ENDMEMBER EXTRACTION FROM HYPERSPECTRAL IMAGERY USING A PARALLEL ENSEMBLE APPROACH WITH CONSENSUS ANALYSIS

F. Ayuso, J. Setoain, M. Prieto, C. Tenllado, F. Tirado

Department of Computer Architecture
Complutense University of Madrid, Spain

J. Plaza, A. Plaza

Department of Technology of Computers
and Communications,
University of Extremadura, Spain

1. ABSTRACT

We have explored in this paper a framework to test in a quantitative manner the stability of different endmember extraction and spectral unmixing algorithms based on the concept of Consensus Clustering. The idea is to investigate if the sensibility of those algorithms to the number of endmembers can be used to estimate this parameter itself. Preliminary results on synthetic data reveal that the proposed scheme, which can be implemented efficiently in parallel, can compete with state-of-the-art schemes.

2. HYPERSPECTRAL UNMIXING AND ENDMEMBER EXTRACTION

A general approach to unmix hyperspectral data consists of (1) an endmember extraction step and (2) the estimation of fractional abundances. The former identifies a collection of hidden pure spectra – the endmembers – from the hyperspectral scene, whereas the latter finds the fractional abundances that indicate the proportion of each endmember material in the mixed pixels.

Endmember extraction is the most critical and challenging step in this process and has received considerable attention in recent years. From a viewpoint of algorithm design, three major issues determine its performance [1]:

1. The **learning rules** that drive the searching process. Most algorithms implicitly or explicitly assume that pure pixels are present in the data to speed up this process, and use different criteria when searching for those pure pixels. Two major criteria are multidimensional geometry-based simplex volume maximization and pixel spectral signature similarity [2]. A promising scheme based on the Non-Negative Matrix Factorization without the pure-pixel assumption was recently proposed in [3].

2. The **stopping criteria** of the process. Most popular endmember extraction algorithms, such as N-FINDR, the Simplex Growing Algorithm (SGA), the Pixel Purity Index (PPI), the Vertex Component Analysis (VCA) or the Automatic Target Generation Process (ATGP) use as stopping criteria the number of endmembers to be searched. This predefined parameter must be estimated beforehand and becomes critical for performance. If it becomes too low, then not all desired endmembers will be extracted – specifically, those being *weak* endmembers –, or, conversely if it becomes too high, some extracted endmembers may turn out to be mixed signatures [1]. A popular approach is to resort to the Virtual Dimensionality (VD) proposed in [4].

3. The **initial conditions** use by the algorithm. Many algorithms use a random restart but a proper initial selection of the endmebers can be very beneficial [1].

Research efforts have mainly focused on the first issue [5, 3, 6, 2] but both the stopping criteria and the initial conditions also have a significant impact on performance [1].

3. RESEARCH GOALS

Our ultimate goal is to devise a feasible and robust method that can set the stopping criteria – the number of endmembers – from the data itself without a priori knowledge about the underlying data distributions. While the VD method is quite efficient, it also suffers from the following issues. First, it depends on an input parameter (false alarm probability) which has to be carefully set

in advance in order to obtain satisfactory performance. Second, the VD may not identify weak endmembers which correspond to anomalous endmembers.

In order to address these issues, we have developed a framework to test in a quantitative manner the stability of different endmember extraction and spectral unmixing algorithms. The idea is to investigate if the sensibility of those algorithms to the number of endmembers can be used to estimate this parameter itself.

Our approach is based on *Consensus Clustering* [7], a methodology proposed in the context of gene expression data analysis to represent the consensus across multiple runs of a clustering algorithm. It has been use effectively to determine the number of clusters in the data and to assess the stability of the discovered clusters [8].

Preliminary results on synthetic data reveal that the proposed scheme can compete with state-of-the-art schemes but its computational cost is very high. Fortunately, this analysis is inherently parallel and we can take advantage of the recent progress in commodity high-performance hardware to mitigate this problem.

In the final contribution we will include the results of our experiments on both simulated and real hyperspectral data aimed at evaluating the effectiveness of the methodology in both estimating the number of endmembers in the data and extracting such endmembers.

4. REFERENCES

[1] A. Plaza and C.-I. Chang, "Impact of initialization on design of endmember extraction algorithms," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.

[2] Qian Du, Nareenart Raksuntorn, Nicolas H. Younan, and Roger L. King, "End-member extraction for hyperspectral image analysis," *Applied Optics*, vol. 47, no. 28, pp. F77–F84, 2008.

[3] Lidan Miao and Hairong Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 3, pp. 765–777, March 2007.

[4] Chein-I Chang and Qian Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 3, pp. 608–619, March 2004.

[5] Chao-Cheng Wu and Chein-I Chang, "Does an endmember set really yield maximum simplex volume?," *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp. 3814–3816, July 2007.

[6] J.M.P. Nascimento and J.M.B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 4, pp. 898–910, April 2005.

[7] Stefano Monti, Pablo Tamayo, Jill Mesirov, and Todd Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," *Machine Learning*, vol. 52, no. 1-2, pp. 91–118, 2003.

[8] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization.," *Proceedings National Academic Scice U S A*, vol. 101, no. 12, pp. 4164–4169, March 2004.