Pixel Unmixing Using Positive Matrix Factorization

Samuel Rosario Torres
Advisor: Miguel Vélez Reyes

Electrical and Computer Engineering Department
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681-5000
{samuel.rosario, mvelez}@ece.uprm.edu, samuel_rt@yahoo.com

Abstract

In *hyperspectral imagery* (HSI), hundreds of images are taken at narrow and contiguous spectral band space providing us with high spectral resolution that can be used to discriminate between objects [Vélez02, Vélez00]. HSI sensors with high spectral resolutions and relative low spatial resolution inherit the problem of mixing pixel because the pixel size is relative big consequently many spectral signatures of near objects could be included in the image scanning process [Keshava02]. Is an interesting and practical problem is *pixel unmixing*, separate the pixel in the corresponding spectral signature, *endmember* and a sets of corresponding fractions, *abundances*. Positive Matrix Factorization (PMF) also know as Non-Negative Matrix Factorization (NNMF) are technique to factorizes a matrix A into matrixes W, H such that A ≅ WH. In this work, we use the PMF to decompose the mixed pixel in the constituent spectra (*endmember*) and a set of corresponding fractions (*abundances*). In addition, we present preliminary results of two PMF iterative algorithms for unmixing problem, based in Euclidean distance and Divergence algorithms.

1. Introduction

Remote sensing imagery used in detection and target recognition in many environmental applications such for vegetation stress, minerals, etc. [Vélez00]. HSI sensors provide high spectral resolution order of hundreds of bands but with relative low spatial resolution. Mixed pixels are consequence of HSI sensor with low spatial resolution the pixel sizes are bigger and consequently pixels in the image are mixed with near objects also we could have mixed pixels as results of different materials combined in a homogeneous mixture [keshava02]. An interesting problem and practical problem is decomposing radiance or reflectance of the pixels in a HSI and separate into the spectral signatures that contribute to the pixels, the *unmixing problem. Spectral unmixing* is know as the procedure of decompose the measure spectrum of mixed pixels into a set of corresponding spectra, *endmember*, and a set of corresponding fractions, *abundances* [keshava02], [Vélez02]. In the literature we see different approach to solve the unmixing problem but most of them assume the Linear Mixing Model (LMM) (see equation 1) for their development of the algorithms [keshava02], [Vélez02], [Plaza02]. The representation of the LMM is shown in the equation (1):

$$\mathrm{b} = \sum_{i=1}^{n} x_i \bar{\mathrm{a}}_i + \mathrm{w} = \mathrm{Ax} + \mathrm{w} \qquad \textbf{(1)}$$

where $\mathbf{A} \in \Re_{+}^{m\times n}$ is the matrix of the endmember where $\overline{a}_i$ is the spectral signature of the i-th endmember; $\mathbf{x} \in \Re_{+}^{n}$ is the vector of the abundances; n is the number of endmember and $\mathbf{w}$ is the noise vector [Boardman94], [keshava02], [Vélez02]. The entries for all the variables has to be positive in order to have physical meaning implying that x_i, a_i, $w_i > 0\ \forall i$. In the LMM assume that the incident light interact with the surface with only one endmember, there is no multiple scattering between endmembers, the total surface area is a linear combination of the abundances of the endmember [Keshava02], [Velez02].

Positive Matrix Factorization (PMF) or Non-negative Matrix Factorization (NMF) is used for the decomposition of multivariable data into smaller sets of the original [Lee99], [Lee00]. The problem of PMF is having a matrix V, find the matrix factors W, H such that V ≅ WH. The matrix V will be represented as a linear combination of W, weighted by the elements of H; this has some similarities to the LMM assumptions.

In this work we used to iterative PMF algorithms described by [Lee02] to find the endmember in a HSI data, Euclidean Distance and Divergence. We create synthetic data in order to observe the performance of the algorithms. In addition, we present some results of the PMF algorithms findings endmembers and the convergence with different number of endmembers.

2. Unmixing Algorithms

The unmixing algorithms can be separated into two areas supervised or unsupervised methods. Supervised methods require of a trained analyst in contrast with the unsupervised that is highly automated. In addition; some methods require dimensional reduction for better performance of the algorithm [Plaza02]. Into the unsupervised methods the most common type algorithms for the unmixing problem are based that the endmember are know for the estimation of the abundances, while other estimate first the endmember and then the abundances and others estimate both quantities simultaneously [Keshava02]. In addition; we can find other algorithms that first estimate the endmembers and then the abundances. Also we can find other methods like morphological operations [Plaza02].

Algorithms that estimate the abundances are the ones developed for the emission topography, EMML and the Image Space Reconstruction Algorithm (ISRA) [Vélez02]. An algorithm that performs endmember determination and pixel purity is the Automated Morphological Endmember Extraction (AMEE) introduced in [Plaza02]. The algorithms mentioned above are unsupervised methods. Some supervised methods are Pixel Purity Index (integrated in the ENVI Software), Manual Endmember Selection Tool (MEST), Multiple Endmember Spectral Mixture Analysis (MESMA) are describe in [Plaza02]. The algorithms of PPI and MEST use band reduction algorithms in order to improve the faster the results, in PPI the band reduction algorithms is Minimum Noise Fraction (MNF) and for MEST is Principal Component Analysis (PCA).

In this work we plan to use Positive Matrix Factorization (PMF) ([Lee00]) knowing that the HSI data is positive (in $\Re_+^{m \times n}$) to find endmember, this algorithms has been used for image reconstruction [Lee99]. The PMF algorithms factorize V into two positive sub matrices W and H such that V ≅ WH. In [Lee00] explain that W is a linear combination weighted by H to approximate V, similar to the equation 1. The two iterative algorithms are based in minimizing a cost function to verify the approximation of V ≅ WH. The costs function for the Euclidean Distance:

$$\|A - B\|^2 = \sum_{i,j} (A_{ij} - B_{ij})^2 \qquad \textbf{(2)}$$

The costs function for the Divergence:

$$D(A \| B) = \sum_{i,j} (A_{ij} \log(\frac{A_{ij}}{B_{ij}}) - A_{ij} + B_{ij}) \quad \textbf{(3)}$$

The iterative rules for the Euclidean Distance are:

$$H_{au} \leftarrow H_{au} \frac{(W^T V)_{au}}{(W^T WH)_{au}}, \qquad \textbf{(4)}$$

$$W_{ia} \leftarrow H_{ia} \frac{(V^T H)_{ia}}{(W^T WH)_{ia}}.$$

The rules for the iterative algorithm for the Divergence are given by:

$$H_{au} \leftarrow H_{au} \frac{\sum_i W_{ia} V_{iu} / (WH)_{iu}}{\sum_k W_{ka}}, \quad \textbf{(5)}$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum_u H_{au} V_{iu} / (WH)_{iu}}{\sum_v H_{av}}$$

In the Section 4 (Simulation Results) we show the results of the simulation experiments with the PMF algorithms.

3. Simulation Experiment

In this work we create synthetic data with know endmembers. We use the spectral signature of different types of leafs: pansy, dock, grass, clover, trefoil and dandelion. The Figure 1 is the graph of the spectral signature of the selected leafs (measured in reflectance). Also we can observe from de Figure 1 is that the endmember of the leaf in many wavelength are very similar. The endmember leafs where used for the W and generate random number to generate the H, then we generate V = WH, as the synthetic HSI data. The purpose of these is to observe the PMF algorithms can find the endmember in the data, knowing the real endmember of the data and later we can contrast the result of the algorithms. In future experiments we will add noise (**w** of the equation 1) to the synthetic data to simulate similar condition when a HSI sensor is taking an image.

4. Simulation Results

The results of the PMF Euclidean Distance (PMFED) and Divergence (PMFD) algorithms finding six endmembers are show in the Figure 2 and Figure 3 respectively. Observing the Figure 1,2,3 we can observe that there are some resemblances of the real endmember and the endmember founds, for example the grass spectral signature is similar to the endmember 1 from the Figure 1 and Figure 2, in the Figure 4 show the normalization (with the maximum value of each one) of the grass endmember, endmember found with the PMFED and PMFD. In addition, we can see others endmembers like dock that has similar endmember founds for the PMF algorithms, the endmember 3 in the Figure 3 and Figure 4.

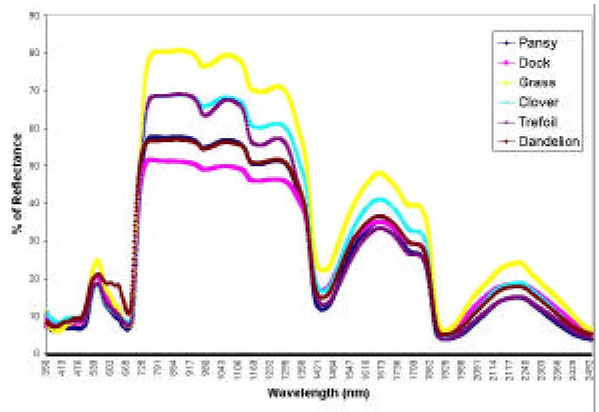

Figure 1: Endmember of leafs

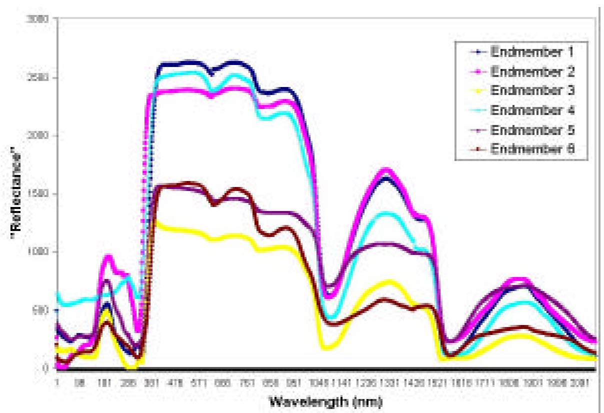

Figure 2: Endmembers obtained with PMFED

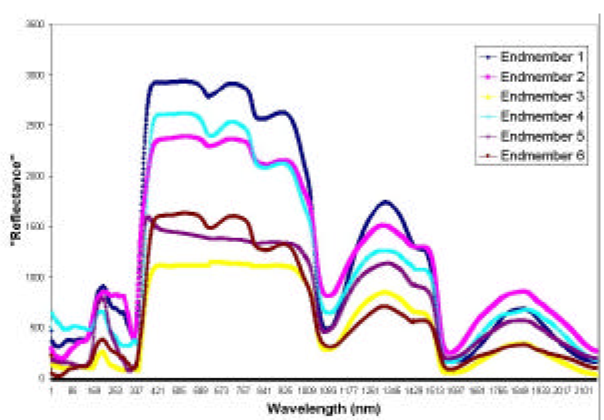

Figure 3: Endmember obtained with PMFD

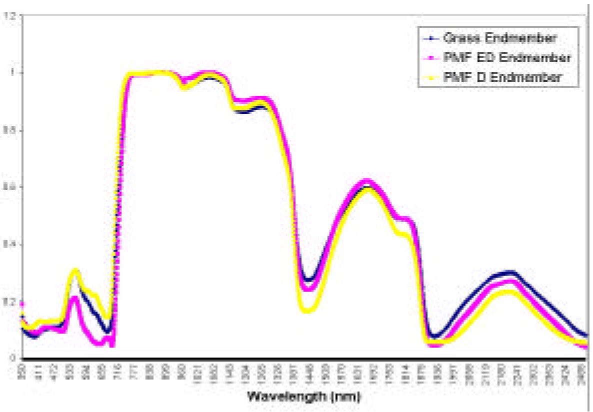

Figure 4: Normalization of the grass endmember and founded endmember with PMFD PMFDE

5. Conclusions

In this work we have presented Positive Matrix Factorization algorithms to solve unmixing problem. We use the iterative methods introduced by [Lee02] based in the Euclidean Distance and Divergence. We generate synthetic data with different types of leafs to verify the consistency of the PMF algorithms finding the endmembers. The results given by the PMFED and the PMFD are relative close with the real endmember. Further analysis has to be done to verify the consistency of the algorithms when we use synthetic data with noise and real data.

References

[Boardman94] Boardman, J.W. "Geometric mixture analysis of imaging spectrometry data". Geoscience and Remote Sensing Symposium. Volume 4, pages 2369 -2371, 1994.

[Keshava02] Keshava N., Mustard J. F. "Spectral Unmixing", IEEE Signal Processing Magazine, volume 19, issue 1, January 2002, pages 44 -57.

[Lee00] Lee D., Seung. H. "Algorithms for non-negative matrix factorization". NIPS, 2000.

[Lee99] Lee D., Seung. H. S. "Learning the parts of objects by non-negative matrix factorization". Nature, volume 401, October 1999, pages 788-791.

[Plaza02] Plaza, A., Martinez, P., Perez, R., Plaza, J. "Spatial/spectral endmember extraction by multidimensional morphological operations". IEEE Transactions on Geosciences and Remote Sensing, volume 40, issue 9, September 2002, pages: 2025 -2041.

[Vélez02] Vélez-Reyes M. "Iterative Algorithms for Unmixing of Hyperspectral Imagery". Air Force Faculty Fellowship Program Final Report, summer 2002.

[Vélez00] Vélez-Reyes M., Jiménez L.O., Linares D.M., and Velázquez H.T. "Comparison of matrix factorization algorithms for band selection in hyperspectral imagery." In *Algorithms for Multispectral, Hyperspectral and Ultraspectral Imagery VI*, Proceedings of SPIE, Vol. 4049, Orlando Florida, 2000.