Morphological Memories for Feature Extraction in Hyperspectral images

M. Graña, X. Albizuri, A. D'Anjou*

Dept. CCIA, UPV/EHU, San Sebastian, Spain

Abstract. In previous papers we proposed Associative Morphological Memories (AMM)as tools for endmember extraction in hyperspectral images. Linear Spectral Unmixing (LSU) based on these endmembers is a kind of unsupervised image segmentation. In this paper we propose that the fractional abundance coefficients may be used as features for the construction of supervised pixel spectra classifiers. Thus we compare them with two well-known linear feature extraction algorithms: Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

1 Introduction

Construction of supervised classifiers often employs some feature extraction algorithms, which are data dimension reduction procedures applied to the experimental data prior to training or operation of the classifier. The goals of feature extraction are both computational efficiency and enhanced discrimination of the data classes. Linear feature extraction algorithms, like Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [1], Independent Component Analysis (ICA) [6] are defined as a linear transformation that minimizes some criterion function, like the mean square error (PCA), a class separability criterion (LDA) or an independence criterion (ICA). The alternative approach we propose in this paper is to characterize the data by a convex region that encloses it or most of it. The features extracted are the coordinates of the data points in this region relative to its vertices: the convex coordinates. Depending on the application, the meaning of these vertices varies. In hyperspectral image processing they are identified with endmember materials in the assumed linear mixing model [8], in which several basic materials (endmembers) are combined according to their abundance coefficients at each image pixel. The computation of the abundance coefficients, given a pixel spectrum and a set of endmembers, is the so-called unmixing procedure. If the endmembers are known *a priori*, the unmixing procedure is equivalent to the parallel detection of the spectral features represented by the endmembers (i.e.: materials). If the endmembers are inferred from the image data, the procedure may be interpreted as an unsupervised segmentation of the image. However the focus in this paper is assuming the abundance coefficients as the feature vectors for the supervised classifier construction. For the endmember induction we apply two alternative methods:

---

the Convex Cone Analysis (CCA) [7] and an approach based on Associative Morphological Memories (AMM) morphological independence [2, 3].

2   Linear feature extraction: ICA and PCA

The Independent Component Analysis (ICA) [6] assumes that the data is a linear combination of nongaussian, mutually independent latent variables with an unknown mixing matix. The ICA reveals the hidden independent sources and the mixing matrix. That is, given a set of observations represented by a d dimensional vector $\mathbf{x}$, ICA assumes a generative model $\mathbf{x} = \mathbf{As}$, where $\mathbf{s}$ is the M dimensional vector of independent sources and $\mathbf{A}$ is the $d \times M$ unknown basis matrix. The ICA searches for the linear transformation of the data $\mathbf{W}$, such that the projected variables $\mathbf{Wx} = \mathbf{s}$ are as independent as possible. It has been shown that the model is completely identifiable if the sources are statistically independent and at least $M - 1$ of them are non gaussian. If the sources are gaussian the ICA transformation could be estimated up to an orthogonal transformation. Estimation of mixing and unmixing matrices can be done maximizing diverse objective functions, among them the non gaussianity of the sources and the likelihood of the sample. We have used the FastICA [5] algorithm available at http://www.cis.hut. /projects/ica/fastica. The Principal Component Analysis (PCA)[1] is a well-known linear dimension reduction procedure that is optimal in the sense of the mean squared error. It consists in the selection of the largest eigenvalue eigenvectors of the data covariance matrix. These eigenvectors constitute the transformation matrix. The selection of the number of eigenvectors or the independent components can be made attending to some quantitative criteria, but in our experiment below we selected the number of components in the ground truth image.

3   Linear mixing model

The linear mixing model [8] can be expressed as follows:

$$\mathbf{x} = \sum_{i=1}^{M} a_i \mathbf{s}_i + \mathbf{w} = \mathbf{Sa} + \mathbf{w}, \tag{1}$$

where $\mathbf{x}$ is the d-dimension pattern vector, $\mathbf{S}$ is the $d \times M$ matrix whose columns are the d-dimension vertices of the convex region covering the data $\mathbf{s}_i, i = 1, .., M$, $\mathbf{a}$ is the M-dimension fractional abundance vector, and $\mathbf{w}$ is the d-dimension additive observation noise vector. The linear mixing model is subjected to two constraints on the abundance coefficients. First, to be physically meaningful, all abundance coefficients must be non-negative $a_i \geq 0, i = 1, .., M$. Second, to account for the entire composition, they must be fully additive $\sum_{i=1}^{M} a_i = 1$. Once the convex region vertices have been determined the Linear Spectral Unmixing (LSU) is the computation of the matrix inversion that gives the coordinates of the point inside the convex region. The simplest approach is the unconstrained

least squared error estimation given by:

$$\hat{\mathbf{a}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{x}. \tag{2}$$

In the setting of hyperspectral image processing, the convex coordinates are interpreted as the fractional abundance coefficients of the endmember materials in the scene pixel.

4 The Convex Cone Analysis (CCA)

The CCA was proposed by [7]. The basic idea is that after PCA of the spectral correlation matrix, the data falls in a cone shaped region in the positive subspace centered in the first eigenvector. Given the $N \times M \times d$ hyperspectral image, it is reorganized as a $NM \times d$ matrix $\mathbf{X}$. The spectral correlation matrix is computed as $\mathbf{C} = \mathbf{X}^T\mathbf{X}$. Let it be $\mathbf{C} = \mathbf{PLP}^T$ the PCA decomposition of the correlation matrix, select the first c eigenvectors $[\mathbf{p}_1, .., \mathbf{p}_c] = \mathbf{P}_c$ and search for the boundaries of the convex region characterized by $\mathbf{x} = \mathbf{p}_1 + a_1\mathbf{p}_2 + .. + \mathbf{a}_{c-1}\mathbf{p}_c \geq \mathbf{0}$. The vertices of this region are the points with exactly $c - 1$ zero components. The CCA algorithm searches among all the $\binom{b}{c-1}$ possible combinations of eigenvectors performing the following test. Let it be $[\mathbf{p}(\gamma_1), .., \mathbf{p}(\gamma_{c-1})] = \mathbf{P}'$ the selected set of eigenvectors. Solve the set of equations $\mathbf{P}'\mathbf{a} = \mathbf{0}$ and compute $\mathbf{x} = \mathbf{P}_c\mathbf{a}$. If $\mathbf{x}$ has exactly $c - 1$ zero components it is a vertex of the convex region data. In practice, each component is tested against a threshold. However, as the combinatorial space grows the problem becomes intractable. We implemented an straightforward random search. Application of more sophisticated random search algorithms like genetic algorithms may be of interest for large problems. The CCA algorithm provides the endmembers that may be used to compute the abundance images.

5 The AMM approach

The Associative Morphological Memories (AMM) [9, 10, 11] are the morphological counterpart of the well known Hopfield Associative Memories [4]. AMM's are constructed as correlation matrices computed by either Min or Max matrix product. Dual constructions can be made using the dual Min and Max operators. The AMM selective sensitivity to specific types of noise (erosive and dilative noise) is of special interest to us. It was established that AMM are able to store and recall morphologically strongly independent sets of patterns. To obtain general noise robustness [9], [11], [12] proposed the kernel method. Related to the construction of the kernels, [11] introduced the notion of morphological independence. Here we distinguish erosive and dilative versions of this definition: Given a set of pattern vectors $X = \left(\mathbf{x}^1, ..., \mathbf{x}^k\right)$, a pattern vector $\mathbf{y}$ is said to be morphologically independent of X in the erosive sense if $\mathbf{y} \not\leq \mathbf{x}^\gamma; \gamma = \{1, .., k\}$, and morphologically independent of X in the dilative sense if $\mathbf{y} \not\geq \mathbf{x}^\gamma; \gamma = \{1, .., k\}$. The set of pattern vectors X is said to be morphologically independent in either

sense when all the patterns are morphologically independent of the remaining patterns in the set.

The region of the space enclosed by a set of vectors that are morphologically independent in both erosive and dilative senses simultaneously is a high dimensional box that approaches the minimal simplex enclosing the data points. The steps in the endmember induction procedure are detailed in previous papers [2, 3]. In short it may be described as follows: At each instant we maintain two erosive and dilative MAM built up from the detected endmembers. The set of endmembers is started with a randomly selected sample data point (image pixel). The data sample (image pixels) is processed sequentially, for each data point $\{\mathbf{f}(i) \in \mathbb{R}^d; i = 1, .., n\}$ we compute the eroded and dilated versions of it previous to using it as inputs to the dilative and erosive MAM respectively. Failing to recover an stored endmember means that the data point has some kind of morphological independence as is a candidate to be a new endember. There is a gain parameter α controls the amount of flexibility in the discovering of new endmembers. Usually $\alpha = 2$ gives the desired results.

6   Experimental results

We have applied the linear dimension reduction methods and the LSU based on the discussed endmember induction methods to two well known real hyperspectral experimental data, obtained by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) developed by NASA JPL which has 224 contiguous spectral channels covering a spectral region from 0.4 to 2.5 mm in 10 nm steps. The first hyperspectral image used for this work correspond to the Indian Pines 1992 image. It is a 145 by 145 pixel image with 220 spectral bands. The available image ground truth designates 16 mutually exclusive classes of land cover [13].

The second is real hyperspectral data collected by the AVIRIS imaging spectrometer in 1998 over Salinas Valley, California. The full scene consists of 512 lines by 217 samples with 224 spectral bands with a spatial pixel resolution of 3.7mx3.7m. The available ground truth has 15 classes. When applying the PCA, ICA and CCA methods we have set the target dimension to the exact number of ground truth components. Our AMM approach needed the setting of the noise gain parameter α. Setting $\alpha = 2$ we obtained 12 endmembers on the Salinas image. Setting $\alpha = 3$ we obtained 6 endmembers on the Salinas image.

The supervised classifiers employed were the Nearest Neighbor (NN), the Gaussian Classifier (GC) using the Euclidean distance, and the Support Vector Machines (SVM) [14] with a Radial Basis Function (RBF) kernel of identical unit variances, using the implementation by Anton Schwaighofer available at http://www.cis.tugraz.at/igi/ aschwaig/software.html. No attempt has been made to fine tune the SVM. The motivation for this selection of classifiers is that they do not introduce additional bias in the experiment, which is aimed to show the value of the LSU as feature extraction algorithm.

The experiment consisted in 30 repetitions of the construction and validation of the clasifiers over 50% random partitions of the data, which preserve the a

	Nearest Neigh.	Gaussian C.	SVM
raw	0.19	0.08	0.23
PCA	0.16	0.05	0.24
ICA	0.14	0.03	0.35
CCA	0.33	0.20	0.48
AMM	0.34	0.25	0.64

Table 1: Correct recognition over the Indian Pines data.

	Nearest Neigh.	Gaussian C.	SVM
raw	0.08	0.11	0.21
PCA	0.33	0.17	0.36
ICA	0.13	0.07	0.38
CCA	0.57	0.41	0.75
AMM	0.61	0.50	0.89

Table 2: Correct recognition over the Salinas data.

priori distributions of the classes. We did not perform any band selection or smoothing of the pixel spectra in the experimental results presented here. The results of the experiment are presented in tables 1 and 2. They consist of the average accuracy of the classifiers.

The SVM improves greatly over the other classifiers, as may be expected from the results in the literature. However, we are more interested in analyzing the results by rows. The results on the raw data are very bad, but, surprisingly, PCA and ICA do not improve very much over them most of the times. Finally, both LSU based methods, CCA and AMM improve substantially over the linear projection methods. The SVM with AMM LSU feature extraction gives an almost state of the art result.

7 Conclusions

We claim that the convex coordinates of the data points based on the vertices of a convex region (approximately)covering the data can be used as features for the construction of supervised classifiers. The experiment on a very noisy hyperspectral image confirm this intuition and opens the way for further systematic experimentation with other hyperspectral images and other kinds of data. Besides, this experiment confirms also that our approach to the induction of endmembers from the data, the AMM approach is, at least, comparable to other well-established methods, like CCA. The idea of using approximations of the convex hull of the data to characterize it may serve as a further justification for the research in morphological learning algorithms and neural networks that go away from denoising in image processing paradigm.

References

[1] Fukunaga K., Introduction to statistical pattern recognition, Academic Press, Boston, MA 1990

[2] Graña M., J. Gallego, Associative Morphological Memories for Spectral Unmixing, In M. Verleysen, editor, *proceedings of the 11th European Symposium on Artificial Neural Networks* (ESANN 2003), d-side pub., pages 481-486, April 28-30, Bruges (Belgium), 2004.

[3] Graña M., J. Gallego, C. Hernandez, Further results on AMM for endmember induction, in IEEE 2003 Workshop on Advances in Techniques for Analysis of Remotely Sensed Data.

[4] Hopfield J.J., (1982) Neural networks and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Sciences, vol. 79, pp.2554-2558,

[5] Hyvarynën A., E. Oja A fast fixed-point algorithm for independent component analysis, Neural Comp. 9:1483-1492, 1999

[6] Hyvarynën A., J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York, 2001

[7] Ifarraguerri A., C.-I Chang, (1999) Multispectral and Hyperspectral Image Analysis with Convex Cones, IEEE Trans. Geos. Rem. Sensing, 37(2):756-770

[8] Keshava N., J.F. Mustard Spectral unimixing, IEEE Signal Proc. Mag. 19(1) pp:44-57 (2002)

[9] Ritter G. X., J. L. Diaz-de-Leon, P. Sussner. (1999) Morphological bidirectional associative memories. Neural Networks, Volume 12, pages 851-867,

[10] Ritter G. X., P. Sussner, J. L. Diaz-de-Leon. (1998) Morphological associative memories. IEEE Trans. on Neural Networks, 9(2):281-292,

[11] Ritter G.X., G. Urcid, L. Iancu, (2003) Reconstruction of patterns from moisy inputs using morphological associative memories, J. Math. Imaging and Vision 19(2):95-112

[12] Sussner P., (2001) Observations on Morphological Associative Memories and the Kernel Method, Proc. IJCNN'2001, Washington DC, July

[13] Tadjudin, S. and D. Landgrebe,(1999) Covariance Estimation with Limited Training Samples, IEEE Trans. Geos. Rem. Sensing, 37(4):2113-2118,

[14] Vapnik V.N. (1995) The Nature of Statistical Learning Theory Springer-Verlag, New York.