DEVELOPMENT OF A NEW ALGORITHM FOR LINEAR UNMIXING PROCESS IN HYPERESPECTRAL IMAGES

Raúl Guerra, Sebastián López and Roberto Sarmiento

Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria, SPAIN

ABSTRACT

This paper presents a new algorithm to perform the endmember extraction and their abundances calculation. It also presents a new method to estimate the number of endmembers and make the dimensional reduction based on the Implicitly Restarted Arnoldi's method.

Goods results have been obtained compared with some well-known algorithm such as the Vertex Component Analysis (VCA) and the NFinder for the endmembers extraction, Fully Constrained Linear Spectral Unmixing (FCLSU) for the abundances calculation and Virtual dimensionality (VD) and HySIME for the estimation of the number of endmembers. Moreover, this is achieved with independence of the amount of noise and/or the number of endmembers of the hyperspectral image under processing.

1. INTRODUCTION

Linear unmixing has rapidly become one of the most popular techniques in order to determine the content of a remotely sensed hyperspectral image. It is based on the idea that each captured pixel $r = [r^1, r^2, \dots r^{Nb}]^T$ in a hyperspectral image composed by $\boldsymbol{N_b}$ spectral bands, can be represented as a linear combination of a finite set of spectrally pure constituent spectra or endmembers, e_i, weighted by an abundance factor, a_i, that establishes the proportion of each endmember in the pixel under inspection, as follows:

$$\boldsymbol{r} = \sum_{i=1}^{p} \boldsymbol{a_i} \times \boldsymbol{e_i} + \boldsymbol{n} \qquad \textbf{(1)}$$

Where p is the total number of endmembers of the image and n represents a source of additive noise. Two physical constraints can be imposed into this linear model, namely the abundance non-negativity constraint (ANC), i.e., $a_i > 0$ for all $1 < i < p$; and the abundance sum-to-one constraint (ASC), i.e., $\sum_{i=1}^{p} \boldsymbol{a_i}=1$. At this point it is worth to mention that while partially constrained solutions imposing only the ANC have found success in the literature, the ASC is however, prone to strong criticisms because of the strong signature variability that normally characterize remotely sensed hyperspectral images. In addition, this linear mixture model assumes that secondary reflections and scattering effects can be neglected from the data collection procedure, and hence, the measured spectra can be expressed as a linear combination of the spectral signatures of materials present in the mixed pixel. If the impact of the secondary reflections or the scattering effects is relevant, more complex non-linear models can be applied but they normally demand a priori information about the geometry and physical properties of the observed objects, which results in an increase of the computational complexity of the unmixing process.

Linear unmixing process is typically divided in four stages. The first stage is a dimensionality reduction of the hyperespectral image. For this stage some well-known algorithm are the Principal Components Analysis (PCA) and the Maximun Noise Fraction (MNF). The second stage is the estimation the numbers of the endmembers. For this stage the most popular algorithms are the VD and HySIME. The third stage is the endmembers extraction. Some of the most popular algorithms for this stage are the VCA and NFinder. Once the endmembers have been extracted, the last stage is to calculate their abundance in each pixel of the image. For this stage the most used algorithm is the FCLSU.

Each of these algorithms has high operational complexity, what makes difficult their hardware implementation for real time processing. Moreover, these algorithms use to have little errors in their results, which makes the whole linear unmixing process to present a bigger error.

This paper presents some new techniques to perform the whole linear unmixing process with a smaller error and less operational complexity, in order to make it viable for real time processing of hyperespectral images.

2. ALGORITHM FOR ENDMEMBERS EXTRACTION AND ABUNDACES CALCULATION

This section aims to describe the proposed algorithm to extract the endmembers present in a hyperspectral image and calculate their abundances.

The algorithm takes as input parameters the number of endmembers to be extracted, p, and the dimensional reduced image, of p bands. The algorithm provides the endmembers and their corresponding abundances. The extracted endmembers need not necessarily to correspond with pixels in the image. Calculated abundances accomplish non-negativity constraint (ANC) and sum-to-one constraint (ASC). These restrictions are applied to every iteration performed by the algorithm for calculating the abundances. To do this, first the negative terms of abundances vectors are replaced by zeros, and then each of these vectors are divided by its norm.

The endmembers are initialized as the identity matrix. Abundances matrix is initialized as the dimensional reduced image, of p bands, applying the ANC and ASC restrictions to it. The algorithm iteratively refines these endmembers and abundances. It refines five times the abundances for each time it refines the endmembers. This sequence is repeated as much times as the iteration limit indicate.

Abundances refinement is performed maintaining fixed endmembers and minimizing function 2, employing the method of gradient descent independently to each of the pixels. The factor Y takes a high initial value, 2, and decreases linearly with the iterations to a reduced final value, 0.005.
Endmembers refinement is performed by holding fixed the abundances and minimizing the function 3, employing the method of gradient descent independently for each of the bands.

$$\boldsymbol{F_{pixel_j}} = \left(\boldsymbol{IMG_{red_j}} - \boldsymbol{e_{red}} \cdot \boldsymbol{a_j}\right)^{\boldsymbol{2}} + \boldsymbol{Y} \cdot \left(\boldsymbol{1} - \boldsymbol{a_j} \cdot\ \boldsymbol{a_j}^{\boldsymbol{t}}\right) \qquad \textbf{(2)}$$

$$\boldsymbol{F_{band_i}} = \left(\boldsymbol{IMG_{red_i}} - \boldsymbol{e_{red_i}} \cdot \boldsymbol{a}\right)^{\boldsymbol{2}} \qquad \textbf{(3)}$$

It is noteworthy that the decent gradient method calculates separately each component of each pixel or band to minimize these functions, so the algorithm is highly parallelizable. Moreover, the operations performed are just products and sums.

3. ALGORITHM FOR DIMENSIONAL REDUCTION AND ESTIMATING THE NUMBER OF ENDMEMBERS

This section aims to describe the proposed algorithm for performing dimensional reduction of the image and the estimated number of endmembers to be extracted. The proposed algorithm for this task is based on the Implicitly Restarted Arnoldi's method for calculating eigenvalues and eigenvectors.

It was noted that if this method is applied to the correlation matrix shown in function 4, and a number of eigenvalues and eigenvectors greater than the number of endmembers present in the image is calculated, the vector with the Ritz values returned by the algorithm has the same number of zeros that the number of endmembers of the image. From here, eigenvectors and eigenvalues which have Ritz values nulls are selected. The dimensional reduction is done with these eigenvectors and eigenvalues as shown in expression 7. The function 8 shows how to undo the dimensional reduction.

$$\boldsymbol{Correlation\,Matrix = IMG \cdot IMG^t / N} \quad (4)$$

$$\boldsymbol{matP = D^{-1/2} \cdot V,\ \ dimension\ p \cdot n_b} \quad (5)$$

$$\boldsymbol{matV = Vinv \cdot D^{-1/2},\ \ dimension\ n_b \cdot p} \quad (6)$$

$$\boldsymbol{IMG_{red} = matP \cdot IMG,\ \ dimension\ p \cdot N} \quad (7)$$

$$\boldsymbol{e = matU \cdot e_{red}} \quad (8)$$

Where N is the number of pixels of the image, D is a pxp diagonal matrix whose diagonal terms are the p eigenvalues obtained, V is a pxn_b matrix with the eigenvectors in rows, and $Vinv$ is the pseudoinverse of matrix V, of dimension n_bxp.

4. RESULTS

4.1. Results for estimating the number of endmembers.

4.1.1. Results with synthetic images.

Simulations were performed varying the dimension, the noise and the number of endmembers of the images. 20 simulations were conducted for each image size and each noise, varying randomly the number of endmembers between 10 and 20 in each case. For each image we estimated the number of endmembers with the proposed algorithm and with the HySIME and VD algorithms. In the VD algorithm we use 10^{-5} as value of false alarm. The average values of the results are found in Table 4.1.

Image		Percentage of correct			Average error		
Dim	SNR	TFM	HS	VD	TFM	HS	VD
150x150	40	100	75	5	0	0.25	16.5
	20	5	0	0	1.75	5.6	207.1
300x300	40	100	85	0	0	0.15	209.5
	20	35	0	0	0.7	4.7	208.5

Table 4.1 Number of endmembers. Synthetic Images.

4.1.2. Results with real images.

Simulations were performed with real images. The results are shown in Table 4.2.

Cup.	250x191x188	SNR 45	I.P.	145x145x200	SNR 35
IMG	TFM	HS	VD (10^{-1})	VD (10^{-3})	VD (10^{-5})
Cup.	13	18	31	20	16
I.P.	18	18	80	48	37

Table 4.2. Number of endmembers. Real images.

4.2. Results for endmembers extraction and abundances calculation.

4.2.1. Results with synthetic images.

Image simulations were performed with a very small number of pixels, with the aim of studying the convergence of the method to images with different noise levels and different number of endmembers. The limit of iterations performed by the algorithm is set as 1000 iterations for each endmember extracted. Moreover, for each image, endmembers are also extracted employing the VCA and NFinder algorithms and in each case abundances were calculated using the FCLSU algorithm. 5 simulations were performed for each type of image. The average results are found in Table 4.3.

Image			Average values					
			SA			SRE		
Dim	p	SNR	TFM	VCA	NF	TFM	VCA	NF
400	5	60	0,46	0,14	3,16	21,9	31,3	-1,5
625	10		0,74	0,41	5,41	20,0	26,9	-3,0
900	15		0,85	0,33	6,55	19,5	26,1	-3,4

Table 4.3. Small synthetic images.

The quality of the results obtained by the algorithm increases when increase the dimension of the image. For this reason, simulations with images of 10000 pixels we performed. 5 simulations were performed for each type of image. The average results are found in Table 4.4.

Image		Average values					
		SA			SRE		
p	SNR	TFM	VCA	NF	TFM	VCA	NF
5	60	0,42	0,01	1,17	21,8	30,0	-2,1
	20	1,20	3,10	6,70	20,2	21,6	-2,2

Table 4.4. Synthetic images of 1000 pixels.

4.2.2. Results with real images.

Simulations were conducted with the real image of Cuprite. In this simulations, the number of iterations to be performed by the algorithm is varied between 1000, 500 and 250 for each endmember extracted. The results are shown in table 4.5.

Algorithm	Average SA	RMSE
TFM 1000	8,79	$23 \cdot 10^{-5}$
TFM 500	9,23	$23 \cdot 10^{-5}$
TFM 250	9,55	$24 \cdot 10^{-5}$
VCA	9,17	$16 \cdot 10^{-5}$
NFinder	7,08	$37 \cdot 10^{-5}$

Tabla 4.5. Cuprite image. Dimension 250x191x188

5. CONCLUSIONS

A new algorithm has been developed in order to perform the endmember extraction and their abundance calculation. This algorithm is highly parallelizable and performs only simples operations. Good results have been obtained with this algorithm compared with the most popular algorithm for these tasks. Moreover, these results are consistent to the variations of the dimension of the image, the number of endmembers and the noise.

Otherwise, a new method has been proposed in order to estimate the number of endmembers and making the dimensional reduction of the hyperespectral image, based on the Implicitly Restarted Arnoldi's method for calculating eigenvalues and eigenvectors. Good results have been obtained with this algorithm compared with the most popular algorithm for these tasks. Moreover, these results are consistent to the variations of the dimension of the image, the number of endmembers and the noise.