A New Digital Repository for Remotely Sensed Hyperspectral Imagery with Unmixing-Based Retrieval Functionality

Jorge Sevilla[a], Sergio Bernabé[a], Antonio J. Plaza[a], Pablo García[a]

[a]Hyperspectral Computing Laboratory
Department of Technology of Computers and Communications
University of Extremadura, Avda. de la Universidad s/n
10003 Cáceres, Spain
E-mail: jorgesece, sergiobernabe, aplaza, pablogr@unex.es

ABSTRACT

Hyperspectral imaging is concerned with the measurement, analysis, and interpretation of spectra acquired from a given scene (or specific object) at a short, medium or long distance by an airbone or satellite sensor. Over the last few years, hyperspectral image data sets have been collected for a great amount of locations over the world, using a variety of instruments for Earth observation. Despite the increasing importance of hyperspectral images in remote sensing applications, there is no common repository of hyperspectral data intended to distribute and share hyperspectral data sets in the community. Quite opposite, the hyperspectral data sets which are available for public use are spread among different storage locations and present significant heterogeneity regarding the storage format, associated meta-data (if any), or ground-truth availability. As a result, the development of a standardized hyperspectral data repository is a highly desired goal in the remote sensing community. In this paper, we take a necessary first step towards the development of a digital repository for remotely sensed hyperspectral data. The proposed system allows uploading new hyperspectral data sets along with meta-data, ground-truth and analysis results, with the ultimate goal of sharing publicly available hyperspectral images within the remote sensing community. The database has been designed in order to allow storing relevant information for the hyperspectral data available through the system, including basic image characteristics (width, height, number of bands, format) and more advanced meta-data (ground-truth information, publications in which the data has been used). The current implementation consists of a front-end to ease the management of images through a web interface, thus containing both synthetic and real hyperspectral images from two highly representative instruments, such as NASAs Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) over the Cuprite Mining District in Nevada. Most importantly, the developed system includes a spectral unmixing-based content based image retrieval (CBIR) functionality which allows searching for images on the spectral unmixing information (spectrally pure components or endmembers and their associated abundances in the scene). This information is stored as meta-data associated to each hyperspectral image instance, and then used to search and retrieve images based on information content. This paper presents the design of the system and a preliminary validation of the unmixing-based retrieval functionality using both synthetic and real hyperspectral images stored in the database.

Keywords: Hyperspectral imaging, content-based image retrieval (CBIR), spectral unmixing, endmember extraction.

1. INTRODUCTION

Content-based image retrieval (CBIR) intends to retrieve, from real data stored in a database, information that is relevant to a query.[1] This is particularly important in large data repositories, such as those available in remotely sensed hyperspectral imaging.[2] For instance, the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS)[3] is able to record the visible and near infrared spectrum of the reflected light of an area several kilometers long (depending on the duration of the flight) using hundreds of spectral bands. The resulting 'image cube' is a stack of images (see Fig. 1), in which each pixel (vector) has an associated spectral signature or 'fingerprint' that uniquely characterizes the underlying objects. The resulting data often comprises several Gigabytes per flight.

Satellite Data Compression, Communications, and Processing VIII,
edited by Bormin Huang, Antonio J. Plaza, Carole Thiebaut, Proc. of SPIE Vol. 8514,
851412 · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.929803

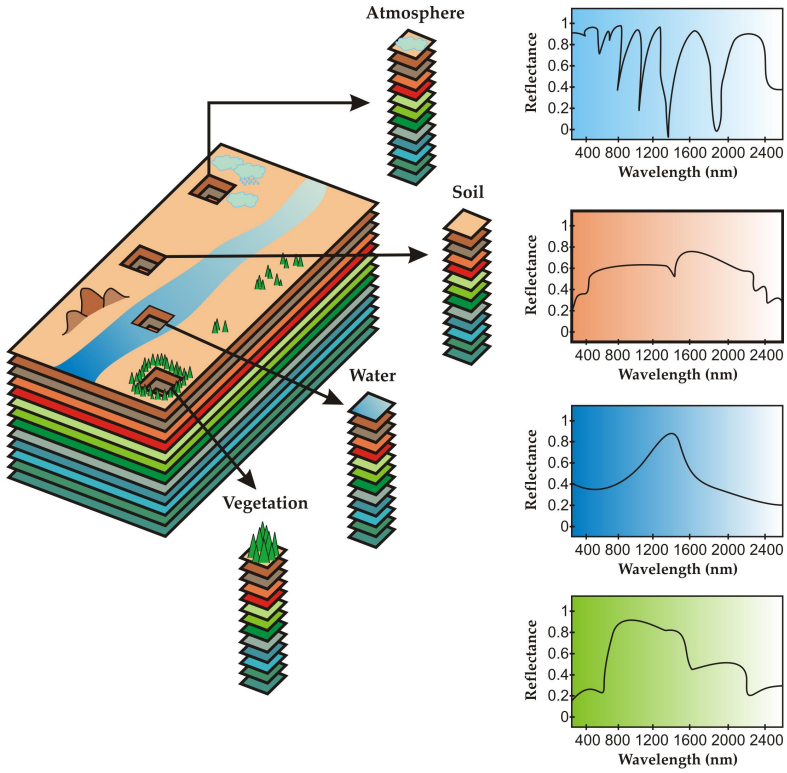


Figure 1. The concept of hyperspectral imaging.

One of the main problems involved in hyperspectral data exploitation is spectral unmixing,[4] as many of the pixels collected by imaging spectrometers such as AVIRIS are highly mixed in nature due to spatial resolution and other phenomena. For instance, it is very likely that the pixel labeled as 'vegetation' in Fig. 1 is actually composed of several types of vegetation canopies interacting at sub-pixel levels. The same comment applies to the 'soil' pixel, which may comprise different types of geological features. As a result, spectral unmixing is a very important task for hyperspectral data exploitation since the spectral signatures collected in natural environments are invariably a mixture of the pure signatures of the various materials found within the spatial extent of the ground instantaneous field view of the imaging instrument. Among several techniques designed to deal with the inherent complexity of hyperspectral images in supervised fashion,[4,5] linear spectral unmixing follows an unsupervised approach which aims at inferring pure spectral signatures, called *endmembers*, and their material fractions at each pixel of the scene.

In this paper, we describe a new digital repository for remotely sensed hyperspectral data with CBIR system which takes advantage of seminal concepts from linear spectral unmixing concepts[6] to perform effective data retrieval. Nowadays, it is estimated that a large fraction of collected hyperspectral data sets are never used but simply stored in a database, whereas these data already available in hyperspectral archives can be readily used in different application contexts or for different purposes than those that motivated the initial data collection, provided that effective CBIR mechanisms are in place to properly retrieve the data.[7] Here, we use the information provided by spectral unmixing (i.e. the spectral endmembers) as effective meta-data to develop a new CBIR system that can assist users in the task of efficiently searching hyperspectral image instances in large data repositories.

The proposed innovative approach is experimentally validated using both synthetic scenes constructed using fractals and a real hyperspectral data set collected by NASA's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the Cuprite Mining District in Nevada. Our results indicate that the proposed system can efficiently retrieve hyperspectral images from a complex image database. The proposed system is expected to increase the value of the data acquired by airborne/satellite hyperspectral imaging instruments, and to improve

the end-user services available in hyperspectral image databases. The remainder of the paper is structured as follows. Section 2 describes the considered spectral unmixing methodology used to implement the core of our CBIR system. Section 3 describes the proposed CBIR system. Section 4 assesses the performance of the system by comparing its retrieval accuracy using sinthetic and real hyperspectral images with differents noise levels. Section 5 concludes with some remarks and future research avenues.

2. SPECTRAL UNMIXING METHODOLOGY

Let us assume that a remotely sensed hyperspectral image with n bands is denoted by $\mathbf{I}$, in which a pixel of the scene is represented by a vector $\mathbf{x} = [x_1, x_2, \cdots, x_n] \in \Re^N$, where $\Re$ denotes the set of real numbers in which the pixel's spectral response x_k at sensor channels $k = 1, \ldots, n$ is included. Under the linear mixture model assumption,[8,9] each pixel vector in the original scene can be modeled using the following expression:

$$\mathbf{x} \approx \mathbf{E} + \mathbf{n} = \sum_{i=1}^{p} \mathbf{e}_i + \mathbf{n}, \tag{1}$$

where $\mathbf{E} = \{\mathbf{e}_i\}_{i=1}^{p}$ is a matrix containing p pure spectral signatures (endmembers) and $\mathbf{n}$ is a noise term. Solving the linear mixture model is identifying a collection of $\{\mathbf{e}_i\}_{i=1}^{p}$ endmembers in the image. These processing is carried out using several automatic endmembers extraction and estimation of number of endmember techniches.

2.1 Estimation of Number of Endmembers

The estimation is defined as the minimum number of spectrally distinct signal sources that characterize the hyperspectral data from the perspective view of target detection and classification. Two different methods have been used in this letter to estimate the number of endmembers in the original hyperspectral image.

VD[10] algorithm defines the number of endmembers where the signal sources are determined only on their distinct spectral properties. It first calculates the correlation eigenvalue vector and covariance eighenvalue vector. For each position of the vector, if the correlation eigenvalue is greater or equal to its corresponding covariance eigenvalue, a new endmember will be considered.

Hysime[11] algorithm first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense.

2.2 Endmembers extraction

The unmixing extraction algorithms locates the endmembers of the image. Several algorithms have been developed over the last decade for automatic or semiautomatic extraction of spectral endmembers directly from the input scene. In this work three well-known algorithms have been include.

N-FINDR[12] algorithm looks for the set of pixels with the largest possible volume by *inflating* a simplex inside the data. The procedure begins with a random initial selection of pixels. Every pixel in the image must be evaluated in order to refine the estimate of endmembers, looking for the set of pixels that maximizes the volume of the simplex defined by selected endmembers. The corresponding volume is calculated for every pixel in each endmember position by replacing that endmember and finding the resulting volume. If the replacement results in a an increase of volume, the pixel replaces the endmember. This procedure is repeated in iterative fashion until there are no more endmember replacements.

Automated Target Generation Process[13] (ATGP) altorithm uses the orthogonal subspace projection. This algorithm looks for the set of pixels with the pixels with the higher orthogonality. The procedure begin with a endmember matrix which includes just the brightest pixel. The orthogonality is calculated for every pixel in the image with the endmember matrix, the pixel with higher orthogonality is include in the endmember matrix. This procedure is repeated in iterative fashion until reaching the number of defined endmembers.

The *Vertex Component Analisis* altorithm[14] (VCA) is a improved version of ATGP algorithm. In the first step, this algorithm performs the noise characterization in the data. By means of the Hysime algorithm[11] is calculated the estimated number of endmembers and the spectrally pure regions. The ATGP[13] algorithm looks for the endmembers in the pixels located in spectrally pure regions.

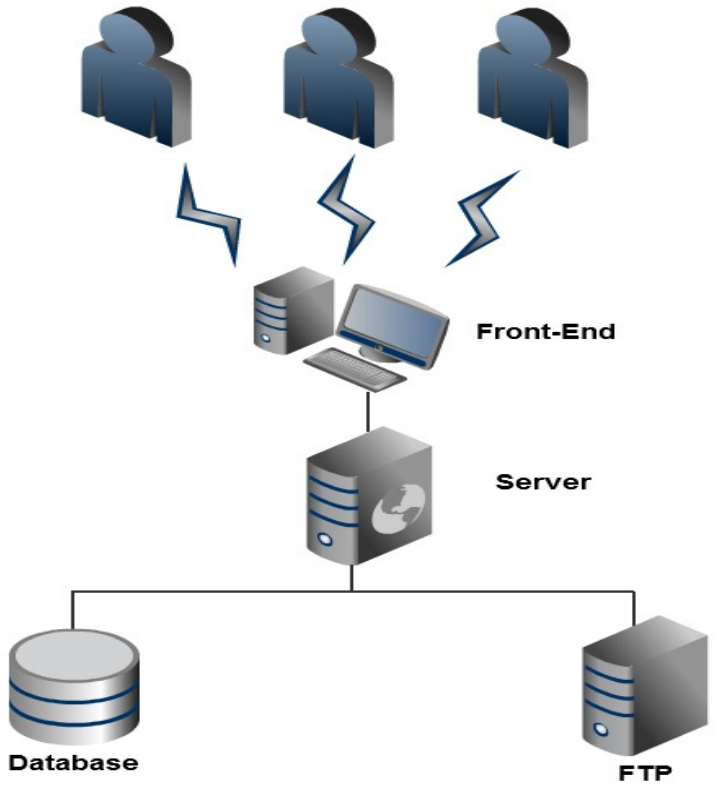


Figure 2. Architecture

3. PROPOSED CBIR SYSTEM

The proposed CBIR system for retrieval of hyperspectral imagery is based on the spectral unmixing methodology described in the previous section. A web tool is developed to ease hyperspectral images management, simple and easy to use, which stores and catalogs images, as well as content-based image retrieval. In this section, we describe the several layers that compose the architecture schema of system and the database characteristics, besides the stages involved in a standard search procedure using the proposed CBIR system from an user's point of view.

3.1 System Design

The proposed system have been designed as web service, where the users can access to the services througth web interface in agnostic way, and the server is in charge for managing metadata anda executing unmising algorithms.

Architecture

As the schema 3.1 shows, it is composed of several connected layers to provide a easy and fast access to the resources:

1. *Front-end*, provides easy access of user from anywhere to the resources.
2. *Web server*, manages the front-end request. It is the responsible for execute data queries and unmixing algorithms.
3. *Mysql server*, database managing system. Keep metadata image.
4. *FTP server*, file managing system. Keep image digital content.

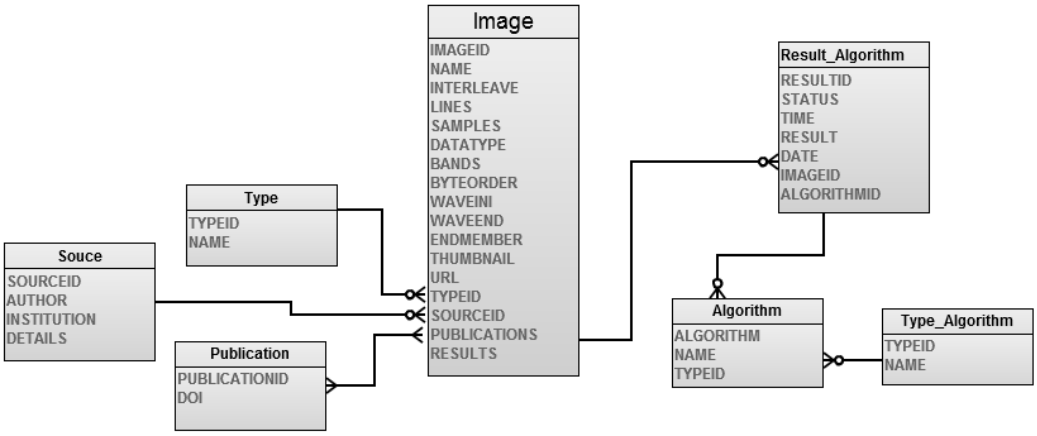


Figure 3. Database Schema

Database schema

The database 3.1 has been designed in order to allow storing relevant information for the hyperspectral data available through the system. The information includes basic image characteristics like width, height, number of bands and formats, and more advance meta-data like extracted endmembers througth cataloging, publications in which the data has been used.

3.2 Searching process

In this section, we describe the searching methodology content-based, whereby the system filters images that contain any endmember equal to a particular ground truth.

The CBIR system contrasts endmember spectral signatures of a image with several ground truth spectral signatures. Image endmembers are obtained through cataloging this image with the desired extraction algorithm (VCA[14] , ATGP[13] or N-FINDR[12]), and the number of endmembers to be extracted from the image is calculated using a specific estimation algorithm (HySime[11] or VD[10]). All cataloging results of a image are stored in the database, and just one of them is taken to contrast in the procedure, the user-selected. Such result contains the spectral signatures of the extracted endmembers. The uncataloged images are not included in the searching.

The *Spectral Angle Distance algorithm*[8] (SAD) is used to contrast the spectral signatures. This algorithm calculates spectral angle distance between two spectral signature vectors. Such angle provides a measure to compare two spectral signature vectors, useful for content-base search.

The searching procedure used in the CBIR sytem follows the steps:

1. *Inicialice* . Select ground truth spectral signatures to match and set a threshold.
2. *Signature comparison.* For each image and ground truth, calculate the spectral angle between ground truth spectral signature and all endmembers of the image so that, select the smaller angle, selecting for each endmember the matching in which the angle is smaller. The image will be matched if all angles do not exceed the threshold.
3. *Return result.* Return all matched images.

Using the tool

This visual web tool is intuitive and easy to use, agnostic to hyperspectral images treatment complexity, the steps to be followed in a standard search procedure, using the web tool from an user's point of view:

Spectral signatures	Noise levels of synthetic image						Real image
	10	30	50	70	110	No-Noise	Cuprite
KaolineiteKGa-l (wxyl)	0.21137	0.02091	0.00279	0.00183	0.00183	0.00183	0.10403
Dumortierite HS190.3B	0.28540	0.02936	0.00389	0.00295	0.00295	0.00294	0.10645
Nontronite GDS41	0.46640	0.04460	0.00638	0.00482	0.00482	0.00482	0.34333
Alunite GDS83 Na	0.22495	0.02489	0.00256	0.00085	0.00084	0.00084	0.19337
Sphene HS189.3B	0.47810	0.05846	0.00871	0.00612	0.00777	0.00777	0.21425
Pyrophyllite PYS1A fine g	0.27288	0.02426	0.00274	0.00117	0.00114	0.00114	0.21443
Halloysite NMNH106236	0.25901	0.02639	0.00379	0.00238	0.00235	0.00235	0.17150
Muscovite GDS108	0.27493	0.03174	0.00904	0.00883	0.00885	0.00885	0.14529
Kaolinite CM9	0.65911	0.02847	0.00318	0.00088	0.00087	0.00087	0.15684

Table 1. Matching results (in radians) obtained after comparing every images of the first type with each one of the pure spectral signatures. Algorithms used were VD and N-FINDR.

1. *Cataloging* hypespectral images, only the cataloged images will be included on the search. A image can be catalog several times with diferents algorithms (VCA[14] , ATGP[13] or N-FINDR[12]), thus the desired catalog result for searching should be set. Fig.4(a).
2. *Uploading spectral library* on the application. Standar spectral libraries are supported as USGS. Fig.4(b).
3. *Selecting ground truths* for search. The web tool allows to select several spectral signatures of the library, in addition to view reflectance graph of the spectral signature. Fig.4(c).
4. *Matching threshold selection*, to filter results.
5. *Executing searching*, the matched images are displayed after a few seconds, furthermore the user can check a matching log for e matched image. Fig.4(d).

4. EXPERIMENTAL RESULTS

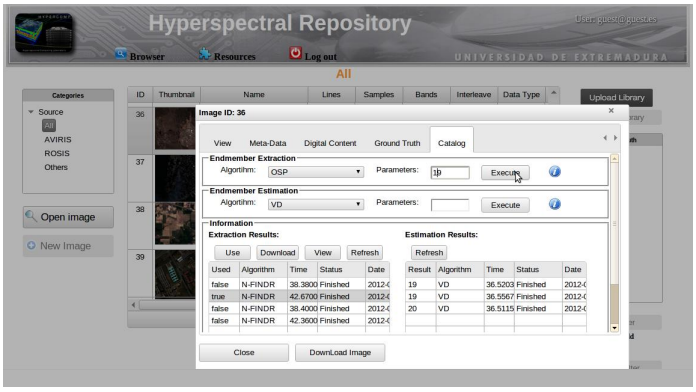
In order to illustrate the performance of our unmixing-based CBIR system, we specfically address a case study of accuracy of the matching results with images of different noise levels, using a collection of 30 synthetic hyperspectral images which are composed of known pure spectral signatures with different noise levels, spectral signatures source is the mineral spectral library from USGS Spectral Lab, version convolved to various remote sensing spectrometers such as the NASA/JPL Airborne Visual and Infra-Red Imaging Spectrometer (AVIRIS).The image collection consists in 5 types with different spectral signatures and for each type there are 6 noise levels (SNR-10,SNR30,SNR,50,SNR70,SNR110 and nonoise), 30 images in total.

In all case, the spectral resolution is of 221 narrow spectral bands between 0.4 and 2.5 micrometers, with 100×100 pixels. The experiment results are based on the images of first type, which are composed of 9 spectral signatures: *KaolineiteKGA-l(wxyl), Dumortierite HS190.3B, Nontronite GDS41, Alunite GDS83 Na, Sphene HS189.3B, Pyrophyllite PYS1A fine, Halloysite NMNH10623, Muscovite GDS108 and Kaolinite CM9.* Fig.5 correponds to no-noise image of such first group, it shows a false color composite formed using the 1682, 1107 and 655 nm channels, displayed as red, green and blue, respectively.

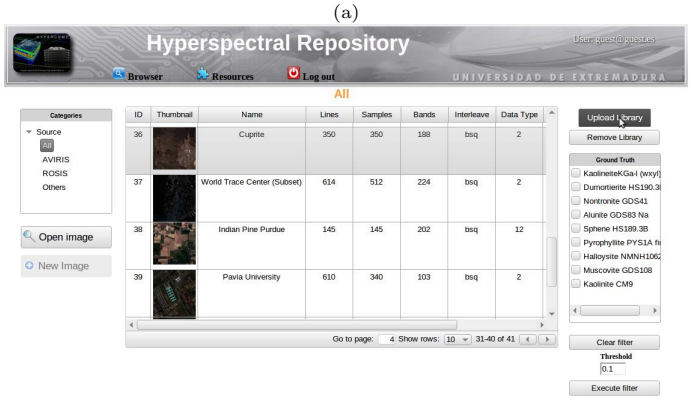
In addition, we use a real hyperspectral data set collected by NASA's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the Cuprite Mining District in Nevada, where the spectral resolution is of 188 narrow spactral bands between 0.4 and 2.5 micrometers, with 350×350 pixels, which composition is well-known and the endmembers cited in the above paragraph are among their own endmember.

Table 1 shows a preliminary validation of unmixing-based retrieval functionality unsing the first type of synthetic images and the real image.

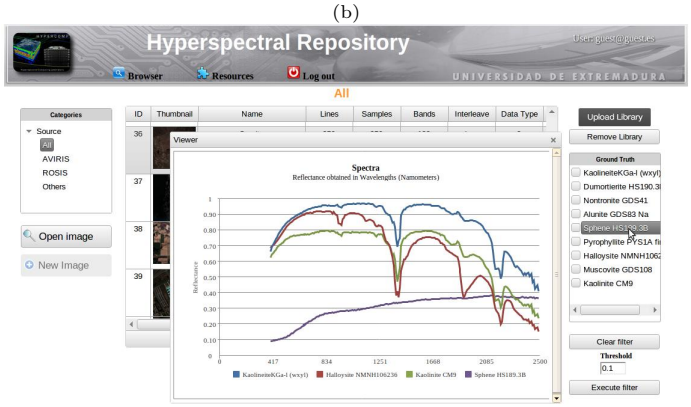
The experiment demostrates the system accuracy, wherein each cell is the minimum value obtained by comparing either image endmembers with a particular spectral signature. Obtained results with the synthetic



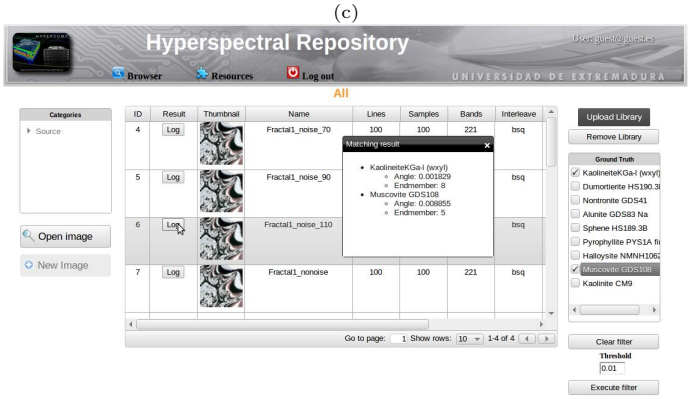
(a)



(b)



(c)



(d)

Figure 4. Web tool matching procedure workflow.

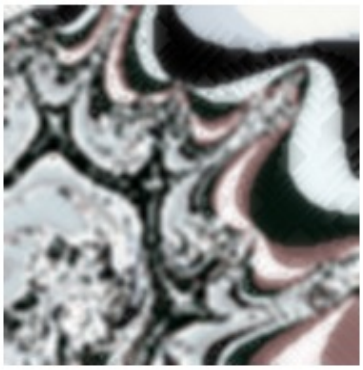


Figure 5. Synthetic hyperspectral image.

images with higher noise are great angles and those obtained from images with little noise are small angles, close to zero. SAD[8] indicates values range between 0 and 1, where 0 is the desired value, therefore the system accuracy is demostrated. On the other hand, the real image (Cuprite) results have corresponding values between SRN10 and SNR30 of synthetic images, meaning real image results are good because the real image contains noise.

5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have developed an innovative hyperspectal image repository that allows uploading and sharing images, in addition an CBIR system for hyperspectral image retrieval based on spectral unmixing. The current implmentation consists of a web application comunicating with a server, in which the users can manage data through a web interface in visual way, while server manages the repository data base and algorithm executions. The system has been implemented using well-known algorithms in the spectral unmixing community, such as HySime[11] or VD[10] for estimation the number of endmember in a given scene or VCA[14] , ATGP[13] or N-FINDR[12] for endmember extraction. Our experimental results, conducted using both synthetic scenes constructed using fractals and a real hyperspectral data set collected by NASA's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the Cuprite Mining District in Nevada, indicate that the proposed CBIR system can accurately extract hyperspectral image instances from a complex image database with sub-pixel precision and quickly enough for practical use. This can be accomplished by resorting to available parallel implementations of the considered spectral unmixing chain in different types of high performance computing architectures. As result, we belive that the proposed system can be a standardized hyperspectral data repository data intended to distribute and share hyperspectral data sets in the community. As future extension of the system, we plan develop a distributed algorithm implementation that may provide competitive advantages in terms of increased availability, time, and quality of service.

REFERENCES

1. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Fast dimensionality reduction and simple PCA," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, pp. 1349–1380, 2000.
2. C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Kluwer Academic/Plenum Publishers: New York, 2003.
3. R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment* **65**(3), pp. 227–248, 1998.
4. A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment* **113**, pp. 110–122, 2009.
5. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* **43**, pp. 1351–1362, 2005.

6. J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site," *Journal of Geophysical Research* **91**, pp. 8098–8112, 1986.
7. A. Plaza, J. Plaza, A. Paz, and S. Sanchez, "Parallel hyperspectral image and signal processing," *IEEE Signal Processing Magazine* **28**(3), pp. 119–126, 2011.
8. N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine* **19**(1), pp. 44–57, 2002.
9. D. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **39**, pp. 529–545, 2001.
10. C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **42**(3), pp. 608–619, 2004.
11. J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.* **46**(8), pp. 2435–2445, 2008.
12. M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proceedings of SPIE* **3753**, pp. 266–277, 1999.
13. J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.* **32**, pp. 779–785, 1994.
14. J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Trans. Geosci. Remote Sens.* **43**(4), pp. 898–910, 2005.