NEW APPROACHES FOR FEATURE EXTRACTION IN HYPERSPECTRAL IMAGERY

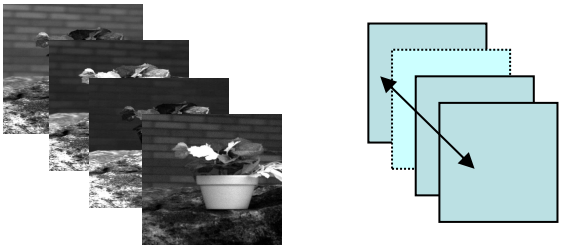
Stefan A. Robila, Lukasz Maciack
(Center for Imaging and Optics, Department of Computer Science, Montclair State University, Montclair, NJ 07043, robilas@mail.montclair.edu, mackiackl1@mail.montclair.edu)

ABSTRACT

In this paper we introduce a novel feature extraction method based on Nonnegative Matrix Factorization (NMF) for hyperspectral image processing. Given the large size of the hyperspectral imagery, feature extraction plays an important role in producing fast and accurate results. Traditional approaches such as Principal Component Analysis and Independent Component Analysis generate the features as a linear combination of the hyperspectral bands emphasizing on the decorrelation or independence of the features. Compared to this, NMF offers a decomposition solution that is less restrictive requiring only the positivity of the features and the associated linear transform. Such scenario has a natural meaning in hyperspectral imagery where each pixel observation is thought to be formed as a linear positive mixture of reflectance values of the materials in the scene (endmembers) covered by the pixel. With hyperspectral imagery spatial resolution ranging from millimeters to kilometers, it likely that the data observed are formed as a mixture. In this case, the linear transform used to generate the features would be associated to the endmembers and the resulting features would be associated to the abundance of each endmember in the pixels.
We present our results on using NMF for feature extraction by performing experiments with hyperspectral digital imagery collection experiment (HYDICE) data as well as in-house imagery collected with a SOC 700 hyperspectral camera. The experiments suggest that NMF outperforms PCA in feature and endmember extraction.

1. INTRODUCTION

Remote sensing is generally described as the measurement, from a distance, of spectral features of the Earth's surface and atmosphere [1] and has developed in parallel with airspace technology. Starting with the first successful flights, scientists have realized the advantage of aerial over ground observations in terms of cost, accuracy and ability to observe inaccessible areas. Applications in agriculture, forestry, mining, military surveillance and resource and emergency management were soon developed further supporting the rapid development of remote sensing techniques [2]. Originally based on humans, aerial reconnaissance soon shifted to photography and video recordings providing considerable amounts of information to be analyzed.
Following these beginnings, advances in optics and imaging have revealed that different materials expose different properties when analyzed under various light wavelength intervals [3]. A classic example is vegetation that provides higher reflectance values for near infrared wavelengths compared with certain visible wavelengths. In contrast, bare soil or man made materials have relatively the same reflectance for both of these wavelength ranges. Using a simple formula (NDVI – Normalized Digital Vegetation Index) [4] that combines the material reflectance in near-infrared and imaging sensors capable of recording data in both ranges, one can easily build a general detection tool for vegetation.
Hyperspectral imagery takes this idea further. The data are collected as hundreds of images (*spectral images* or *spectral bands*), with each image corresponding to narrow contiguous wavelength intervals (Fig. 1a). Hyperspectral sensors cover wavelengths from the visible range (0.4μm-0.7μm) to the middle infrared range (2.4μm) [5].
In hyperspectral data, it is a common practice to define *pixel vectors* (or *spectra*) as the vectors formed of pixel intensities from the same location, across the bands (Fig. 1b) [6]. Each pixel corresponds to a certain region of the scene surveyed and will represent the spectral information for that region.



a) b)

Figure 1 – a) Examples of spectral bands for visible (green, blue and red) and near infrared wavelength intervals b) Formation of pixel vectors or spectra.

Due to the narrow bandwidth and the abundance of observations, the pixel vector for each pixel location resembles a continuous function of wavelengths. This function describes the reflectance of the material for wavelengths within the frequency interval covered by the sensor. Figure 2 shows the spectra for a leaf and a rock collected using a hyperspectral camera over 120 bands ranging from 0.4μm to 0.7μm. Note the significant difference in reflectance values for the near-infrared bands.

Due to its richness in information, hyperspectral imagery allows for detection of targets covering areas smaller than a pixel or separation of objects and shapes otherwise undistinguishable in regular images. Most of the hyperspectral image processing techniques have complexity that depends directly on the number of spectral bands in the acquired data. Since this is usually large, it is of interest to find methods that transform the data cube into one with reduced dimensionality while, at the same time, maintaining as much information content as possible. These techniques are known under the general name of feature extraction [7].

In this paper we provide a new direction for feature extraction in hyperspectral data by relaxing the conditions imposed on features by traditional techniques. Our paper is organized as follows. In the next section we provide a brief introduction to feature extraction and discuss the techniques available for unsupervised feature extraction. In Section 3 we present a new approach for feature extraction based on Nonnegative Matrix Factorization. In Section 4 we discuss our experimental results followed by Future Work (Section 5), Conclusions and References.

2. FEATURE EXTRACTION

Feature extraction is generally defined as the process of reducing the data to a lower dimension without significant information loss. In hyperspectral imagery, this is done by either selecting certain bands of by using a transform that produces the features as combinations of bands (Fig. 3). The algorithms focus on the increase of the separation between classes within each feature. The separation is measured using class information such as distance between means, distance between probabilities, etc. Supervised extraction is employed when prior class information is available (such as training pixel vectors), and unsupervised extraction is used when a priori information is lacking. In the unsupervised case the statistics and distance measures between classes cannot be computed or estimated. Instead, the main goal becomes the reduction of the data redundancy. The narrow bandwidth associated with the spectral bands leads to correlation between the adjacent bands yielding a relatively high level of redundancy in the data [7].

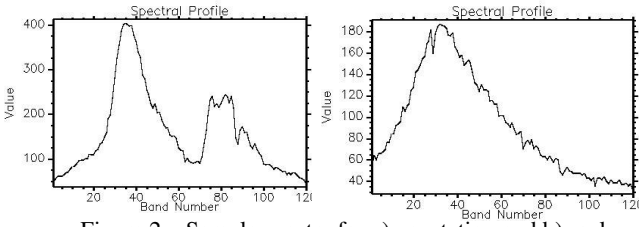


Figure 2 – Sample spectra for a) vegetation and b) rock

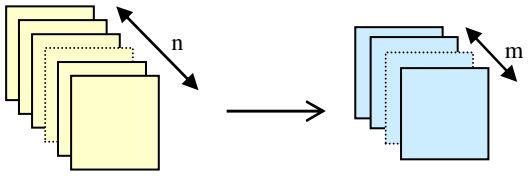


Figure 3. Feature Extraction reduces the data dimensionality

Based on this observation, one can simply proceed to perform feature selection by analyzing the correlation matrix and selecting only a few from each group of highly correlated bands. A better approach is to transform the data such that the resulting features are decorrelated. The variance of the individual frames is considered to be an indicator of information content; large values suggest high levels of information and low values indicate the presence of mostly noise. Based on this, only the features with high variance are selected for further processing.

The traditional choice for decorrelation is Principal Component Analysis (PCA). For the multidimensional random vector $\mathbf{x}$, PCA finds a linear transform $\mathbf{W}$ such that the obtained components are uncorrelated [8]:

$$\mathbf{Y}=\mathbf{W}\mathbf{x} \quad (1)$$

The transform is obtained as:

$$\mathbf{W}=\mathbf{A}_{\mathbf{x}}^{\mathbf{T}} \quad (2)$$

Where $\mathbf{A}_{\mathbf{x}}$ is the matrix formed of the normalized eigenvectors for the covariance matrix $\mathbf{\Sigma}_{\mathbf{x}}$[9].

The components of $\mathbf{y}$ are called principal components and the eigenvector matrix $\mathbf{A}_{\mathbf{x}}$ is called the principal component transform (PCT). The eigenvalues of the covariance matrix for $\mathbf{x}$ correspond to the variances of the principal components. When these eigenvalues are sorted in decreasing order (along with the corresponding permutation of the eigenvectors), we get the components of $\mathbf{y}$ sorted in the decreasing order of their variance, the approach suggested above for feature extraction.

Decorrelation based feature extraction has been observed to be less efficient when dealing with small classes in the image. These classes tend to have little influence on the band variance leading to the possibility of being discarded in the lower variance frames. In the context of target detection, losing information regarding small targets (that correspond to small classes in the image) affects the accuracy of the algorithms.

An alternative approach is provided by Independent Component Analysis (ICA). Here, given a random vector **s**, and a matrix **A** the problem is to recover this pair (**s, A**) from the available observations **x** defined as [10]:

$$\mathbf{x=As} \quad (3)$$

knowing that the vector s is formed of independent non-Gaussian components [10]:

$$p(\mathbf{s}) = \prod_{i=1}^{m} p(s_i) \quad (4)$$

where $p(.)$ refers to the probability density function and s_i refers to the components of the vector **s**.

Previous studies have shown that a solution to equation (3) exists. The components of **s** are called independent components. Algorithms solving ICA are derived from either minimizing the mutual information [11]:

$$I(s_1,...,s_m) = E\{\log \frac{p(\mathbf{s})}{\prod_{i=1}^{m} p(s_i)}\} \quad (5)$$

or by maximizing the non-Gaussianity of the components while keeping them decorrelated. In all the cases, the algorithms use gradient based optimization that often produce local optima.

Compared to PCA, ICA imposes a stronger condition on the resulting features. While both methods: try to minimize the dependencies between the components of the data, in PCA dependence minimization is achieved when the covariance between the components is zero while in ICA it is achieved through the components' independence. When the data are Gaussian, decorrelation is equivalent to independence, and PCA can be considered to be equivalent to ICA. When the data are not Gaussian, ICA becomes stronger than PCA. Here dependence in the data needs to be characterized through third and fourth order statistical measures, not present in PCA [11].

While feature extraction can be construed as a general method for any multidimensional data processing, in hyperspectral imagery it is strongly linked to the unmixing problem. Due to the spatial resolution, it is often the case that a pixel covers an area that contains more than one material and the measured value is the result of mixing the reflectance values of different materials. If the materials are physically placed adjacent to each other and the scattering of light radiation is dominated at any point by a single material, then the observed values can be considered to be a linear mixture of the reflectances of the materials in the corresponding pixel.

The linear mixing model (LMM) can be described in the following manner. Each *n*-dimensional observed pixel vector **x** present in the image cube can be expressed as [12]:

$$\mathbf{x} = \sum_{i=1}^{m} a_i \mathbf{s_i} + \mathbf{w} = \mathbf{Sa} + \mathbf{w} \quad (6)$$

where **S** is the *n*x*m* matrix of spectra ($\mathbf{s_1}$, .., $\mathbf{s_m}$) of the individual composing materials (also called *endmembers*), **a** is an *m*-dimensional vector describing the fractional abundances of the endmembers in the mixture (*abundance vector*) and **w** is the additive noise vector. The elements of the abundance vector are assumed to be positive and with unit sum [12]:

$$a_i \geq 0, i = 1,....,m \quad (7)$$

$$\sum_{i=1}^{m} a_i = 1 \quad (8)$$

Identifying various materials in the image means finding the endmembers and their abundances. The endmember linear unmixing problem plays an important role in hyperspectral image analysis [13].

Feature extraction can be seen as performing unmixing. The abundance determination is done at the same time with endmember selection. In this case, the endmembers and abundances are assumed deterministic and unknown and a maximum likelihood estimation approach is employed to determine them. Assuming this, any further processing of the feature extracted data will be efficient since the produced bands correspond to the endember contributions to each pixel.

The unmixing cannot be however directly obtained from feature extraction techniques we presented above. First, one must determine the number of endmembers present in the image. In PCA this is done by computing the sum of the eigenvalues and considering components that make up to certain percentage (ranging from 90 to 99%). In ICA, one can consider the most non-Gaussian components (according to their kurtosis) as associated to endmember abundances and discarding the rest [11].

A second problem is related to the relationship among the endmembers. In PCA and ICA, the endmembers correspond to rows in the linear transform and are orthogonal. This may be a condition too strict for endmember extraction where the base materials may be only slightly different from each other.

Finally, the biggest hurdle comes when we analyze the abundance images. The feature extraction methods we discuss do not ensure that equations (7) and (8) will hold.

3. NONNEGATIVE MATRIX FACTORIZATION (NMF) FOR FEATURE EXTRACTION

Given the observed data **x**, the goal of NMF is to find **s** and a linear mixing transform **W** both positively defined such that [14]:

$$\mathbf{x=Ws} \quad (9)$$

This approach can be understood as factorizing a data matrix subject to positive constraints. Solutions to NMF are based on constraining positivity and the gradient optimization (minimizing the distance between x and

iterations of Ws). The optimization is done by repeatedly updating **W** and **s** using [14]

$$\mathbf{W} = \mathbf{W} - \frac{\partial f(\mathbf{W},\mathbf{s})}{\partial \mathbf{W}} \quad (10)$$

$$\mathbf{s} = \mathbf{s} - \frac{\partial f(\mathbf{W},\mathbf{s})}{\partial \mathbf{s}} \quad (11)$$

where:

$$f(\mathbf{W},\mathbf{s}) = \left\| \mathbf{x} - \mathbf{W}\mathbf{s} \right\|_F^2 \quad (12)$$

and $\left\| . \right\|_F$ designates the Frobenius (or Euclidean) norm. At each step we also ensure that **W** and **s** are positive and s is normalized.

An algorithm including the positivity restrictions is presented in [14] and was used to separate a limited number of hyperspectral spectra in [15]. In Fig. 4 we present a similar approach. We note that steps 4-6 will be repeated until a convergence criterion is satisfied. This can be based on a stability measure (such as change in the values for **W** and/or **s**) or on the convergence of $f(\mathbf{W},\mathbf{s})$. In each iteration, steps 4 and 5 are applied to each component of **s** and **W** respectively.

Compared with NMF algorithms described in the literature our change is to enforce the restrictions on **s** such that they will satisfy equation (8). We note that since the two matrices were initialized to positive values and since the update step maintains the positivity, equation (7) will hold for the results. The value of ε is relatively small and is mainly used to limit the effect of the local optima. The convergence criterion is based on equation (12). When $f(\mathbf{W},\mathbf{s})$ reaches a value close to 0 (convergence condition), or no significant change is noticed (stability condition), the algorithm will stop.

We also note that the current version of the algorithm needs to have the number of features predetermined apriori. Since the purpose of the study was to investigate the value of NMF for feature extraction, we considered the number of features to be equal to the number of original spectral bands.

4. EXPERIMENTAL RESULTS

The algorithm was implemented in Matlab 6.5 and run a on a Dell Workstation with dual Intel Xeon 2.39GHz processors and 2GB of RAM. The dual processor technology, while not directly used in the code run allowed us to test the execution speed. For the first experiment we used a hyperspectral image from the Hyperspectral Digital Imagery Collection Experiment (HYDICE). It corresponds to a foliage scene taken from with a spatial resolution of 1.5m at wavelengths between 400nm and 2.5 micron. The wavelength interval assures that mostly the reflective energy of the objects is recorded

1. Randomly initialize **W** and **s** to positive values
2. Scale the columns of **s** to sum up to one
3. Repeat:
4. $s_{ij} = s_{ij} \dfrac{\left(\mathbf{W}^{\mathbf{T}}\mathbf{x}\right)_{ij}}{\left(\mathbf{W}^{\mathbf{T}}\mathbf{W}\mathbf{s}\right)_{ij} + \varepsilon}$
5. $W_{ij} = W_{ij} \dfrac{\left(\mathbf{x}\mathbf{s}^{\mathbf{T}}\right)_{ij}}{\left(\mathbf{W}\mathbf{s}\mathbf{s}^{\mathbf{T}}\right)_{ij} + \varepsilon}$
6. Scale the columns of **s** to sum up to one

while non convergence

Figure 4. Nonnegative Matrix Factorization algorithm

(the emissive energy of the objects can be ignored). The data set uses 16 bit pixel calibrated sub-scenes provided by the Spectral Information Technology Application Center.

The scene (Fig. 5) contains panels made of 8 different materials (from top row down): dark olive parachute, light olive parachute, nomex kevlar (woodland), green tenting, cotton/nylon (green woodland), nylon (green woodland), cotton (green), and desert BDU. For each of the material, there were 3 panels of sizes 1 by 1, 2 by 2 and 3 by 3 meters. A small forest patch is also present in the image.

A subset of 85 bands uniformly extracted from the data was used as input to the NMF algorithm with an epsilon value of 0.05. We ran the algorithm for 300 runs although the stability criterion was satisfied after approximately 150 runs. The average iteration execution time was 2.3s. Fig. 6 displays the graph of the error value for all the iterations. Fig. 7 displays the abundance values for the 10 most representative elements in the image a)-h) panels in the rows 1 through 8 respectively (top to bottom), i) vegetation and j) the ground. The graphs associated to panels were produced by manually selecting a pixel from the center of the first row (the largest). The vegetation abundance spectra was obtained from the center of the vegetation patch (lower left corner in Fig 5) and the exposed ground graph is based on a generic pixel.

The graphs suggest that NMF accurately separates the materials and associates them to distinct endmembers. The panels in row 1,2, 6 and 8, while composed of different colored materials all share a common predominant endmember in band 27 and vary in other less significant endmembers. We suggest that this is associated to the nylon that was used in all four cases. Similarly, the fifth and seventh row panels share two endmembers, although with different abundance levels. This is probably to cotton being used in both cases. The third (nomex Kevlar) and fourth row (nomex-kevlar) are the most distinct. The vegetation is clearly separated in a single endmember

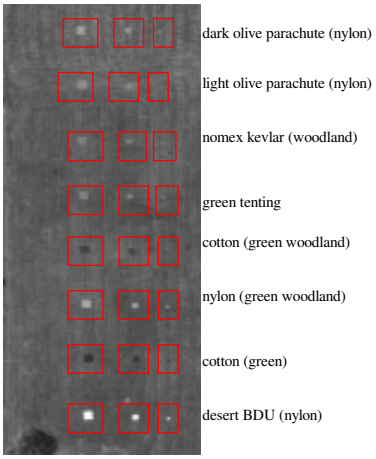


Figure 5. Hydice data scene with the panels highlighted.

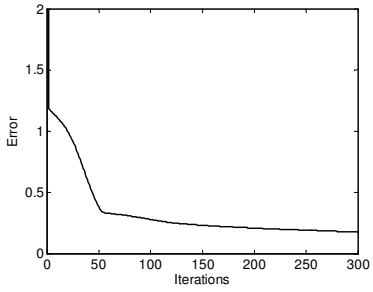


Figure 6. Plot of the error value for NMF on the Hydice data for 300 iterations

(Fig 7i) while the exposed ground does not correspond to any predominant endmember. This can be explained by the complexity of the ground as well as the presence of high noise levels in the data.

The second experiment uses data produced using a SOC 700 hyperspectral sensor currently available in our lab. The camera, pictured in Fig. 8 is able to produce 640x640 pixel images on 120 bands equally spaced within the 400nm and 900nm (i.e. visible to near-infrared range). Forty bands uniformly extracted from the image cube were used for the experiment. Fig. 9 shows the selected image. The set-up is formed of an artificial plant arranged in a light brown ceramic pot. Several real leaves (shown in enhanced green in the picture) were placed between the artificial leaves (left side, top and lower right side of the green area). To benefit from full spectrum illumination, the arrangement was placed outside on a large rock formation. The background is formed of a brick wall.

Fig. 10 displays the error graph for the NMF algorithm run on 300 iterations. We note that stability was achieved after approximately 100 rounds with minor changes after that. The average iteration execution time was 4.8s. Fig. 11 shows the abundance graph for four of the elements present in the image. While regular imaging was unable to distinguish among the real and artificial vegetation, this is easily differentiable when the NMF results are analyzed.

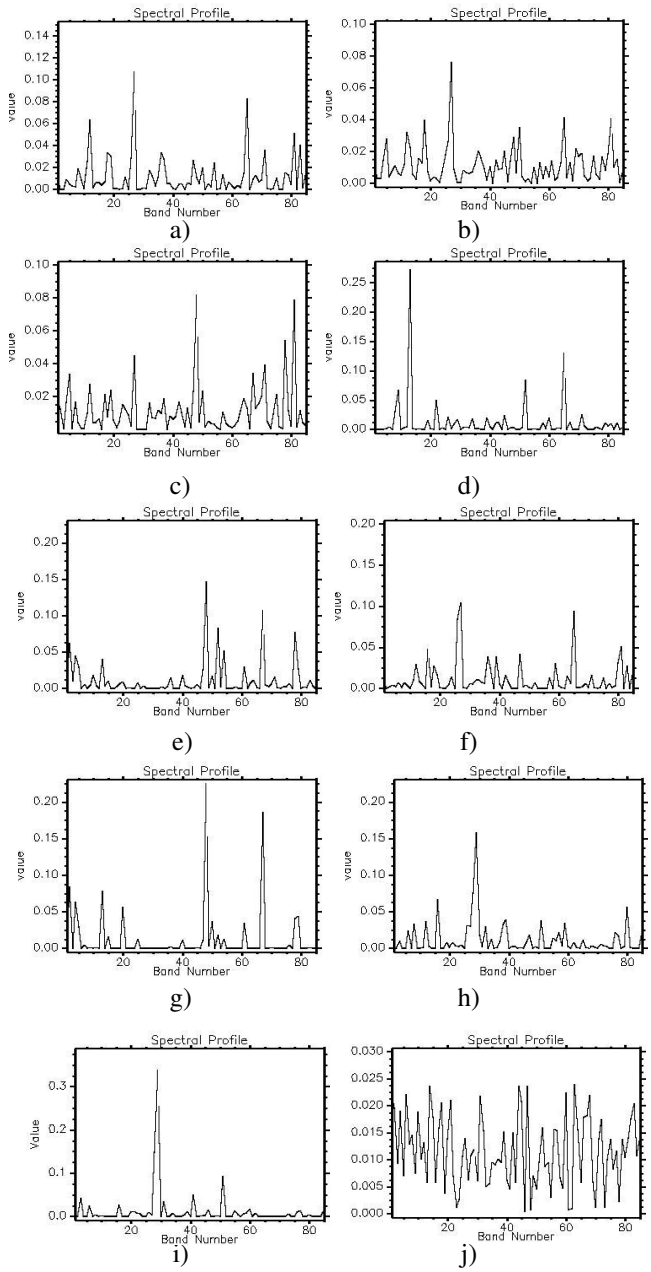


Figure 7. Abundance graphs for ten different materials present in the Hydice scene after the stability was achieved. a) – h) panels in rows 1-8 (top down), i) vegetation patch, j) exposed ground

Figure 8. Experimental setup using SOC 700 hyperspectral camera

Figure 9. SOC 700 image used in the experiments

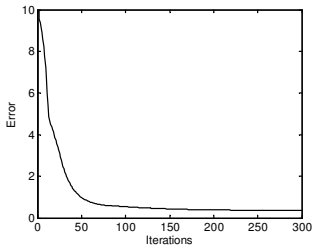


Figure 10. Plot of the error value for NMF on the SOC700 data for 300 iterations

The real vegetation exposes one endmember with significant abundance in band 30 and several other endmembers with limited contributions. Compared to that, the artificial plant information is characterized by two significant endmember contributions, associated to bands 23 and 31. The ceramic pot is characterized by two endmembers while the rock present in the image does not provide any distinct endmembers and seems to be formed as combination of various endmembers.

The rock behavior is similar to the one of the exposed ground in the Hydice image. We conjecture that this is due to the inadequacy of the linear model for this type of material. In the case of the exposed ground, the various materials forming it are no longer distinct when sensed at the current Hydice spatial resolution. Similarly, the rock in the SOC700 image was heavily speckled by various minerals that make distinct separation using linear unmixing unfeasible. Overall, however, NMF is able to accurately separate the materials composing the ground elements whenever distinct separation is possible. This is a significant step forward in unsupervised endmember extraction and nonlinear unmixing. Compared to it, neither PCA nor ICA can be employed.

5. FUTURE WORK

While the results using the NMF algorithm are encouraging, several research questions remain.

First, a significant factor affecting the processing accuracy is the number of features to be extracted. Our choice (the same as the number of original bands) while providing accurate results is prone to introduce errors. We intend to investigate possible estimation of the number of NMF components.

Second, the risk of local optima remains. Our runs, while reaching stable point have most likely also not reached the global optima. To counter this, we intend to use the value of ε in Fig. 4 as a variable parameter in a modified NMF based on simulated annealing. Whenever a stability point is achieved, ε will be modified to force a move away from the local optima.

Third, the size of the original data ranged from 10 to 80 MB. This prompted our decision to uniformly sample the bands. While this is acceptable for many instances, it is also likely that slight spectral differences among the materials present in the scene were no longer present in the sampled image. Our future research plans to redevelop the NMF algorithm such that a distributed processing is possible, allowing for increased data size, without adding significant execution time. Changes in the programming language from Matlab to C or Java are likely.

Finally, we intend to couple the development of NMF feature extraction with the design of a real time unmixing tool. Here, the currently available SOC700 camera will be connected to either a computer cluster or a multiprocessor machine (such as a currently available Sun v40z 8-way server) to provide a significant speedup. We have already modified the camera drivers to allow for partial data extraction and processing.

6. CONCLUSION

Feature extraction remains an attractive approach in processing hyperspectral images. The current techniques are focused on strong restrictions on the separability of the resulting bands and do not have a natural interpretation for the nature of hyperspectral data. Compared to them, NMF provides an elegant approach that simply assumes that the features must be separable and positively defined. When adding the condition that the features must also sum up to one pixelwise we discover that NMF also provides a solution to the classical linear unmixing problem.

Based on these initial considerations we employed a modification of the classic NMF algorithm to extract endmembers and abundances (and thus features) from two hyperspectral data sets, however without a reduction of data size. In both cases, the results suggest that NMF reaches optimal solutions that clearly separate endmember information for the data. Several open questions persist, including the ability of the algorithm to reach the global optima or what modifications are necessary for the extraction to yield a data reduction. This initial investigation prompts us to search for solutions to the questions and encourages us to consider NMF as a viable approach for feature extraction.

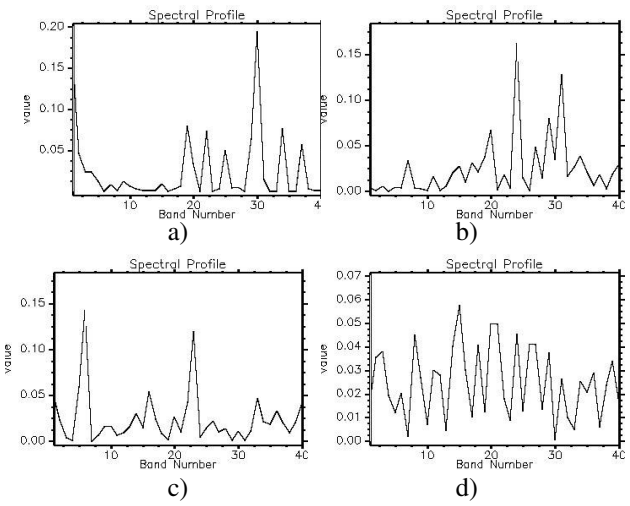


Figure 11. Abundance graphs for ten different materials present in the SOC700 scene after the stability was achieved. a) vegetation, b) artificial plant, c) ceramic pot, d) base rock

7. REFERENCES

[1] T. M. Lillesand, R.W. Kiefer, *Remote sensing and image interpretation*, John Wiley and Sons, New York, 2000

[2] J. A. Richards, X. Jia. *Remote Sensing Digital Image Analysis*, Springer, New York, 1999

[3] P.M. Mather, *Computer processing of remotely-sensed images*, John Wiley & Sons, 1987

[4] F.F. Sabins, *"Remote sensing: principles and interpretation*", W. H. Freeman and Company, New York 1987

[5] S. A. Robila, "Using Spectral Distances for Speedup in Hyperspectral Image Processing", *International Journal of Remote Sensing*, in press, 2005

[6] P.J. Ready, and P.A. Wintz, "Information extraction, SNR improvement, and data compression in multispectral imagery", *IEEE Transactions on Communications*, Vol. Com-21, 1973, 1123-1130.

[7] S. A. Robila, P. K. Varshney, "Extracting Features from Hyperspectral Data Using ICA", in P.K. Varshney, M.K. Arora editors. *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer, New York, 2004, pp. 199 - 216.

[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.

[9] T. Achalakul, and S. Taylor, "A distributed spectral-screening PCT algorithm", *Journal of Parallel and Distributed Computing*, **63**, no. 3, 373-384, 2003

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[11] S. A. Robila, "Independent Component Analysis (ICA)", in P.K. Varshney, M.K. Arora editors. *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer, New York, 2004, pp. 109 - 132.

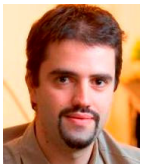
[12] N. Keshava and J. Mustard. \Spectral Unmixing", IEEE Signal Processing Magazine, pp. 44-57, January 2002.

[13] A. Plaza, P. Martinez, R. Perez and J. Plaza. \A Quantitative and Comparative Analysis of Endmember Extraction Algorithms from Hyperspectral Data", IEEE Transactions on Geoscience and Remote Sensing, Vol. 42, No. 3, pp. 650-663, 2004

[14] D. Lee and H. Seung. "Algorithms for Non-Negative Matrix Factorization", *Advances in Neural Processing*, 2000

[15] P. Pauca, J. Piper, and R. Plemmons, "Nonnegative Matrix Factorization for Spectral Data Analysis", to appear in *Linear Algebra and Applications*, 2006

8. AUTHOR'S BIOGRAPHY



Dr. Stefan Robila is Assistant Professor of Computer Science and Director of the Center for Imaging and Optics at Montclair State University, Montclair, NJ. Stefan received a B.S. in Computer Science from University of Iasi in 1997, followed by an M.S. in Computer Science and a Ph.D. in Computer Information Science in 2000 and 2002 respectively, both from Syracuse University. His current research interests are remote sensing, signal and image processing, multispectral / hyperspectral imagery, data mining and pattern recognition.

Lukasz Grzegorz Maciak is a graduate student at Montclair State University, working towards a M.S. in Computer Science (expected Summer 2006). Lukasz received a B.S. in Computer Science from Montclair State University in 2004. He is currently doing research in hyperspectral imaging and Feature Extraction.