

TOWARDS THE ACCELERATION OF SEQUENTIAL ENDMEMBER EXTRACTION ALGORITHMS FOR TIME CRITICAL APPLICATIONS

Raul Guerra, Sebastian Lopez, Gustavo M. Callico, Jose F. Lopez and Roberto Sarmiento

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

ABSTRACT

This paper presents a new strategy in order to accelerate the execution of sequential endmember extraction algorithms without compromising their performance in terms of the accuracy of the estimated endmembers. In particular, our proposal takes advantage of the correlation between pixels located in adjacent spatial positions as well as of the information provided by the dimensionality reduction step that takes place prior to the endmember extraction itself. The results demonstrate that when the proposed strategy is applied to the well-known Vertex Component Analysis (VCA) sequential algorithm, almost half of the computing time is saved with negligible variations in the quality of the endmembers extracted. Moreover, this is achieved with independence of the amount of noise and/or the number of endmembers of the hyperspectral image under processing.

Index Terms— Sequential endmember extraction algorithms, VCA, SGA, low computational cost.

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 [1]. It is based on the idea that each captured pixel $r = [r^1, r^2, \dots, r^N]^T$ in a hyperspectral image composed by N 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:

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

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., $\sum_{i=1}^p a_i = 1$. 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.

One of the most challenging tasks within the whole process of unmixing a hyperspectral image is the extraction of its endmembers. Due to this reason, a bunch of endmember extraction algorithms have been published during the last decade. These algorithms can be classified according to many different criteria, being one of them the way in which the extraction procedure is implemented. According to this, it is possible to distinguish between sequential and parallel endmember extraction algorithms [2]. Sequential algorithms are based on iteratively computing one endmember after another. Remarkable examples are the well-known Vertex Component Analysis (VCA) algorithm [3] and the Simplex Growing Algorithm (SGA) [4]. On the other hand, parallel algorithms compute all the endmembers at one time, being Winter's NFINDR [5] the most popular algorithm belonging to this category. Despite their nature, the fact is that all these algorithms incur in a formidable computational effort which becomes prohibitive for applications under real-time or near real-time constraints.

In this paper, we introduce a new strategy for accelerating the execution of sequential endmember extraction algorithms without compromising their results in terms of endmember extraction accuracy. The performance of our proposal will be demonstrated in this paper by applying it to the VCA algorithm, although it can be easily adapted for accelerating other types of sequential endmember extraction algorithms.

Our proposal is based on two issues. The first one is the inherent spatial correlation that exists between pixels in adjacent regions of remotely sensed hyperspectral images.

As it will be demonstrated in this work, this fact allows the introduction of spatial subsampling patterns which significantly speed up the execution of the endmember extraction stage without affecting its performance. The second one is motivated by the empirical fact uncovered by the inventors of the SGA which states that a pixel which has either a maximum or a minimum value in the first component of the dimensionality reduction transform applied priors the extraction is always an endmember [4]. In particular, as it will be further detailed, this has been used in the scope of this work for speeding up the first iteration of the targeted endmember extraction algorithm.

The remainder of this paper is organized as follows. Section II reviews the VCA algorithm while Section III explains how the proposed acceleration strategy has been applied to such endmember extraction algorithm. Section IV presents the most significant results obtained with synthetic and real hyperspectral images and finally, Section V outlines the conclusions extracted from this work.

2. VCA ALGORITHM

The VCA algorithm is based on the algebraic fact that the endmembers are the vertices of a simplex, being the affine transformation of a simplex also a simplex. It uses a positive cone defined by the hyperspectral data to be processed, which projected on a properly chosen hyperplane gives a simplex with vertices corresponding to the endmembers. After projecting the data onto the selected hyperplane, the VCA algorithm projects all image pixels to a random direction, obtaining the first endmember as the pixel with the largest projection. The other endmembers are identified by iteratively projecting the data onto a direction orthonormal (given by a vector named f) to the subspace spanned by the endmembers already determined. The new endmember is then selected as the pixel corresponding to the extreme projection, and the procedure is repeated until the whole set of p endmembers is found. This procedure is described in the commented pseudocode shown in Fig. 1, where $Y = [r_1, r_2, \dots, r_R]$ contains the R hyperspectral pixels to be processed, E is a $p \times p$ auxiliary matrix that stores the endmembers extracted at each iteration of the algorithm, E^+ is the pseudoinverse matrix of E , w is a $p \times 1$ zero-mean random Gaussian vector and I stands for the identity matrix. Finally, it is worth to mention that priors to the extraction itself, the VCA algorithm considers a transform operation that reduces the number of bands of the hyperspectral image to be processed from N to p with $N \gg p$.

3. PROPOSED ACCELERATION STRATEGY

This section is devoted to describe the proposed acceleration strategy as well as how it has been applied to the VCA algorithm within the scope of this work.

VCA ALGORITHM COMMENTED PSEUDOCODE

(comments begin with %)

INPUTS: p , $Y = [r_1, r_2, \dots, r_R]$ % Y is composed by R hyperspectral pixels of p frequency bands each

1: $E = [e_0 \mid 0 \mid \dots \mid 0]$; % $e_0 = [0, \dots, 0, 1]^T$ and E is a $p \times p$ auxiliary matrix

2: for $i = 1$ to p do

3: $f = ((I - EE^+)w) / (\|(I - EE^+)w\|)$; % f is a random vector orthonormal to the subspace spanned by the columns of E

4: $v = f^T Y$; % Y is projected onto the direction indicated by f

5: $\text{index} = \arg \max_{\text{index} = 1, \dots, R} |v[\text{index}]|$; % the projection extreme is found

6: $E[:, i] = Y[:, \text{index}]$; % endmembers are updated

7: end for

OUTPUT: $E = [e_1, e_2, \dots, e_p]$

Fig. 1. VCA algorithm commented pseudocode

As it was previously mentioned, it is based on the following two issues:

- i. **Information obtained after the dimensionality reduction stage.** Due to the fact mentioned in Section 1 about the properties of an hypothetical pixel which has either a maximum or a minimum value in the first component of the dimensionality reduction transform, it is possible to heavily reduce the computational burden associated to one of the iterations of a sequential endmember extraction algorithm. This has been taken into account within this work by establishing that in the first iteration of the VCA algorithm, only the pixels of higher and lower value of the most influential band of the reduced image will be projected. Hence, the first endmember will be selected from them.
- ii. **Spatial subsampling.** Due to the local spatial coherence of real hyperspectral images, to extract the endmembers from one of these images by means of a sequential algorithm, it is not necessary to consider all the pixels in all the iterations. For the case of the VCA algorithm, this means that in the first $\alpha \cdot (p-1)$ iterations of the last $(p-1)$ algorithm iterations, with $\alpha \in \mathbb{R}$ and $0 \leq \alpha \leq 1$, VCA will project only one pixel of each $(\beta + 1)$ pixels, with $\beta \in \mathbb{N}$. That is, β pixels will not be projected for each pixel projected. Thanks to this strategy, the endmembers that are spatially more frequent in the targeted image are likely to be extracted during the first $\alpha \cdot (p-1)$ iterations while the others will be extracted during the remaining $(1-\alpha) \cdot (p-1)$ iterations.

The combination of these two actions leads to the achievement of the following theoretical time savings with respect to the original VCA algorithm:

$$Time\ savings(\%) = \left(\frac{IR-1}{p} \cdot \frac{\beta}{1+\beta} + \frac{1}{p} \right) \cdot 100 \quad (2)$$

where IR represents the integer part of $\alpha \cdot p$.

According to this, high values of α and β obviously guide to high speed-up factors but they may also conduct to significant losses in terms of the accuracy of the endmembers extracted. Fortunately, as it will be demonstrated in the next section of this paper, it is possible to find pairs of α and β values able to significantly accelerate the execution of the VCA algorithm without compromising the accuracy of the endmembers extracted with independence of the amount of noise and/or the number of endmembers of the hyperspectral image under processing.

Finally, it is worth to mention that this strategy, as far as it based on discarding pixels at each algorithm's iteration, could be also easily applied to other sequential endmember extraction algorithms, such as the SGA. For this last algorithm, rather than iteratively projecting a fewer amount of pixels (as for the VCA algorithm), the proposed strategy would allow to solve less determinants per SGA's iteration.

4. RESULTS

This section discloses the most significant results achieved when applying the proposed acceleration technique to the VCA endmember extraction algorithm. For this purpose, a set of artificially generated hyperspectral images as well as the real Cuprite hyperspectral image collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) have been used.

4.1. Results with synthetic hyperspectral images

Artificial hyperspectral images represent an excellent test bench in order to demonstrate the goodness of our proposal, due to the reason that their signature endmembers, as well as their abundances, are known in advance. In particular, the hyperspectral images used in this work were generated with the *demo_VCA* software tool available in [6]. This MATLAB tool allows us to create a hyperspectral image of a spatial size defined by the user from p spectral signatures selected from the U.S. Geological Survey (USGS) digital spectral library [7] that are mixed according to abundance fractions generated with a properly tuned Dirichlet distribution. In addition, a certain amount of Gaussian noise can be added so that the generated image has a signal-to-noise ratio (SNR) value that can also be defined by the user. In particular, we have first generated four sets of 20 artificial images of 300×300 pixels, 224 spectral bands, and a SNR of 60 dBs each. The main difference between the

images from one set with respect to the other three sets is the number of endmembers, which takes the value of 5, 10, 15 and 20. The most relevant simulation results obtained with different values of α and β have been tabulated in Table I, where the performance of the VCA algorithm is compared against its accelerated version by means of the technique proposed in this work in terms of endmember extraction accuracy (measured as the average spectral angle (SA) in degrees between the estimated endmembers and the real ones) and computational cost (measured as the average time in seconds in order to extract the endmembers from the input image).

As it can be seen from Table I, all the selected α and β values allow obtaining significant time savings with respect to the original VCA algorithm without compromising its inherent endmember extraction accuracy.

p	α	β	Average SA (°)		Average time (s.)		Time savings (%)
			VCA	Proposed	VCA	Proposed	
5	0.4	1	0.15	0.17	0.51	0.38	25.85
		2	0.13	0.13	0.51	0.36	29.14
		3	0.16	0.14	0.51	0.35	30.22
	0.5	1	0.20	0.19	0.51	0.38	25.43
		2	0.17	0.15	0.52	0.37	28.59
		3	0.19	0.12	0.51	0.36	30.17
	0.6	1	0.10	0.11	0.51	0.34	33.31
		2	0.13	0.12	0.51	0.31	39.12
		3	0.11	0.13	0.52	0.29	42.38
10	0.4	1	0.49	0.52	1.74	1.33	23.86
		2	0.49	0.58	1.74	1.25	28.25
		3	0.85	0.67	1.74	1.22	30.14
	0.5	1	0.62	0.60	1.75	1.25	28.68
		2	0.79	0.63	1.74	1.14	34.40
		3	0.56	0.55	1.75	1.08	38.40
	0.6	1	0.51	0.67	1.72	1.16	32.32
		2	0.67	0.61	1.74	1.03	40.59
		3	0.57	0.64	1.74	0.97	44.40
15	0.4	1	0.84	0.85	3.80	2.94	22.59
		2	0.93	0.86	3.78	2.71	28.19
		3	1.09	0.99	3.79	2.63	30.58
	0.5	1	0.99	1.02	3.79	2.83	25.30
		2	1.19	1.00	3.79	2.56	32.52
		3	0.87	0.97	3.75	2.43	35.18
	0.6	1	1.02	1.05	3.80	2.56	32.50
		2	0.89	0.99	3.79	2.23	41.06
		3	1.02	1.15	3.79	2.07	45.32
20	0.4	1	1.57	1.60	6.55	5.12	21.83
		2	1.73	1.78	6.57	4.78	27.20
		3	1.56	1.56	6.58	4.56	30.59
	0.5	1	1.89	1.95	6.57	4.56	30.59
		2	1.60	1.52	6.56	4.35	33.61
		3	1.57	1.64	6.56	4.07	37.80
	0.6	1	1.57	1.57	6.57	4.48	31.69
		2	1.75	1.66	6.56	3.88	40.83
		3	1.60	1.59	6.56	3.58	45.37

Table I. Simulation results for different α and β values.

In order to study the influence of the level noise of the input hyperspectral image within the acceleration strategy proposed in this work, further simulations have been performed by considering two different SNR values: 40 and 80 dBs. The results obtained with the pair of α and β values that produced the best time savings in the previous simulations, i.e. $(\alpha, \beta) = (0.6, 3)$, are shown in Table II. As it is seen from the data shown in this table, the proposed method is again able to obtain endmembers with a similar level of accuracy than the one offered by the VCA algorithm but significantly reducing its computational cost. Moreover, it is worth to mention that these advantages are obtained with independence of the level of noise of the image.

SNR (dB)	p	α	β	Average SA (%)		Average time (s.)		Time savings (%)
				VCA	Proposed	VCA	Proposed	
40	5	0.6	3	0.24	0.27	0.51	0.28	43.56
	10	0.6	3	0.71	0.72	1.73	0.96	44.67
	15	0.6	3	7.17	7.28	3.81	2.06	45.93
	20	0.6	3	1.80	2.02	6.55	3.57	45.43
80	5	0.6	3	0.24	0.19	0.51	0.29	43.37
	10	0.6	3	0.65	0.65	1.79	0.95	45.53
	15	0.6	3	0.92	0.96	3.78	2.05	45.38
	20	0.6	3	1.43	1.49	6.55	3.58	45.24

Table II. Simulation results with different levels of noise.

4.2. Results with real hyperspectral images

In order to test the proposed acceleration strategy in a more realistic scenario, the AVIRIS Cuprite image has also been used in this work. This scene, is well understood from the mineralogically point of view and consists of 224 spectral bands between 0.4 and 2.5 μm . Prior to the analysis, different bands have been removed due to water absorption and low SNR resulting in a total of 188 spectral bands. Particularly, we have used the portion of the full Cuprite image which has a spatial size of 350×350 pixels. In order to determine the number of endmembers of the image, the virtual dimensionality (VD) has been estimated by means of the Noise Whitenened Harsanyi–Farrand–Chang eigenthresholding method [8] using the Neyman–Pearson test with the false-alarm probability set to 10^{-5} , resulting in a total number of 19 different pure materials.

Table III reports the average results obtained in terms of extraction accuracy, measured as the average spectral angle score obtained after comparing the USGS library spectra of alunite, buddingtonite, calcite, kaolinite and muscovite with the corresponding endmembers extracted by the different algorithms from the Cuprite scene, and in terms of processing time after ten tests. These results, together with the ones shown in Tables I and II, definitely corroborates the goodness of our proposal in terms of endmember extraction accuracy vs. computational cost.

p	α	β	Average SA (%)		Average time (s.)		Time savings (%)
			VCA	Proposed	VCA	Proposed	
19	0.6	3	7.51	8.13	8.20	4.63	43.68

Table III. Results obtained with the Cuprite dataset.

5. CONCLUSIONS

The computational burden inherent in sequential endmember extraction algorithms represents a major obstacle in order to apply them in applications under tight time constraints. To tackle this problem, we have presented a new acceleration strategy that when coupled to the VCA sequential endmember extraction algorithm allows reducing its computational cost to almost a half without sacrificing the accuracy of the endmembers extracted and with independence of the amount of noise and/or the number of endmembers of the hyperspectral image under processing. Moreover, the proposed technique could also be applied to other sequential algorithms, such as the SGA.

6. ACKNOWLEDGMENTS

This work has been carried out under the research activities of the project DREAMS (TEC2012-28666-C04-04) funded by the Spanish Government.

7. REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354-379, 2012.
- [2] Q. Du, "A New Sequential Algorithm for Hyperspectral Endmember Extraction," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no.4, pp. 695-699, 2012.
- [3] J. M. P. Nascimento and J. M. Bioucas-Dias "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898 - 910, April 2005.
- [4] C.-I. Chang , C.-C. Wu , W.-M. Liu and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2804-2819, 2006.
- [5] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266-275, 1999.
- [6] VCA algorithm MATLAB code for Windows. Available: <http://www.lx.it.pt/~bioucas/code.htm>.
- [7] USGS digital spectral library. Available: <http://speclab.cr.usgs.gov/spectral-lib.html>.
- [8] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608-619, 2004.