Sparse Unmixing-Based Content Retrieval of Hyperspectral Images on Graphics Processing Units

Jorge Sevilla, Luis Ignacio Jiménez, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Content-based image retrieval (CBIR) systems have gained significant importance in the remotely sensed hyperspectral imaging community due to the increasing availability of hyperspectral data collected from different instruments. Spectral unmixing has been a popular technique for not only interpreting hyperspectral images but also retrieving them precisely from databases based on information content. This is due to the fact that the information provided by unmixing (i.e., the spectrally pure components of the scene or *endmembers*, and their corresponding *abundance* fractions) provides a very intuitive way to describe the content of the scene in both the spectral and the spatial sense. In this letter, we present a new computationally efficient CBIR system for hyperspectral data (available online: http://hypercomp.es/repositorySparse) which uses sparse unmixing concepts to retrieve hyperspectral scenes, based on their content, from large repositories. The search is guided by a spectral library, which is used as a guide to retrieve the data in a robust and efficient way. Given the large size of libraries and the sparsity of the unmixing solutions, the incorporation of sparse unmixing to the CBIR engine brings significant advantages. To optimize its performance in computational terms, the system has been implemented in parallel by taking advantage of the computational power of commodity graphics processing units. The proposed system is validated using a collection of synthetic and real hyperspectral images, exhibiting state-of-the-art performance.**

***Index Terms*—Content-based image retrieval (CBIR), graphics processing units (GPUs), hyperspectral imaging, sparse unmixing.**

I. Introduction

OVER the last years, hyperspectral remote sensing sensors have collected a large amount of data from different locations, and there are several new missions under development [1]. Hence, the incorporation of content-based image retrieval (CBIR) [2] techniques into remote sensing data repositories offers significant advantages from the viewpoint of effectively managing, storing, and retrieving large volumes of remotely sensed data. Spectral unmixing [3], [4] has been shown to be effective in the task of not only interpreting hyperspectral images but also retrieving them from databases [5].

For instance, a spectral/spatial CBIR system for hyperspectral images was described in [6] which exploits *endmembers* (spectrally pure signatures [3], in spectral unmixing jargon) and per-pixel endmember abundance fractions. These sources of information are integrated into a dissimilarity measure that guides the search for answers to database queries [7]. In this context, each hyperspectral image can be characterized by a tuple given by the set of extracted endmembers and the set of fractional abundance maps resulting from an unmixing process conducted using three stages [8].

A similar strategy is employed in [9], which presents a parallel heterogeneous CBIR system for efficient hyperspectral image retrieval using spectral mixture analysis. This system extracts *endmembers* and estimates abundances from each image of the database and from an example image used for searching. Therefore, the query is designed to compare the spectral information of an example image for searching to that of the images from the database. In addition, the system is efficiently implemented for heterogeneous networks of computers, possibly distributed among different locations.

Recently, Sevilla and Plaza [2] presented a web-based system which manages a digital repository of hyperspectral image data, allowing for the upload and retrieval of images through a CBIR engine based on spectral unmixing concepts. The system has been efficiently implemented in parallel using graphics processing units (GPUs) [5]. The techniques used to implement the CBIR engine are based on the extraction of endmembers directly from the available scenes, which presents some challenges.

1) First, if the spatial resolution of the sensor is not high enough to separate different pure-signature classes at a macroscopic level, the resulting spectral measurement can be a composite of individual pure spectra which correspond to materials that jointly occupy a single pixel. In this case, the use of image-derived endmembers may not result in accurate fractional abundance estimations since, in this case, it is likely that such endmembers may not be completely pure in nature.
2) Second, mixed pixels can also result when distinct materials are combined into a microscopic (intimate) mixture, which is independent from the spatial resolution of the sensor. In this scenario, the use of image-derived spectral

Manuscript received May 16, 2015; revised August 23, 2015; accepted September 24, 2015. Date of publication October 19, 2015; date of current version November 11, 2015. This work was supported in part by the CEOS-SPAIN project AYA2011-29334-C02-02, funded by the Spanish Ministry of Science and Innovation. Funding from "Tools for Open Multi-Risk Assessment using Earth Observation Data" (TOLOMEO) under the Marie Curie International Research Staff Exchange Scheme (PIRSES-GA-2009) and from project FP7-REGPOT-CT-2011-284595-HOST is also acknowledged. This work was also supported in part by the computing facilities of Extremadura Research for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund. The CETA-CIEMAT belongs to the Spanish Ministry of Science and Innovation.

The authors are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, 1000 Caceres, Spain (e-mail: jorgesece@gmail.com; luijimenez@unex.es; aplaza@unex.es).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2015.2483679

TABLE III
PROCESSING TIMES (IN SECONDS) AND SPEEDUPS ACHIEVED FOR THE GPU IMPLEMENTATION OF THE SPARSE CBIR SYSTEM (USING THE FULL USGS SPECTRAL LIBRARY)

	AVIRIS Cuprite Scene			Synthetic scene		
	Init.	**Sparse**	**Total**	**Init.**	**Sparse**	**Total**
CPU	0.460 ±0.008	77447.910 ±1642.342	77448.370 ±1642.342	0.012 ±0.004	1958.481 ±196.183	1958.493 ±196.183
GPU	0.517 ±0.009	10.382 ±0.059	10.899 ±0.059	0.106 ±0.008	5.456 ±0.015	5.562 ±0.015
Speedup	7106x			352x		

TABLE IV
PROCESSING TIMES (IN SECONDS) AND SPEEDUPS ACHIEVED FOR THE GPU IMPLEMENTATION OF THE SPARSE CBIR SYSTEM (USING DIFFERENT NUMBERS OF SIGNATURES FROM THE USGS SPECTRAL LIBRARY)

Spectral Signatures		AVIRIS Cuprite Scene		
		Init.	**Sparse**	**Total**
10	**CPU**	0.271	27.163	27.503
	GPU	0.340	2.113	2.384
	Speedup	11x		
80	**CPU**	0.302	2500.540	2500.842
	GPU	0.356	2.032	2.388
	Speedup	1047x		
200	**CPU**	0.342	14871.910	14872.252
	GPU	0.396	3.687	4.083
	Speedup	3642x		
300	**CPU**	0.384	32024.070	32024.454
	GPU	0.432	5.449	5.881
	Speedup	5545x		
481	**CPU**	0.460	77447.910	77448.370
	GPU	0.517	10.582	10.899
	Speedup	7106x		

conducting the cataloging process over the AVIRIS Cuprite scene using different numbers of signatures in the USGS library. As shown by Table IV, the cataloging times (and speedups) increase notably with the number of spectral signatures used in the library. These results can be compared with those reported in [2], in which the unmixing chain used to catalog the AVIRIS Cuprite scene took 4.012 s in the same GPU architecture. These results suggest that the sparse CBIR system implemented in GPUs is more computationally effective, particularly for a very large number of spectral signatures. Moreover, the sparse CBIR system can obtain all the needed information relevant to a query in just one execution, while the unmixing-based approach in [2] needs to execute several algorithms (which implies several downloading/uploading transactions).

IV. CONCLUSIONS AND FUTURE LINES

In this letter, we have described a CBIR system which takes advantage of sparse unmixing techniques for the process of cataloging and retrieving hyperspectral scenes from large hyperspectral repositories. The use of sparse unmixing offers an important advantage: The generation of metadata and the CBIR process can be guided by a spectral library of laboratory-measured endmembers instead of endmembers extracted from the hyperspectral image. Since the endmembers in the library are acquired in ideal conditions, this allows for a more reliable process for the generation of metadata and image retrieval, circumventing problems related to the mixture problem that have traditionally prevented the extraction of completely pure spectral signatures from hyperspectral images. The sparse CBIR approach also allows for a more robust catalog since the metadata is not generated from a given image instance but from a previously available spectral library with high-quality spectral signatures measured in ideal conditions. As future extension of the system, we will include other efficient implementations of sparse unmixing algorithms and exploit other strategies for GPU optimization, such as parallel streams. In addition, we are working toward including the possibility of performing queries based on the abundance of a given material in the database for future developments of the system.

REFERENCES

[1] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[2] J. Sevilla and A. Plaza, "A new digital repository for hyperspectral imagery with unmixing-based retrieval functionality implemented on GPUs," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2267–2280, Jun. 2014.
[3] J. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[4] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
[5] J. Sevilla, S. Bernabe, and A. Plaza, "Unmixing-based content retrieval system for remotely sensed hyperspectral imagery on GPUs," *J. Supercomput.*, vol. 70, no. 2, pp. 588–599, Nov. 2014.
[6] M. Veganzones and M. Grana, "A spectral/spatial CBIR system for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 5, no. 2, pp. 488–500, Apr. 2012.
[7] M. Grana and M. Veganzones, "An endmember-based distance for content based hyperspectral image retrieval," *Pattern Recognit.*, vol. 45, no. 9, pp. 3472–3489, Sep. 2012.
[8] M. Veganzones, J. Maldonado, and M. Grana, "On content-based image retrieval systems for hyperspectral remote sensing images," in *Computational Intelligence for Remote Sensing*, ser. Studies in Computational Intelligence, M. Grana and R. J. Duro, Eds. Berlin, Germany: Springer, vol. 133, 2008, pp. 125–144.
[9] A. Plaza, J. Plaza, and A. Paz, "Parallel heterogeneous CBIR system for efficient hyperspectral image retrieval using spectral mixture analysis," *Concurrency Comput.: Pract. Exp.*, vol. 22, no. 9, pp. 1138–1159, Jun. 2010.
[10] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
[11] J. Nascimento, J. Bioucas-Dias, J. Rodriguez Alves, V. Silva, and A. Plaza, "Parallel hyperspectral unmixing on GPUs," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 666–670, Mar. 2014.
[12] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd WHISPERS*, Jun. 2010, pp. 1–4.
[13] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[14] J. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[15] M. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorithm suitable for volume etc," *IEEE Trans. Med. Imag.*, vol. 5, no. 2, pp. 61–66, Jun. 1985.
[16] R. Clark, G. Swayze, A. Gallagher, T. King, and W. Calvin, "The U.S. geological survey, digital spectral library: Version 1: 0.2 to 3.0 microns," U.S. Geological Survey Open File Report, Washington, DC, USA, 1993, pp. 93–592.