions.

I. INTRODUCTION

Most of the pixels collected by hyperspectral imagers contain the resultant mixed spectra from the reflected surface radiation of various subpixel constituent materials. As a result, mixed pixels may exist for several reasons. First, if the spatial resolution of the sensor is not fine enough to separate different pure signature classes at a macroscopic level, these can jointly occupy a single pixel, and the resulting spectral measurement will be a composite of the individual pure spectra [1] (often called *endmembers* in hyperspectral analysis terminology), weighted by a set of scalar endmember abundance fractions.

The linear mixture model assumes that the collected spectra are linearly mixed [2]. For instance, a linear (macroscopic) mixture is obtained when the endmember substances are sitting side-by-side within the field of view of the imager (the linear model assumes minimal secondary reflections and/or multiple scattering effects in the data collection procedure). The resultant mixed spectrum can be expressed as follows:

$$\mathbf{r} = \mathbf{E}\alpha + \mathbf{n} = \sum_{i=1}^{p} \mathbf{e}_i \alpha_i + \mathbf{n}, \qquad (1)$$

where $\mathbf{r}$ is a pixel vector given by a collection of values at different wavelengths, $\mathbf{E} = \{\mathbf{e}_i\}_{i=1}^{p}$ is a matrix containing p endmember signatures, α is a vector containing the fractional abundance values for each of the p endmembers in $\mathbf{r}$, and $\mathbf{n}$ is a noise vector.

Although the linear mixture model has several advantages, such as ease of implementation and flexibility in different applications, there are many naturally occuring situations where nonlinear mixture models may best describe the resultant mixed spectra for certain endmember distributions [3]. In particular, nonlinear mixtures occur in situations where endmember components are randomly distributed throughout the field of view of the instrument [2]. In those cases, the resultant mixed spectra is better described by assuming that part of the source radiation is multiply scattered before being collected by the imager. A general expression for the nonlinear mixture model is given by:

$$\mathbf{r} = f(\mathbf{E}, \alpha) + \mathbf{n}, \qquad (2)$$

where f is an unknown nonlinear function that defines the interaction between $\mathbf{E}$ and α. Various learning-from-data techniques have been proposed in the literature to estimate f. In particular, neural networks have demonstrated great potential to decompose mixed pixels due to their inherent capacity to approximate complex functions [4]. Although many neural network architectures exist, for decomposition of mixed pixels in terms of nonlinear relationships mostly feed-forward networks of various layers, such as the multi-layer perceptron (MLP), have been used [5]. It has been shown in the literature that MLP-based neural models, when trained accordingly, generally outperform other nonlinear models such as regression trees or fuzzy classifiers.

In this paper, we explore in detail the two standard mixture models for abundance estimation in hyperspectral imagery and further propose a combined linear/nonlinear model that results from the advantages and disadvantages of each one. The paper is structured as follows. Section II identifies some important aspects of neural network-based spectral mixture analysis. Section III describes the combined linear/nonlinear model. Section IV conducts real data experiments using hyperspectral data sets collected by the DAIS 7915 and the ROSIS imaging spectrometers, operating at multiple resolutions in the context of a real agriculture and farming application. Finally, Section V summarizes the main contributions of this work.

II. Issues for Mixed Pixel Interpretation Using Artificial Neural Networks

A variety of issues have been investigated in order to evaluate the impact of training and initialization on mixed pixel interpretation using artificial neural networks. Two aspects will be particularly addressed in this work:

1) *Training issues*. Most of the attention when analyzing the impact of training on the interpretation of mixed pixels has been focused on the size of the training set. However, even if the endmembers participating in mixtures in a certain area are known, proportions of these endmembers on a per-pixel basis are difficult to be estimated *a priori*. Therefore, a challenging aspect in the design of neural network-based techniques for spectral mixture analysis is to reduce the need for very large training sets.

2) *Initialization issues*. A second important issue has to do with initial model conditions. For instance, the MLP neural network is typically trained using the error back-propagation algorithm, a supervised technique of training with three phases. In the first one, an initial vector is presented to the network, which leads to the activation of the network as a whole. The second phase computes an error between the output vector and a vector of desired values for each output unit, and propagates it successively back through the network. The last phase computes the changes for the connection weights, which are randomly generated in the beginning. According to algorithm design, an effective learning algorithm should not depend on initial conditions, which can only affect the convergence rate but should not alter the final results. Often, this is not the case of learning algorithms used for neural networks. In order for a mixture model to be effective, initial values must be representative and cannot be arbitrary.

In this work, we explore and further propose solutions to resolve the issues addressed above. To address the first issue, we develop intelligent training sample selection algorithms which can greatly minimize the number of required training samples. To address the second issue, we develop a joint linear/nonlinear mixture model in which linear estimations are used as the initial condition for a nonlinear neural network-based mixture model.

III. Combined Linear/Nonlinear Mixture Model

In this section, we describe a combined linear/nonlinear model which consists of the following main steps:

1) Initialization via a fully constrained linear mixture model based on automatic endmember extraction; and
2) nonlinear refinement using a MLP neural network. This step is supported by a pool of unsupervised algorithms for intelligent selection of training samples.

Fig. 1 shows a schematic block diagram of the proposed method, which will be further developed in this section.

A. Initialization Using a Linear Mixture Model

Linear SMA assumes that the pure (endmember) signatures of materials in the image scene are known. Therefore, estimation of the number of endmembers, p, is a key issue. Here, we use the concept of virtual dimensionality (VD) can be used to accurately estimate the number of endmembers in the scene [6]. Once the number of endmembers is estimated, we use the automated morphological endmember extraction (AMEE) algorithm [7], one of the few available endmember extraction algorithms which combine spatial and spectral information, to extract a set of p spectral endmembers from the input data. Once a final set of endmembers $\{\mathbf{e}_i\}_{i=1}^P$ has been found, we use fully constrained linear spectral unmixing (FCLSU) [6] to produce a vector of abundance fractions $\alpha^{(\mathbf{r})} = \{\alpha_i^{(\mathbf{r})}\}_{i=1}^p$ for each pixel vector $\mathbf{r}$, subject to $\alpha_i \geq 0$ for all $1 \leq i \leq p$, and $\sum_{i=1}^p \alpha_i^{(\mathbf{r})} = 1$. The fully constrained linear estimation above is used in this work as an initial condition for a subsequent nonlinear refinement stage, which is described below.

B. Nonlinear Refinement Using a Multi-layer Neural Network

The neural architecture used for nonlinear refinement is composed of three layers [8]. The neuron count at the input and output layers, p, equals the number of spectral constituents found by the AMEE algorithm in the initialization stage. The input patterns to the input layer are vectors of endmember fractional abundances $\alpha^{(\mathbf{r})} = \{\alpha_i^{(\mathbf{r})}\}_{i=1}^p$ for each sample vector $\mathbf{r}$, first estimated by FCLSU. The number of neurons in the hidden layer has been empirically set to twice the number of endmembers found in the initialization stage. Information flows from the input layer to the hidden layer, and then to the output layer via a set of network connections. Each connection multiplies the value coming from its origin node by the weight assigned to that arc and sends the result to the destination node, which adds the values presented to it by all the incoming connection, transforms it with a nonlinear activation function (the sigmoid function in this work [9]), and then sends the result along all of its outgoing connections.

If we denote by $\alpha^{(\mathbf{t}_l)} = \{\alpha_i^{(\mathbf{t}_l)}\}_{l=1}^t$ the vector of fractional abundances estimated by linear mixture model for the l-th training sample $\mathbf{t}_l$ used in the backpropagation algorithm (where t is the total number of training patterns), and by $\hat{\alpha}^{(\mathbf{t}_l)} = \{\hat{\alpha}_i^{(\mathbf{t}_l)}\}_{l=1}^t$ a known vector of contributions (ground-truth) for that training pattern, then the backpropagation algorithm computes the difference between the true function value and the prediction and propagates it successively back through the network so that the matrix of connection weights $\mathbf{W}$ is adjusted until the network approximates the desired output closely enough. A measure of how much the network is deviating from the desired performance can be expressed in terms of the root mean square error (RMSE) as follows:

$$\text{RMSE}(\mathbf{W}) = \sqrt{(1/t) \sum_{l=1}^{t} \sum_{i=1}^{p} (\alpha_i^{(\mathbf{t}_l)} - \hat{\alpha}_i^{(\mathbf{t}_l)})^2} \qquad (3)$$
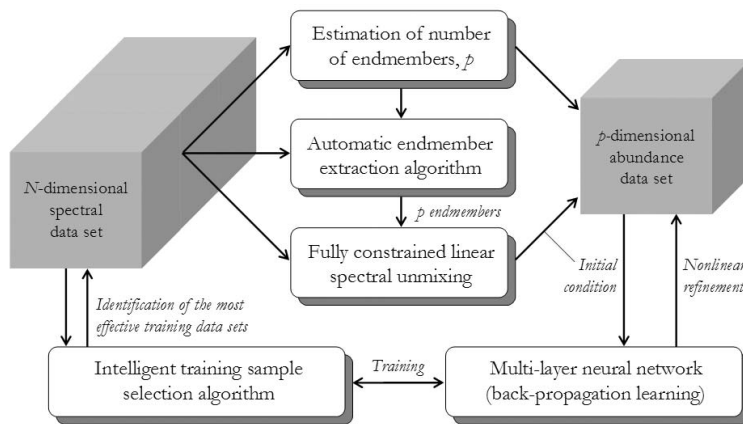
Fig. 1. Block diagram of the proposed linear/nonlinear method for spectral mixture analysis.

C. Algorithms for Intelligent Selection of Training Samples

To conclude this section, we outline several automatic algorithms which have been specifically developed to automatically search for the most representative training samples from the data set according to different criteria, such as the *borderness* (convexity) of those samples or the degree of spectral similarity to other spatially adjacent training samples.

1) Border-Training Sample Selection Algorithm (BTA): The separation of a training set into border and non-border patterns in the context of a pure pixel classification problem was first explored by Foody [10], who expressed borderness as the difference between the two smallest distances measured for each training pattern. A border-training pattern is expected to be almost as close to its actual class of membership as it is to any other class. Therefore, the difference in the Mahalanobis distances between the two most likely classes of membership would be small for a border pattern. Using this concept, we have developed a border-training sample selection algorithm (BTA) which consists of a two-stage process, in which a set of pure training samples are first automatically extracted from the input data (using the AMEE algorithm), and then a degree of borderness related to those samples is used to identify highly mixed training samples.

2) Mixed-Signature Selection Algorithm (MSA): As an alternative to the BTA algorithm, we have developed a mixed-signature selection algorithm (MSA) that iteratively seeks for the most highly mixed training samples first. This is done by first calculating the centroid of the data cloud and then computing an eccentricity score for each pixel in the input scene using the spectral angle as the baseline distance. The pixels with lowest eccentricity score are the most suitable candidates for being selected by MSA to be used as training samples. It should be noted that the algorithm above is designed to search for the most highly mixed signatures first. The concept implemented by this algorithm can be viewed as the opposite to that used by convex geometry-based endmember extraction methods.

3) Morphological erosion algorithm (MEA): This algorithm makes use of an extended morphological erosion op-

eration [7], which can be very useful for the interpretation of mixed pixels since it takes into account both the spatial and the spectral properties of the image data in simultaneous fashion. The idea is to define a spatial search area around each pixel vector, typically, a 5×5 to 15×15-pixel neighborhood, and then compute the spectral angle between the pixel and the most highly mixed pixel in the neighborhood. The resulting scores are accumulated and used as a measure of the eccentricity of the pixel (as in the case of MSA, the pixels with lowest eccentricity score are the most suitable candidates for being selected by MEA to be used as training samples).

IV. EXPERIMENTAL RESULTS

The data used in this study was collected over a so-called *Dehesa* test site in Cáceres, SW Spain. Dehesa ecosystems are formed by *quercus ilex* (cork-oak trees), soil and pasture. Their exploitation has an important impact on the economies of several European countries (most notably, Spain and Portugal). The image data is formed by a ROSIS scene collected at high spatial resolution, with 1.2-meter pixels, and its corresponding DAIS 7915 scene, collected at low spatial resolution with 6-meter pixels. Several field techniques were applied to obtain reliable estimates of the true fractional land cover for each DAIS pixel in the considered Dehesa test site. First, the ROSIS image was roughly classified into the three land-cover components above using a maximum-likelihood supervised classification approach based on image-derived spectral endmembers. Second, the classified ROSIS image was registered with the DAIS image using an automated ground control point-based method with sub-pixel accuracy. Most importantly, the abundance maps at the ROSIS level described above were thoroughly refined using field data before obtaining the final reference proportions.

Fig. 2(a-c) shows the training areas extracted from the DAIS scene by the three developed training sample selection algorithms (BTA, MSA and MEA, respectively). In order to assess the performance of the considered mixed sample selection algorithm, two unsupervised algorithms were also used in experiments. The first one [see Fig. 2(d)] is an
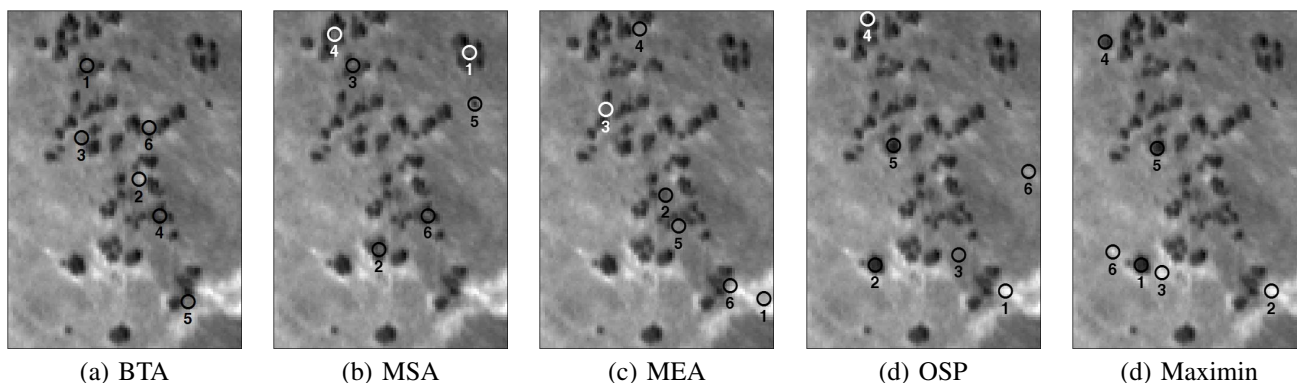
Fig. 2. Training samples extracted by five different training sample selection algorithms from the DAIS 7915 hyperspectral scene.

# Training samples	BTA	MSA	MEA	OSP	Maximin
1	0.146	0.116	0.125	0.147	0.150
2	0.121	0.087	0.096	0.143	0.143
3	0.093	0.040	0.046	0.140	0.142
4	0.093	0.039	0.044	0.136	0.139
5	0.093	0.040	0.043	0.131	0.137
6	0.090	0.041	0.044	0.124	0.127

automated target generation process based on an orthogonal subspace projection (OSP) approach [6]. The second one [see Fig. 2(e)] is the Maximin algorithm commonly used in pattern recognition applications [9]. In all cases, the number of training samples was limited to six on purpose.

Table I quantitatively compares the performance of the proposed linear/nonlinear model trained with different algorithms and number of training samples, where the number of endmembers in the considered DAIS scene was estimated to $p = 10$. Specifically, the table reports the RMSE scores in fractional abundance estimation (with regards to measured ground-truth) of *pasture*. The RMSE score produced by the linear mixture model for this constituent was 0.153.

From results in Table I, it is clear that using only three training samples generated by the BSA, MSA and MEA erosion algorithms always introduced a significant improvement in abundance estimation with regards to the cases where less training samples were considered. It was also apparent that using additional samples selected by these three algorithms did not significantly improve the quality of abundance estimation. Quite opposite, we observed that the OSP and Maximin algorithms produced rather unstable results, with only moderately acceptable scores when all six training samples were used. These results illustrate the advantages of intelligent initialization and training for the proposed linear/nonlinear model. In particular, intelligent generation of training samples appears to play a very significant role, thus showing the potential to direct training data collection to target the most useful sites.

V. SUMMARY

Linear mixture models and artificial neural networks can be viewed as two ends of a spectrum of mixture models in remote sensing. On the one hand, linear models are simple to implement and can provide a rough estimation of the abundance of endmember classes, but generally result in poor accuracy when more complex mixtures are involved, mainly because those models cannot capture the variation in the data well. On the other hand, neural network models are highly nonlinear, and hence are able to capture complex structure in the data better. Their use is gaining popularity, but the lack of commonly accepted initialization and, particularly, training procedures represents a major obstacle. In this paper, we have developed a first attempt to bridge the gap between the two models, with the purpose of exploiting their main advantages in combined fashion.

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 1 site," *Journal of Geophysical Research*, vol. 91, pp. 8098-8112, 1986.
[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44-57, 2002.
[3] W. Liu and E. Y. Wu, "Comparison of non-linear mixture models: sub-pixel classification," *Remote Sensing of Environment*, vol. 18, pp. 1976-2003, 2004.
[4] G. G. Wilkinson, "Open questions in neurocomputing for Earth observation," in: I. Kanellopoulos, G. G. Wilkinson, F. Roli and J. Austin (Eds.), *Neurocomputation in remote sensing data analysis* (pp. 3-13), Berlin: Springer, 1999.
[5] A. Baraldi, E. Binaghi, P. Blonda, P. A. Brivio and A. Rampini, "Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 994-1005, 2001.
[6] C.-I Chang, *Hyperspectral imaging: Techniques for spectral detection and classification*. Kluwer Academic Publishers: NY, 2003.
[7] A. Plaza, P. Martnez, R. Prez and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, pp. 2025-2041, 2002.
[8] J. Plaza, P. Martinez, A. Plaza and R.M. Perez, Nonlinear Neural Network-based Mixture Model for Estimating the Concentration of Nitrogen Salts in Turbid Inland Waters Using Hyperspectral Imagery. *Proceedings of SPIE, Chemical and Biological Standoff Detection II*, vol. 5584, pp. 165-173, 2004.
[9] C. M. Bishop, *Neural networks for pattern recognition*, Oxford: Oxford University Press, 1995.
[10] G. M. Foody, "The significance of border training patterns in classification by a feedforward neural network using backpropagation learning," *International Journal of Remote Sensing*, vol. 20, pp. 3549-3562, 1999.