Endmember Extraction Methods and Unmixing Techniques: a short Review

Miguel A. Veganzones, Manuel Graña

Computational Intelligence Group, Basque Country University,
http://www.ehu.es/computationalintelligence

Abstract. The analysis of hyperspectral images on the basis of the spectral decomposition of their pixels through the so called spectral unmixing process, has applications in tematic map generation, target detection and unsupervised image segmentation. The critical step is the determination of the endmembers used as the references for the unmixing process. We give a comprehensive enumeration of the methods used in practice, because of its implementation in widely used software packages, and those published in the literature. We have structured the review according to the basic computational approach followed by the algorithms: those based on the computational geometry formulation, the ones following lattice computing ideas and heuristic approaches with a weak formal foundation.

1 Introduction

In the field of hyperspectral image processing, Spectral Unmixing [15] is the computation of the fractional contribution of elementary spectra, called endmembers because they constitute the vertices of a convex polytope covering (most of) the image data points in high dimensional space. Spectral Unmixing can be used for taget detection, tematic map building and unsupervised segmentation. The underlying image model is a linear mixture of the endmembers, with positive coefficients that sum up to one. Given the endmembers, enforcing these conditions when performing Spectral Unmixing involves constrained non-negative least squares estimation, which can be a very computationally expensive process by itself. When the convex polytope defined by the provided endmembers does not cover all te data points, it is not possible to enforce these conditions. Solving the problem of providing the appropriate set of endmembers is a precondition to the realization of the Spectral Unmixing. Early approaches to endmember determination were based on human expertise. The prior knowledge about the contents of the imaged terrain was used by the expert to select some candidate endmember spectra from a provided library. The spectra in the library must have some correspondence with the sensor characteristics, in order to perform the matching and unmixing. Besides the methodological questions, this approach is not feasible when trying to process large quantities of image data.

Current approaches try to induce the endmembers from the image data. They either try to select some image pixel spectra as the best approximation to the endmembers in the image [20,9], or to compute estimations of the endmembers on the basis of the transformations of the image data (i.e. [6,13]). The latter is the predominant class of tecniques in the literature. Previous reviews found in the literature [21] make some emphasis on the degree of automation to classify the algorithms. In this review the emphasis will be on the computational foundations, assuming that user interaction must be minimal or null. We distinguish three fundamental approaches:

- Geometric approaches, that try to find a simplex that covers the image data.
- Lattice computing approaches, that use some kind of lattice theoretic formalism or mathematical morphology approach.
- Heuristic approaches, that are not very rigorously formalized under a theoretical framework.

There are some problems that we will not touch in deep in this review. The problem of the endmember induction algorithms initialization is discussed in [19], where an Endmember Initialization Algorithm (EIA) is proposed. The number of spectral signatures that form an hyperspectral image is usually unknown. Recently, a new concept denoted Virtual Dimensionality (VD) [5,3,19] has been used for automated search of the optimal number of endmembers in an image. Most endmember induction algorithms are quite computationally expensive, so a mention is due to the efforts to obtain distributed implementations [22] that may help them to be a feasible approach for real-life applications. The use of off-the-shelf Graphical Processing Units [24] are a low cost way to obtain substantial speed-ups. The outline of the paper is as follows: section 2 describe some geometrically oriented metods, section 3 describe methods based on lattice computing or mathematical morphology, section 4 describe some heuristic methods. Finally, we give some concluding remarks in section 5.

2 Geometric enmember induction methods

Geometric methods follow the formal definition of the endmbers, they search for the vertices of a convex set that covers the image data. Because the distribution of the data in the hyperspace is usually tear-shaped they look for the minimum simplex that covers all the data. Unless said otherwise, the algorithms search for a prefixed number of endmembers, defined by the user. The first such methods is the Minimum Volume Transform, proposed by [7] that introduces two nonortohonormal transforms, the dark-point-fixed (DPF) transform and the fixed-point-free (FPF) transform that map the data onto the minimal simplex that contain all the data points.

One of the earliest approaches is the N-FINDR algorithm proposed in [26]. The N-FINDR algorithm is a selection algorithm. Its works are described as follows: it starts with a random collection of image pixel spectra, corresponding to the initial set of endmembers. Then, each of the remaining image pixels is considered as a candidate to replace each endmember, if doing so the volume of the simplex increases, then it is accepted as the new endmember. The process ends when no more replacements are possible. The N-FINDR algorithm requires a dimension reduction step, originally an orthogonal subspace projection (OSP) to an space of dimension N-1, where N is the number of endmembers. This set of endmembers found by the N-FINDR would not allow the nonnegative unmixing of the pixel spectra in general.

The Convex Cone Analysis (CCA) [13] is based on the fact that the vectors formed by discrete radiance spectra are linear combinations of nonnegative components, and they lie inside a nonnegative convex region. The object of CCA is to find the boundaries of this convex region, which can be used as endmember spectras. The algorithm performs a Principal Component Analysis (PCA) dimension reduction based on the sample spectral correlation matrix of the image. In this reduced space, the endmembers must define a convex cone on the positive hyperquadrant of the space, whose apex is in the space origin. Endmembers are points with exactly $c-1$ zero coefficients in the PCA decomposition, c being the number of eigenvectors selected.

The approach followed in [1] search for the optimal simplex using a simulated annealing algorithm (SA) whose state configuration is given by the partition of the faces of the convex hull of the f image pixel spectra, after a reduction to N-1 dimensions by the Minimum Noise Fraction (MNF) algorithm. The partition in the configuration space defines a simplex convering the image data whose vertices are the candidate endmembers. The objective function minimized is the simplex volume. This approach is followed by the generation of endmember bundles that allow the computation of bounds on the abundance images.

The Iterated Constrained Endmembers (ICE) [2]algorithm performs the minimization of a regularized residual sum of squares (RSS). The regularization term is the volume of the simplex. The name of the algorithm comes from the minimization schema applied. Given that the free parameters are the endmembers and the proportions (abundances) for each pixel the algorithm iterates the solution of the tow interleaved and interdependen minimization problems (much like in an Expectation Maximization process): first the proportions are computed by quadratic programming problem solving assuming that the endmembers are known, then the endmembers are computed as the direct minimization of the RSS functional. The addition of an sparsity promoting term in the RSS functional gives way to SPICE [27]. This sparsity promoting term is derived as the substitution of a Gaussian prior by a Laplacian prior in a bayesian formulation of the RSS functional. The SPICE algorithm allows the selection of the appropriate number of endmembers based on the sparsity measure. The ICE algorithm does need a dimension refuction step, performed by the MNF algorithm.

The Vertex Component Analysis algorithm (VCA) is presented in [18]. The algorithm is unsupervised and exploits that the affine transformation of a simplex is also a simplex. It works with projected and unprojected data. The algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted.

In [4] a simplex-based endmember extraction algorithm, called Simplex Growing Algorithm (SGA), is presented. It is a sequential algorithm to find a simplex with the maximum volume every time a new vertex is added. Virtual Dimensionality (VD) is applied as stopping rule to determine the number of vertices required. SGA improves N-FINDR by including a process of growing simplexes one vertex at a time until the desired number of vertices is reached, which results in a high computational complexity reduction; and by selecting an appropiate initial vector to avoid the use of random vectors as initial condition, which produces different sets of final endmembers if different sets of randomly generated initial endmembers are used.

In [16] a method for endmember extraction for highly mixed data, when there are not pure pixels in the hyperspectral image, is presented. The proposed method, called Minimum Volume Constrained Nonnegative Matrix Factorization (MVC-NMF) takes advantage of the fast convergence of NMF schemes and at the same time eliminates the pure-pixel assumption. It consists in the reformulation of an NMF cost function introducing an volume regularization term, much like the ICE, substituting the RSS by the NMF criteria.

3 Lattice computing endmember induction methods

Lattice computing can be defined as the collection of computational methods that either are defined on the algebra of lattice operators inf and sup, with the addition, or employ lattice theory to generalize previous approaches. Mathematical Morphology is a very successfull case of this paradigm, but it also emcompasses some fuzzy systems approaches and neural networks. The Automated Morphological Endmember Extraction (AMEE) method [20] is a mathematical morphology inspired algorithm for the extraction of the endmembers from the data. It is based on the definition of multispectral erosion and dilation operators, which are then used to compute the Morphological Eccentricity Index (MEI) over kernels of increasing size that are computed over all the pixels in the image. The result is a MEI image whose maxima correspond to the endmember pixels. The method does not need a dimension reduction step.

The concept of morphological independence, later reformulated as lattice independence, was the basic tool in the approach proposed in [9,12,11,10]. The set of endmembers was formulated a set of morphologically independent vectors, either in a dilative or erosive sense, or both. There the Associative Morphological Memories, later renamed Lattice Associative Memories, are proposed as detectors of morphologically independent vectors. The algorithm works in a single pass over the sample data.

This approach has been followed by the one proposed in [8]. The relationship between strong lattice independence and affine independence was proven. Then it was found that most vectors in the erosive and dilative lattice memories are strong lattice independent. Therefore, the mere construction of the lattice memories provide a way to obtain the convex hull of the data. Provided an endmember selection mechanism, the algorithm can obtain in a single pass over the image a set of endmembers.

4 Heuristic endmember extraction methods

The heuristic methods collects a set of heterogeneus endmember extraction methods that use different approaches not grouped under a strict theorical background for endmember induction. The most famous and widely used method, due to its inclusion in the ENVI software package, it is the Pixel Purity Index (PPI) algorithm introduced in [14]. The algorithm reduce the data dimensionality and makes a noise whitened process by MNF method, and then it determines the pixel purity by repeatedly projecting data onto random unit vectors. The extreme pixel in each projection is counted, identifying the purest pixels in scene. PPI requires the human intervention to select those extreme pixels that best satisfy the target spectrum.

Althought PPI has been intesively used, its implementation aspects are kept unknown due to the limited published results. In [6] PPI is investigated and a fast iterative algorithm to implement PPI is proposed. The Fast Iterative PPI algorithm (FIPPI) improves PPI in several aspects. FIPPI produces an appropiate initial set of endmembers to speed up the process. Additionally, it estimates the number of endmembers to be generated by Virtual Dimensionality (VD). FIPPI is also an unsupervised and iterative algorithm, where an iterative rule is developed to improve each of the iterations until it reaches a final set of endmembers.

In [25] the well known Independent Component Analysis (ICA) method is the base of the proposed approach for endmember extraction and abundance quantification. The algorithm, called ICA-based Abundance Quantification Algorithm (ICA-AQA), is a high-order statistics-based technique, that can accomplish endmember extraction and abundance quantification simultaneously in one-shot operation. [17] analyzes the use of ICA and Independent Factor Analysis (IFA) for unmixing tasks, showing that the statistically independent of the sources, assumed by ICA and IFA, is violated in the hyperspectral unmixing, compromising the performance of ICA/IFA algorithms for this purpose. It concludes that the accuracy of this ICA/IFA-based methods tends to improve with the increase of the signature variability and the signal-to-noise ratio.

The Spatial-Spectral Endmember Extraction algorithm (SSEE) proposed in [23] is another projection based method that works by analyzing a scene in parts (subsets), such that it increases the spectral contrast of low contrast endmembers, thus improving the potential for these endmembers to be selected. The SSEE method uses a singular value decomposition (SVD) to determine a set of basis vectors that describe most of the spectral variance for subsets of the image. Then the full image dataset is projected onto the locally defined basis vectors to determine a set of candidate endmember pixels from where the final endmembers are selected. For that, it searchs for spectrally similar but spatially independent endmembers. This is realized by imposing spatial constraints for averaging spectrally similar endmembers.

5 Conclusions

The field of hyperspectral image processing has been an applicaction domain for many pattern recognition techniques. Among them, spectral unmixing offers the appealing of a physical image formation model with an easy interpretation. It also allows subpixel resolution results. Therefore, increasingly Spectral Unmixing will be a tool of hyperspectral image analysis. The requisite for this analysis is the determination of the endmembers. The current approaches reviewed in this paper favor the endmember induction from the image data. It is desirable that the endmembers have some physical meaning, which is more likely in the case of approaches that perform a selection from the image pixel spectra. However, these approaches usually do not produce convex polytopes that cover all the image data points, so that the candidate set of endmembers do not fit into the formal definition of endmembers. Geometrically oriented methods are the best theretically grounded ones, however they ask for great computational resources and the endmembers that they obtain do not have a clear physical meaning.

Acknowleddgements

This work is partially supported by grant MEC DPI2006-15346-C03-03. Miguel A. Veganzones has a predoctoral grant from the Basque Goverment. The Computational Intelligence Group has earned a six year grant as University Research Group from the Basque Goverment.

References

1. C.A. Bateson, G.P. Asner, and C.A. Wessman. Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, 38:1083–1094, 2000.
2. M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J.F. Huntington. Ice: a statistical approach to identifying endmembers in hyperspectral images. *Geoscience and Remote Sensing, IEEE Transactions on*, 42:2085–2095, 2004.
3. Autor Chein-I Chang. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. 2003.
4. C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang. A new growing method for simplex-based endmember extraction algorithm. *Geoscience and Remote Sensing, IEEE Transactions on*, 44:2804–2819, 2006.
5. Chein-I Chang and Qian Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 42:608–619, 2004.
6. Chein-I Chang and A. Plaza. A fast iterative algorithm for implementation of pixel purity index. *Geoscience and Remote Sensing Letters, IEEE*, 3:63–67, 2006.

7. M.D. Craig. Minimum-volume transforms for remotely sensed data. *Geoscience and Remote Sensing, IEEE Transactions on*, 32:542–552, 1994.
8. Mark S Schmalz Gerhard X Ritter, Gonzalo Urcid. Autonomous single-pass endmember approximation using lattice auto-associative memories. 2008.
9. M. Grana and J. Gallego. Associative morphological memories for endmember induction. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, volume 6, pages 3757–3759 vol.6, 2003.
10. M. Grana and J. Gallego. Hyperspectral image analysis with associative morphological memories. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III–549–52 vol.2, 2003.
11. M. Grana, J. Gallego, and C. Hernandez. Further results on amm for endmember induction. In *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, pages 237–243, 2003.
12. M. Grana, P. Sussner, and G. Ritter. Associative morphological memories for endmember determination in spectral unmixing. In *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on*, volume 2, pages 1285–1290 vol.2, 2003.
13. A. Ifarraguerri and C.-I. Chang. Multispectral and hyperspectral image analysis with convex cones. *Geoscience and Remote Sensing, IEEE Transactions on*, 37:756–770, 1999.
14. Robert Green Joseph Boardman, Fred Kruse. Mapping target signatures via partial unmixing of aviris data. 1995.
15. N. Keshava and J.F. Mustard. Spectral unmixing. *Signal Processing Magazine, IEEE*, 19:44–57, 2002.
16. 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*, 45:765–777, 2007.
17. J.M.P. Nascimento and J.M.B. Dias. Does independent component analysis play a role in unmixing hyperspectral data? *Geoscience and Remote Sensing, IEEE Transactions on*, 43:175–187, 2005.
18. 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*, 43:898–910, 2005.
19. A. Plaza and C.-I. Chang. Impact of initialization on design of endmember extraction algorithms. *Geoscience and Remote Sensing, IEEE Transactions on*, 44:3397–3407, 2006.
20. A. Plaza, P. Martinez, R. Perez, and J. Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *Geoscience and Remote Sensing, IEEE Transactions on*, 40:2025–2041, 2002.
21. A. Plaza, P. Martinez, R. Perez, and J. Plaza. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 42:650–663, 2004.
22. A. Plaza, D. Valencia, J. Plaza, and Chein-I Chang. Parallel implementation of endmember extraction algorithms from hyperspectral data. *Geoscience and Remote Sensing Letters, IEEE*, 3:334–338, 2006.
23. D.M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng. Integration of spatial-spectral information for the improved extraction of endmembers. *Remote Sensing of Environment*, 110:287–303, October 2007.
24. J. Setoain, M. Prieto, C. Tenllado, A. Plaza, and F. Tirado. Parallel morphological endmember extraction using commodity graphics hardware. *Geoscience and Remote Sensing Letters, IEEE*, 4:441–445, 2007.

25. J. Wang and C.-I. Chang. Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 44:2601–2616, 2006.
26. Michael E Winter. N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. 1999.
27. A. Zare and P. Gader. Sparsity promoting iterated constrained endmember detection in hyperspectral imagery. *Geoscience and Remote Sensing Letters, IEEE*, 4:446–450, 2007.