A Spectral/Spatial CBIR system for hyperspectral images

Miguel Angel Veganzones, Manuel Graña
Grupo de Inteligencia Computacional, Universidad del Pais Vasco, Spain

***Abstract*—This paper introduces a novel Content-Based Image Retrieval (CBIR) system for hyperspectral image databases using both spectral and spatial features computed following an unsupervised unmixing process which minimizes human intervention. The set of endmembers obtained from the image by an Endmember Induction Algorithm provides the image spectral features. Spatial features are computed as abundance image statistics. Both kinds of information are functionally combined into a dissimilarity measure between two hyperspectral images. This dissimilarity measure guides the search for answers to database queries. The system allows the user to retrieve hyperspectral images containing materials similar to the query image, and in a similar proportion. We provide validation results using both synthetic hyperspectral datasets and real hyperspectral data.**

***Index Terms*—Hyperspectral images, CBIR systems, linear unmixing, endmember induction, image synthesis, CBIR quality measures.**

I. Introduction

A Content Based Image Retrieval (CBIR) system [1], [2], [3] is able to retrieve the images stored in an image database using as image indexing values feature vectors extracted from the images by means of computer vision and digital image processing techniques. Conventional search in image databases is performed on the basis of metadata provided by the sensor or the human interaction. Sensor metadata may include time and position data, while human metadata refer to the human interpretation of the image contents. In CBIR approaches, image metadata is automatically generated by the machine interpretation of the image contents, with minimal human intervention. Specifically, in query-by-example CBIR, the interrogation to the database is done through the presentation of a query image, and the query's answer consists of the most similar images in the database according to some similarity measure defined on the image feature space.

The increasing amount of Earth Observation data provided by hyperspectral sensors, motivates research in some technological problems raised by the sheer scale of the amount of available data. Among them, searching through these huge databases using CBIR techniques has not been properly addressed for the case of hyperspectral images. Approaches to CBIR in remote sensing image databases reported in the literature are focused on panchromatic images, SAR data or low dimensional multispectral images such as LANDSAT [4], [5], [6]. There are few works in the literature dealing explicitly with the spectral information to guide the search [7], [8], [9], [10]. These works define the image features as the endmembers induced from the hyperspectral data by some Endmember Induction Algorithm (EIA). One inconvenience of these CBIR systems is that they can not discriminate among images with the same induced endmembers but very different spatial distributions. In this paper, we propose a Spectral-Spatial feature extraction process, that makes this discrimination possible, and we provide an evaluation of its effectiveness on synthetic and real hyperspectral data.

The contents of the paper are the following. Section II provides a description of the Spectral-Spatial CBIR system. Section III describes the methodological issues of its validation. Section IV gives the validation results using synthetic hyperspectral images. Section V gives validation results using a real hyperspectral dataset. Section VI gives our conclusions and directions for further work.

II. Spectral-Spatial CBIR system

We first describe the Spectral-Spatial CBIR system's architecture, next we detail the Spectral-Spatial feature characterization, and the Spectral-Spatial dissimilarity function over which the proposed CBIR system is built on.

A. System's architecture

Figure 1 shows the Spectral-Spatial CBIR system scheme. The core of the CBIR system is the Spectral-Spatial dissimilarity function between two hyperspectral images by means of their spectral and spatial features. The system interacts with a feature database where the Spectral-Spatial features of the images are stored. These features have been previously extracted by offline application of an Endmember Induction Algorithm (EIA) and a spectral unmixing method using the endmembers extracted by the EIA from the image. System interrogation is done using a query example approach. First, the query example is processed to extract its Spectral-Spatial features and second, it is compared to the images in the database using the Spectral-Spatial dissimilarity measure. A ranking of the images in the database is elaborated by ascending order of dissimilarity to the query. Finally, the system returns the k images in the database corresponding to the first k ranking positions, where k is known as the query's *scope*.

B. Spectral-Spatial feature extraction process

Hyperspectral images in the database are indexed by low level feature vectors, that are extracted and stored offline. To that effect we assume the linear mixing model [11], where a hyperspectral image is the result of the linear combination

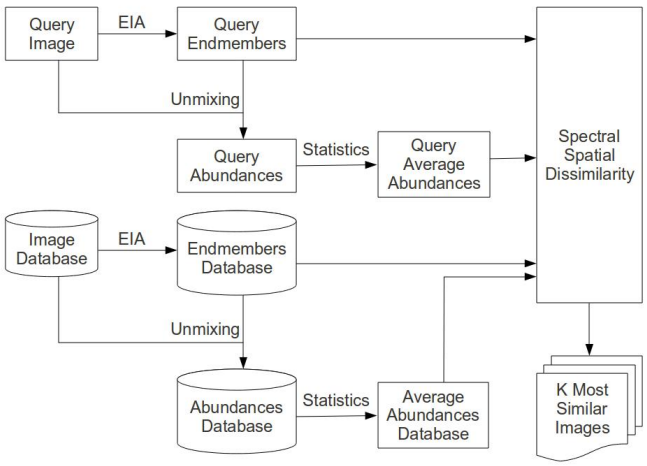

Figure 1. Spectral-Spatial CBIR system's schema

of the pure spectral signatures of ground components, named endmembers, with a fractional abundance matrix.

Let $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m]$ be the pure endmember signatures (normally corresponding to macroscopic objects in scene, such as water, soil, vegetation, ...) where each $\mathbf{e}_i \in \mathbb{R}^q$ is a q-dimensional vector. Then, the hyperspectral signature $\mathbf{r}$ at each pixel in the image is defined by the expression

$$\mathbf{r} = \mathbf{s} + \mathbf{n} = \sum_{i=1}^{m} \mathbf{e}_i \phi_i + \mathbf{n}, \tag{1}$$

where the hyperspectral signature $\mathbf{r}$ is formed by the sum of the pixel's signal $\mathbf{s}$ and an independent additive noise component $\mathbf{n}$; and, ϕ is the m-dimensional vector of fractional abundances at the given pixel subject to constraints: $\phi_i \geq 0$, $\forall i = 1, \ldots, m$, and $\sum_{i=1}^{m} \phi_i = 1$. This equation can be extended to the full image as $\mathbf{H} = \mathbf{E}\mathbf{\Phi} + \mathbf{n}$, where $\mathbf{H}$ is the hyperspectral image and $\mathbf{\Phi}$ is a matrix of fractional abundances.

We characterize an hyperspectral image H_α by a tuple (E_α, Φ_α), where $E_\alpha = \left\{\mathbf{e}_1^\alpha, \mathbf{e}_2^\alpha, \ldots, \mathbf{e}_{m_\alpha}^\alpha\right\}$ is the set of m_α induced endmembers from the α-th image; and, $\Phi_\alpha = \left\{\phi_1^\alpha, \phi_2^\alpha, \ldots, \phi_{m_\alpha}^\alpha\right\}$ is the set of fractional abundance maps resulting from the unmixing process, where each ϕ_i^α represents the spatial distribution of the induced endmember $\mathbf{e}_i^\alpha$, $i = 1, \ldots, m_\alpha$, in the image. We refer to the tuple (E_α, Φ_α) as the *Spectral-Spatial features* of the hyperspectral image H_α.

To implement this approach, an EIA is first used to induce the spectral signatures (the endmembers) of the image; and then, an unmixing method calculates from the image the spatial distribution of each endmember (the fractional abundances).

C. Spectral-Spatial dissimilarity measure

The Spectral-Spatial CBIR system of figure 1 uses a dissimilarity measure between the Spectral-Spatial features of hyperspectral images described above. The aim is to declare images sharing some of their constituent materials as similars. We detect this similarity through the similarity between image spectra and their relative abundance proportions. The spectral-spatial dissimilarity is a version of the Integrated Region Matching (IRM) dissimilarity function [12] used for region matching-based image retrieval. The overall IRM dissimilarity between two images depends on two aspects: the similarity between each region pair from the two images, and the significance of each region pair matching. To build our Spectral-Spatial dissimilarity measure, we establish a correspondence between IRM regions and spectral signatures. Region pair matching corresponds to the matching of pairs of spectral signatures built on the distance between each pair of the endmembers from both images. Significance corresponds to endmember average abundances.

Let $E_\alpha = \left\{\mathbf{e}_1^\alpha, \mathbf{e}_2^\alpha, \ldots, \mathbf{e}_{m_\alpha}^\alpha\right\}$ be the set of endmembers induced from the hyperspectral image H_α in the database, where m_α is the number of induced endmembers from the α-th image; and $\bar{\Phi}_\alpha = \left\{\bar{\phi}_1^\alpha, \bar{\phi}_2^\alpha, \ldots, \bar{\phi}_{m_\alpha}^\alpha\right\}$ the average abundances of H_α, where $\bar{\phi}_i^\alpha = \frac{1}{N}\sum_{j=1}^{N} \phi_i^\alpha(j)$, $i = 1, \ldots, m_\alpha$, being $\phi_i^\alpha(j)$ the i-th endmember fractional abundance for a pixel j, and N the number of pixels in the image. Given two images, H_α and H_β, we first compute the Spectral Distance Matrix, $D_{\alpha,\beta}$, as

$$D_{\alpha,\beta} = [d_{ij};\ i = 1, \ldots, m_\alpha;\ j = 1, \ldots, m_\beta], \tag{2}$$

whose elements d_{ij} are the distances between the endmembers $\mathbf{e}_i^\alpha, \mathbf{e}_j^\beta \in \mathbb{R}^q$ of each image. Any spectral distance function $d : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}^+$ can be used, for instance, the Euclidean distance or the angular pseudo-distance (SAM) [13].

Then, we define the Spectral-Spatial dissimilarity measure, $s(H_\alpha, H_\beta)$, as

$$s(H_\alpha, H_\beta) = \sum_{i,j} r_{ij} d_{ij}, \tag{3}$$

where d_{ij} is the spectral distance between endmembers $\mathbf{e}_i^\alpha$ and $\mathbf{e}_j^\beta$, and r_{ij} is the significance associated to d_{ij}.

The formulation of the measure needs a definition for the significance matrix $R_{\alpha,\beta} = [r_{ij};\ i = 1, \ldots, m_\alpha;\ j = 1, \ldots, m_\beta]$. We follow the *most similar highest priority* (MSHP) principle [12], making use of the average abundances $\bar{\Phi}_\alpha$ and $\bar{\Phi}_\beta$. The average abundances represent "significance credits" assigned to the endmember spectral distances by Algorithm 1.

The algorithm for the assignment of significance credits starts by initializing the set of all possible endmember pairs, $\mathcal{M} = \{(i,j) :\ i = 1, \ldots, m_\alpha;\ j = 1, \ldots, m_\beta\}$, where m_α and m_β indicate the number of endmembers for images H_α and H_β respectively, and the set of previously selected endmember pairs, $\mathcal{L} = \{\}$ (Steps 1, 2). In each subsequent iteration, the algorithm first selects the pair of endmembers $(i,j) :\ i = 1, \ldots, m_\alpha;\ j = 1, \ldots, m_\beta$, with minimum spectral distance, d_{ij}, from the set of available pairs, $(i,j) \in \mathcal{M} - \mathcal{L}$, (Step 3). Let (i', j') denote the selected pair. Second, the value of the minimum of the average fractional abundances is assigned as the pair's corresponding significance, $r_{i'j'} = \min\left\{\bar{\phi}_{i'}^\alpha, \bar{\phi}_{j'}^\beta\right\}$ (Step 4). Notice that the average abundances are always equal to or greater than zero. If $\bar{\phi}_{i'}^\alpha < \bar{\phi}_{j'}^\beta$, then the elements of significance matrix row i' are set to zero;

Algorithm 1 Significance credits assignment algorithm.

1) Set $\mathcal{L} = \{\}$.
2) Denote $\mathcal{M} = \{(i,j) : \ i = 1,\ldots,m_\alpha;\ j = 1,\ldots,m_\beta\}$.
3) Choose the minimum d_{ij} for $(i,j) \in \mathcal{M} - \mathcal{L}$. Label the corresponding (i,j) as (i',j').
4) $r_{i'j'} = \min\left\{\bar{\phi}^\alpha_{i'}, \bar{\phi}^\beta_{j'}\right\}$.
5) If $\bar{\phi}^\alpha_{i'} < \bar{\phi}^\beta_{j'}$, set $r_{i'j} = 0$, for all $j \neq j'$; otherwise, set $r_{ij'} = 0$, for all $i \neq i'$.
6) If $\bar{\phi}^\alpha_{i'} < \bar{\phi}^\beta_{j'}$, set $\bar{\phi}^\alpha_{i'} = 0$ and $\bar{\phi}^\beta_{j'} = \bar{\phi}^\beta_{j'} - \bar{\phi}^\alpha_{i'}$; otherwise, set $\bar{\phi}^\beta_{j'} = 0$ and $\bar{\phi}^\alpha_{i'} = \bar{\phi}^\alpha_{i'} - \bar{\phi}^\beta_{j'}$.
7) $\mathcal{L} = \mathcal{L} + \{(i',j')\}$.
8) If $\sum_{i=1}^{p_\alpha} \bar{\phi}^\alpha_i > 0$ and $\sum_{j=1}^{p_\beta} \bar{\phi}^\beta_j > 0$, go to step 3; otherwise, stop.

otherwise, the elements of significance matrix column j' are set to zero (Step 5). Then, the pool of significance credits is reduced. If $\bar{\phi}^\alpha_{i'} < \bar{\phi}^\beta_{j'}$, then set $\bar{\phi}^\alpha_{i'} = 0$ and $\bar{\phi}^\beta_{j'} = \bar{\phi}^\beta_{j'} - \bar{\phi}^\alpha_{i'}$; otherwise set $\bar{\phi}^\beta_{j'} = 0$ and $\bar{\phi}^\alpha_{i'} = \bar{\phi}^\alpha_{i'} - \bar{\phi}^\beta_{j'}$ (Step 6). Finally, (i',j') is added to the set of previously selected pairs, $\mathcal{L}$ (Step 7). When the stopping condition, $\sum_{i=1}^{m_\alpha} \bar{\phi}^\alpha_i = 0$ or $\sum_{j=1}^{m_\beta} \bar{\phi}^\beta_j = 0$, is met the algorithm ends; otherwise a new iteration starts (Step 8).

III. Spectral-Spatial CBIR system validation

There are two key elements of the validation process. First, the performance measures used to compare the diverse instances of the system obtained by different parametrizations. Second, the strategy followed to obtain the reference measures. Namely, how the ground truth references are obtained and used to compute the benchmarking performance measures.

A. CBIR performance measures

Evaluation metrics from information retrieval field have been adopted to evaluate CBIR systems quality. The two most used evaluation measures are *precision* and *recall* [1], [5]. Precision, p, is the fraction of the returned images that are relevant to the query. Recall, q, is the fraction of returned relevant images respect to the total number of relevant images in the database according to *a priori* knowledge. If we denote T the set of returned images and R the set of all the images relevant to the query, then

$$p = \frac{|T \cap R|}{|T|} \tag{4}$$

$$r = \frac{|T \cap R|}{|R|} \tag{5}$$

Precision and recall follow inverse trends when considered as functions of the scope of the query. Precision falls while recall increases as the scope increases. To evaluate the overall performance of a CBIR system, the Average Precision and Average Recall are calculated over all the query images in the database. For a query of scope k, these are defined as:

$$P_k = \frac{1}{N}\sum_{\alpha=1}^{N} P_k(H_\alpha) \tag{6}$$

and

$$R_k = \frac{1}{N}\sum_{\alpha=1}^{N} R_k(H_\alpha). \tag{7}$$

The Normalized Rank [14] is a performance measure used to summarize system performance into an scalar value. The normalized rank for a given image ranking Ω_α, denoted as $\text{Rank}\,(H_\alpha)$, is defined as:

$$\text{Rank}\,(H_\alpha) = \frac{1}{NN_\alpha}\left(\sum_{i=1}^{N_\alpha}\Omega^i_\alpha - \frac{N_\alpha\,(N_\alpha - 1)}{2}\right), \tag{8}$$

where N is the number of images in the dataset, N_α is the number of relevant images for the query H_α, and Ω^i_α is the rank at which the i-th image is retrieved. This measure is 0 for perfect performance, and approaches 1 as performance worsens, being 0.5 equivalent to a random retrieval. The average normalized rank, ANR, for the full dataset is given by:

$$ANR = \frac{1}{N}\sum_{\alpha=1}^{N} Rank\,(H_\alpha)\,. \tag{9}$$

B. Validation methodology

We perform separate validation experiments using a synthetic dataset and real data. In both cases we follow a similar methodology, explained below. However, each dataset serves to validate different aspects of the Spectral-Spatial CBIR system. We make use of a synthetic dataset to test the system robustness against changes in its internal components, that is, the selected spectral distance and methods used to extract the Spectral-Spatial features; as well as the system robustness against noisy data. In the experiments with real data we test the applicability of the proposed system in a real scenario.

The general methodology followed in all the experiments has two main phases. First, we perform offline the Spectral-Spatial feature extraction of the hyperspectral images in the given dataset. For each image, we apply independently either the ILSIA [15], the N-FINDER [16] or the FIPPI [17] endmember induction algorithms (EIAs) to induce the set of endmembers. Therefore, we have three competing spectral features of each image. A brief review of these three EIAs is given in an appendix. ILSIA is fully automatic and do not require any input parameter. N-FINDR and FIPPI requires as input the estimation of the number of endmembers in the image, m. For the latter EIAs, m is estimated by means of the Harsanyi-Farrand-Chang (HFC) virtual dimensionality method [18] setting the false alarm rate to $\alpha = 10^{-5}$. The number of endmembers induced by ILSIA and FIPPI algorithms may be different from the HFC method estimation. For each set of induced endmembers, we calculate their respective spatial fractional abundances using the Full Constrained Least Squares Unmixing (FCLSU) method [19].

In the second phase, for each hyperspectral image H_α in a dataset we calculate the dissimilarity measure of equation (3) between H_α and each of the remaining images in the dataset. These dissimilarities are represented as a vector $\mathbf{s}_\alpha = [s_{\alpha 1},\ldots,s_{\alpha N}]$, where N is the number of images in the dataset and $s_{\alpha,\beta}$ is the dissimilarity between

the images H_α and H_β, with $\alpha, \beta = 1, \ldots, N$. We can define the ranking of the dataset relative to the query image, $\Omega_\alpha = [\omega_{\alpha,p} \in \{1, \ldots, N\}; p = 1, \ldots, N]$, as the set of image indexes ordered according to increasing values of their corresponding entries in the dissimilarity vector $\mathbf{s}_\alpha$. That is, we sort in increasing order the components of $\mathbf{s}_\alpha$, and the resulting shuffled image indexes constitute Ω_α, so that $s_{\alpha,\omega_{\alpha,p}} \leq s_{\alpha,\omega_{\alpha,p+1}}$. The ranking Ω_α can be computed either on the ground truth information or on the induced endmembers and estimated abundances.

Finally, we estimate the Spectral-Spatial CBIR system performance measures, average precision, average recall and average normalized rank, as follows. For each hyperspectral image H_α, a query $Q_k(H_\alpha)$ is formulated returning the k most similar (less dissimilar) images H_β in the dataset relative to the image H_α, where k is the scope of the query and takes values in the range $1 \leq k \leq N$. The set of returned images $T_k(H_\alpha)$ and the set of relevant images $V_k(H_\alpha)$ for a query $Q_k(H_\alpha)$ used to calculate the CBIR performance measures can be obtained from the rankings computed from the induced endmembers and the ground truth, respectively.

IV. Validation using synthetic datasets

We use a collection of synthetic hyperspectral images to test the robustness of the proposed Spectral-Spatial CBIR system to variations on the selected spectral distance, and the applied endmember induction algorithms. We also test the influence in the system performance of random noise added to the synthetic images. The section starts by describing how we construct the synthetic hyperspectral images. Next, we comment on the specific methodological details, and, finally, we give the results of the experiments.

A. Synthetic dataset

The synthetic hyperspectral images are generated as linear mixtures of a set of spectra (the groundtruth endmembers) according to synthesized fractional abundance coefficients. The generation of the abundance coefficients is a spatial process performed independently for each desired endmember, which does not ensure for each pixel in the image the normalization properties required by the linear mixing model. To ensure them, the resulting collection of abundance images for all endmembers are processed imposing normalization conditions independently for each pixel.

The pool of ground-truth endmembers is a subset of ten spectral signatures selected from the USGS spectral library[1]. The synthetic groundtruth multidimensional fractional abundance images are generated in a two-step procedure. First, we simulate the abundance images corresponding to each endmember as 2D products of 1D Legendre polynomials with randomly generated parameters. Second, to ensure that there are regions of almost pure endmembers, we select for each pixel the abundance coefficient with the greatest value and we normalize the remaining to ensure that the abundance coefficients sum up to one. Figure 2 gives an instance of synthetic fractional abundance images generated using random Legendre polynomials to build a hyperspectral image composed of three endmembers.

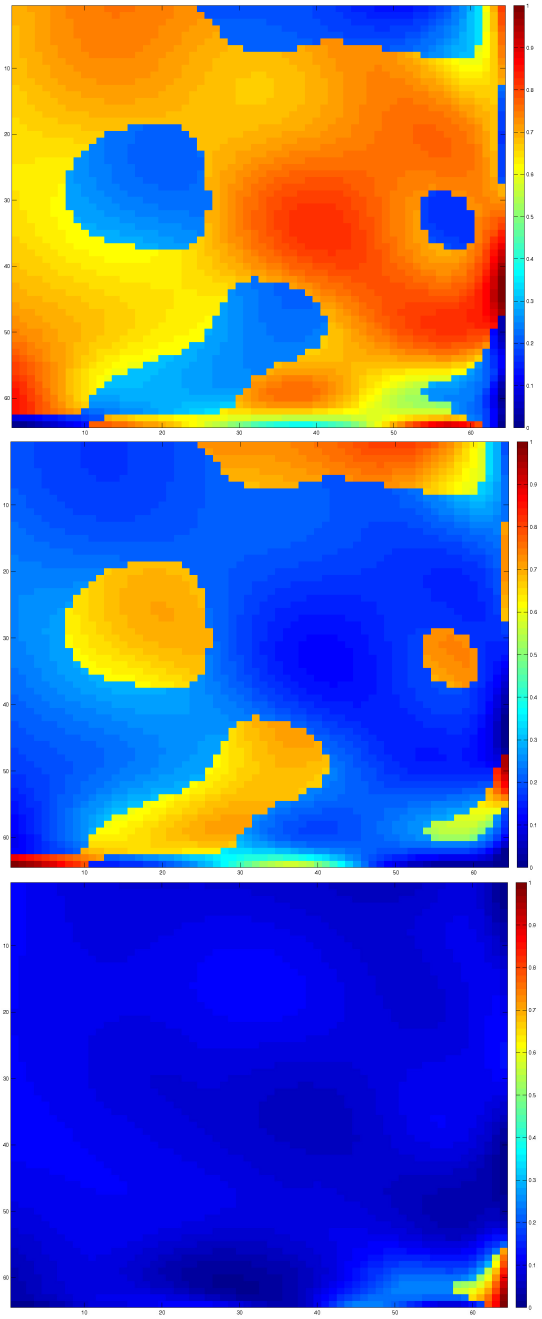

Figure 2. An instance of synthetic fractional abundance images generated using random Legendre polynomials.

We have synthesized a total of 1000 hyperspectral images. All the synthesized hyperspectral images have 269 spectral bands *per* pixel, and their size is 64×64 pixels. Each synthetic image is built with a number of ground-truth endmembers, $m = \{2, 3, 4, 5\}$, randomly selected from the pool of available ground-truth endmembers. We have generated for the experiments in this paper, 250 images of each number of ground-truth endmembers.

In addition to this dataset, hereafter named as the clean

[1] http://speclab.cr.usgs.gov/spectral-lib.html

dataset, three more datasets have been built by adding random Gaussian noise to each of the clean dataset images, resulting in noise images with signal to noise ratios (SNR) 30dB, 40dB and 50dB. We denote clean dataset as D_o, and the noisy datasets as D_{30dB}, D_{40dB} and D_{50dB}.

B. Methodology

We have performed independent experiments[2] over each of the four hyperspectral synthetic datasets, following the methodology explained in section III-B. Let us distinguish between $\mathbf{s}_\alpha^{\text{GT}}$, the vector of dissimilarities computed using the known ground truth endmembers and fractional abundances used to synthesize the images, and $\mathbf{s}_\alpha^{\text{IND}}$, the vector of dissimilarities computed using the endmembers induced by one of the EIAs (either ILSIA, N-FINDER or FIPPI) and their estimated abundances. We distinguish as well rankings $\Omega_\alpha^{\text{GT}}$ and $\Omega_\alpha^{\text{IND}}$ corresponding to the ground truth and induced dissimilarities, respectively.

The set of returned images $T_k(H_\alpha)$ and the set of relevant images $V_k(H_\alpha)$ for a query $Q_k(H_\alpha)$ are defined as follows:

$$T_k(H_\alpha) = \Omega_{\alpha,k}^{\text{IND}} = \left[\omega_{\alpha,p}^{\text{IND}} \text{ s.t. } s_{\alpha,\omega_{\alpha,p}^{\text{IND}}} \leq s_{\alpha,\omega_{\alpha,k}^{\text{IND}}}\right], \quad (10)$$

$$V_k(H_\alpha) = \Omega_{\alpha}^{\text{GT}} = \left[\omega_{\alpha,p}^{\text{GT}} \text{ s.t. } s_{\alpha,\omega_{\alpha,p}^{\text{GT}}} \leq t\right], \quad (11)$$

where $t = \bar{s}_\alpha^{\text{GT}} - 2\sigma_{\mathbf{s}_\alpha^{\text{GT}}}$, and $\bar{s}_\alpha^{\text{GT}}$ and $\sigma_{\mathbf{s}_\alpha^{\text{GT}}}$ are respectively the mean and standard deviation of $\mathbf{s}_\alpha^{\text{GT}}$. This definition allows for the inclusion in the query's answer of images whose dissimilarity is equal to the maximum one, thus allowing that the cardinality of both returned and relevant sets may be bigger than k. Now $T_k(H_\alpha)$ and $V_k(H_\alpha)$ can be used to calculate the average precision and recall measures of the system, as well as the average normalized rank.

C. Performance results

Figures 3, 4, 5 and 6 show the plots of the precision-recall curves for the noise free dataset and the 30dB, 40dB and 50dB datasets respectively. It can be appreciated that for low levels of noise the performance is similar to the noise free case. In general, the plots show a rather high insensitivity to the choice of EIA and individual endmember spectral distance, because corresponding curves are not very different, except in the limit of low recall values where a clear sensitivity to the individual endmember distance is made apparent. The Euclidean distance systematically improves the SAM distance, giving higher precision at the same recall value. Increasing recall value range reverses the picture, so that the SAM distance improves systematically over the Euclidean distance in the noise free case. One effect of the noise is the cancellation of this effect. For the highest noise (30dB) the SAM never improves the Euclidean distance. Other effect of the noise is the disappearance of the effect of the EIA chosen. In the noise free data set, the ILSIA and N-FINDR show a small improvement on the FIPPI, however these relations change according to the noise level, disappearing at the highest noise level.

[2]The Matlab code for the hyperspectral image synthesis and endmember induction is publicly available from http://www.ehu.es/ccwintco/index.php/GIC-source-code-free-libre

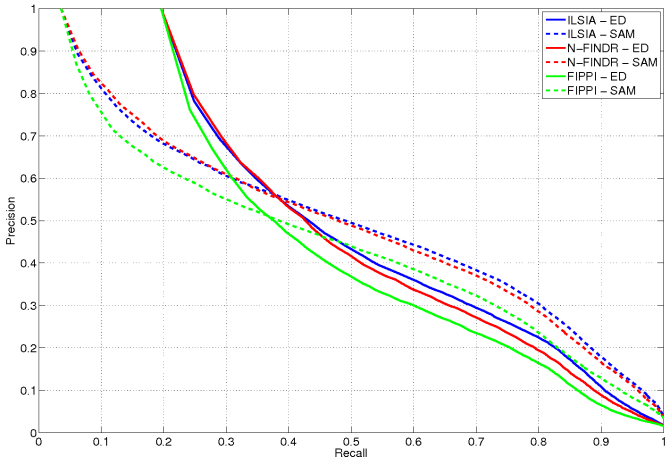

Figure 3. Precision-recall curves for D_o synthetic dataset.

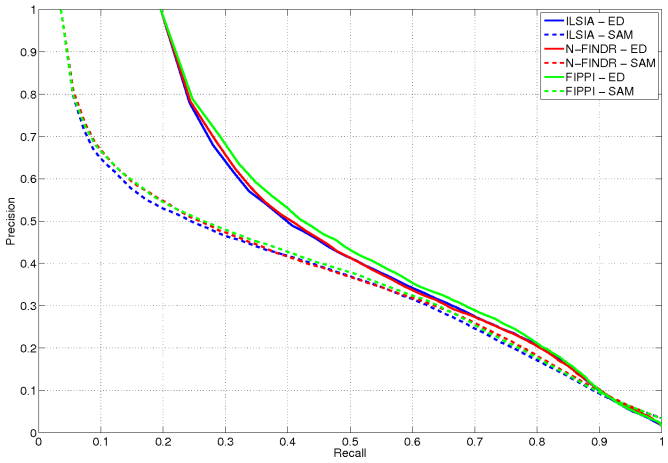

Figure 4. Precision-recall curves for D_{30dB} synthetic dataset.

Figure 5 gives the Averaged Normalized Rank values obtained. The results confirm the conclusions from the figures. The SAM distance is most affected by noise, showing the worse results for the noisiest data. The Euclidean distance is much more robust relative to noise. Comparing the effect of the EIA chosen, the ILSIA gives the best result in the noise free data, and the differences between algorithms disappear for the noisiest data.

	Averaged Normalized Rank (ANR)					
	ILSIA		N-FINDR		FIPPI	
Dataset	ED	SAM	ED	SAM	ED	SAM
D_o	**0.043**	0.053	0.050	0.058	0.064	0.074
D_{30dB}	0.045	0.101	**0.043**	0.097	**0.043**	0.099
D_{40dB}	0.042	0.057	**0.036**	0.048	0.037	0.052
D_{50dB}	0.042	0.053	**0.038**	0.045	0.041	0.051

Table I
ANR RESULTS FOR SYNTHETIC DATASETS.

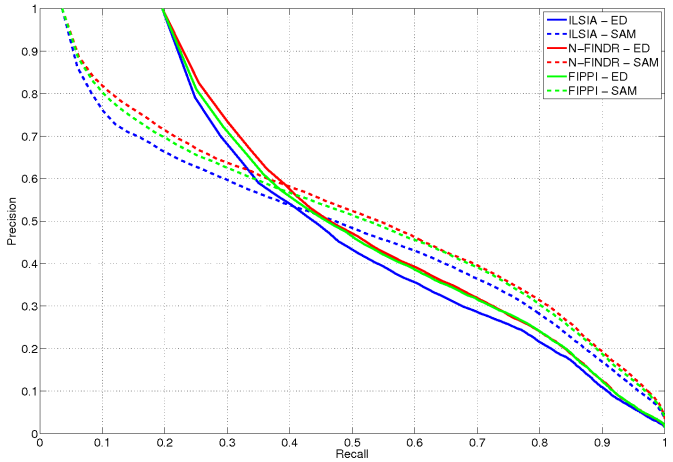

Figure 5. Precision-recall curves for D_{40dB} synthetic dataset.

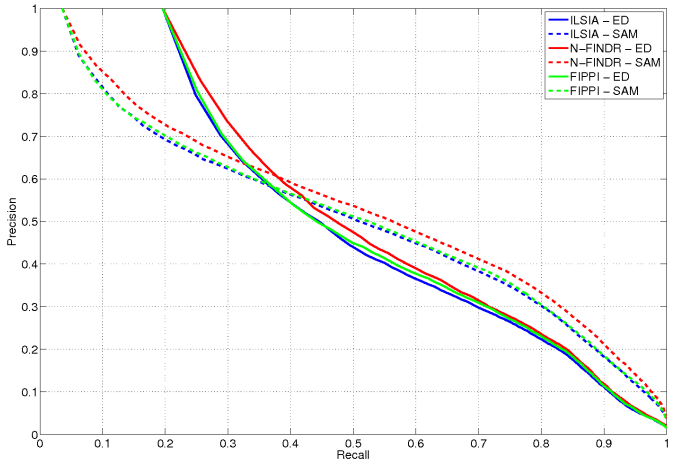

Figure 6. Precision-recall curves for D_{50dB} synthetic dataset.

V. Validation using real data

Here we test the proposed Spectral-Spatial CBIR system over a dataset of real hyperspectral images to validate the system usage in a real scenario. We first introduce the hyperspectral dataset used in the experiments, we follow explaining the specific methodological aspects of the experiments and finally, we give the results.

A. HyMAP dataset

The hyperspectral HyMAP data was made available from HyVista Corp. and German Aerospace Center's (DLR) optical Airbone Remote Sensing and Calibration Facility service[3]. The sensed scene corresponds to the radiance captured by the sensor in a flight line over the facilities of the DLR center in Oberpfaffenhofen (Germany) and its surroundings, mostly fields, forests and small towns. Figure 7 shows the scene captured by the HyMAP sensor. The data cube has 2878 lines, 512 samples and 125 bands; and the pixel values are represented by 2-bytes signed integers.

[3]http://www.OpAiRS.aero

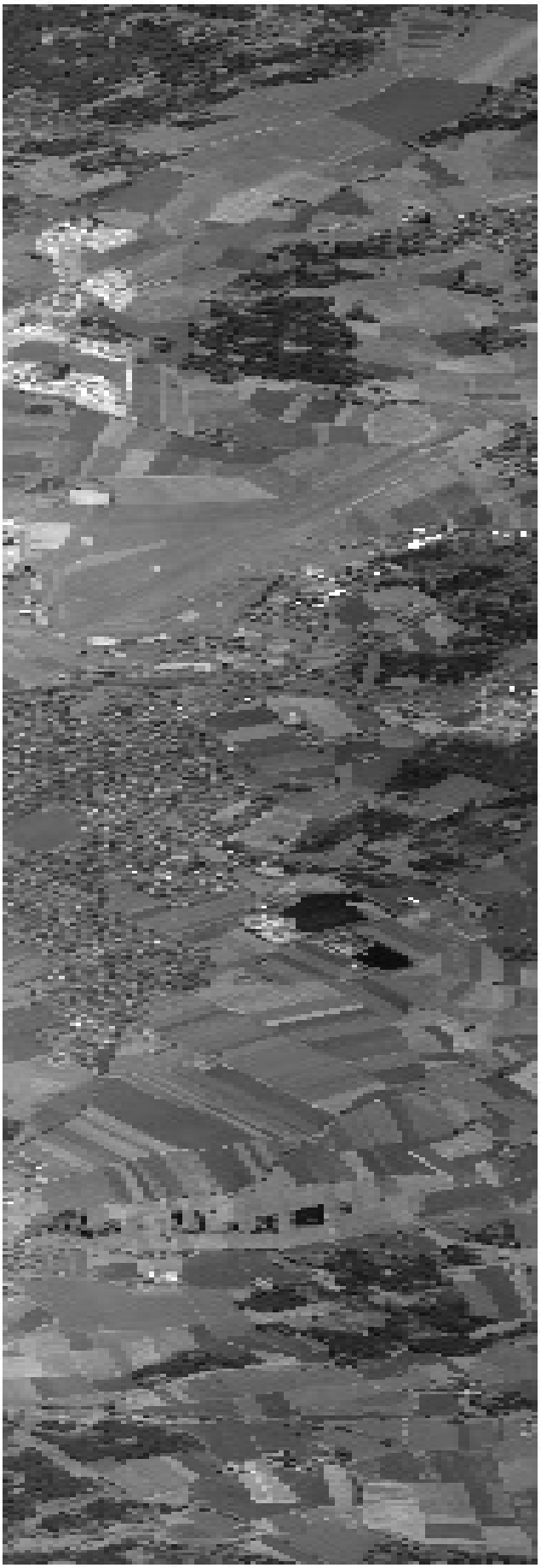

Figure 7. Hyperspectral scene by HyMAP sensor capturing the DLR facilities in Oberpfaffenhofen and its surroundings.

We cut the scene in patches of 64×64 pixels size for a total of 360 patches forming the hyperspectral database used in the experiments. We grouped the patches by visual inspection in five rough categories. The three main categories are 'Forests', 'Fields' and 'Urban Areas', representing patches that mostly belong to one of this categories. A 'Mixed' category was defined for those patches that presented more than one of the three main categories, being not any of them dominant. Finally, we defined a fifth category, 'Others', for those patches that didn't represent any of the above or that were not easily categorized by visual inspection. The number of patches per category are: (1) Forests: 39, (2) Fields: 160, (3) Urban Areas: 24, (4) Mixed: 102, and (5) Others: 35.

B. Methodology

We perform three experiments to validate the use of the proposed Spectral-Spatial CBIR system in a real life scenario. In the first experiment we tested the system using the patches belonging to the three main categories: Forests, Fields and Urban Areas. In second experiment we added patches from the fourth category: Mixed. Finally, in third experiment we used the full patches database.

Let us denote $\mathbf{s}_{\alpha}^{\mathrm{IND}}$ the vector of dissimilarities computed using the endmembers induced by one of the EIAs (either ILSIA, N-FINDER or FIPPI) and Ω_{α}^{IND} to their respective rankings. The groundtruth is given by the a-priori categorization made by visual inspection, and the set of relevant images is composed in a different way to previous experiments using synthetic data. Given a query $Q_k(H_\alpha)$, the set of returned images $T_k(H_\alpha)$ and the set of relevant images $V_k(H_\alpha)$ are defined as follows:

$$T_k(H_\alpha) = \Omega_{\alpha,k}^{IND} = \left[\omega_{\alpha,p}^{IND} \text{ s.t. } s_{\alpha,\omega_{\alpha,p}^{IND}} \leq s_{\alpha,\omega_{\alpha,k}^{IND}}\right] \quad (12)$$

$$V_k(H_\alpha) = \Omega_{\alpha}^{GT} = \left[\beta \,\text{s.t.}\, \mathcal{C}\left(\beta\right) = \mathcal{C}\left(\alpha\right)\right] \quad (13)$$

where $\mathcal{C}\left(\gamma\right)$ indicates the category to which the patch H_γ belongs. This way, the relevant set for a query patch H_α is formed for all those patches belonging to its same category $\mathcal{C}\left(\alpha\right)$. Now $T_k\left(H_\alpha\right)$ and $V_k\left(H_\alpha\right)$ can be used to calculate the average precision and recall measures of the system, as well as the average normalized rank.

C. Performance results

Figure 8 shows the precision-recall curve for the first experiment, with varying EIA and endmember distance. Surprisingly, the best result correspond to the SAM distance, when using the N-FINDR endmember induction. For low recall values, the differences between algorithms and distances are negligible. In all combinations, the precision-recall curves are very high, showing that the approach is feasible for real-life applications. In table II we have the ANR results for the first experiment. Notice that most values when computing the SAM among endmembers induced by N-FINDR are below 0.1.

Adding new categories, the performance of the system degrades gracefully, as shown in figures 9 and 10, maintaining high precision for low recall values. The ANR results in tables III and IV still provide the best results for SAM distance. This must be due to the magnitude normalization performed by the SAM that removes some illumination effects that are stronger for the Euclidean distance. In all cases, the non-homogeneous categories, such as Urban Area, Mixed and Other, are the most difficult to retrieve, and the inclusion of new categories does not help to improve retrieval figures.

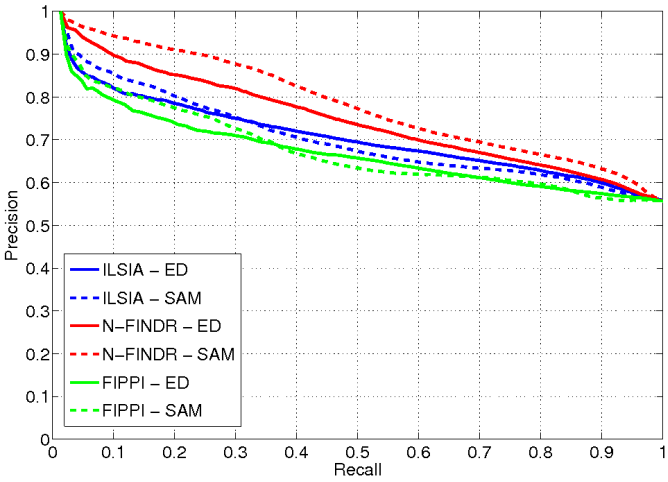

Figure 8. Precision-recall curves for HyMAP experiment 1.

	Averaged Normalized Rank (ANR)					
	ILSIA		N-FINDR		FIPPI	
Category	ED	SAM	ED	SAM	ED	SAM
Forests	0.115	0.069	0.083	**0.023**	0.143	0.082
Fields	0.093	0.109	0.090	**0.079**	0.119	0.128
Urban Areas	0.334	0.250	0.152	**0.010**	0.255	0.220
Average	0.181	0.143	0.108	**0.068**	0.172	0.143

Table II
ANR RESULTS FOR HYMAP EXPERIMENT 1.

VI. CONCLUSIONS

This paper introduces a spatial-spectral CBIR system, providing validation results on synthetic and real hyperspectral

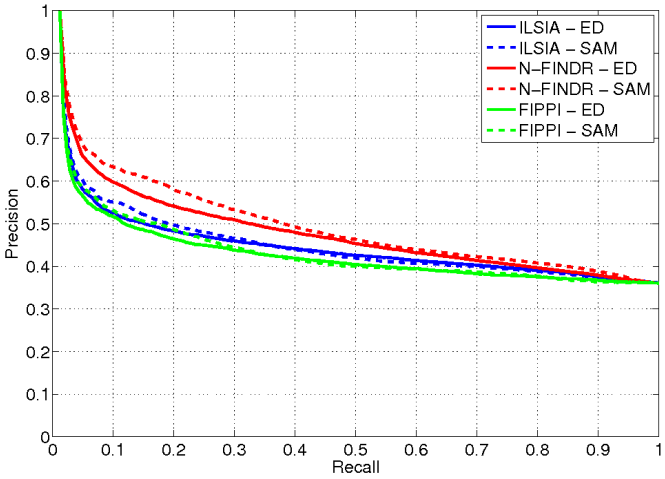

Figure 9. Precision-recall curves for HyMAP experiment 2.

	Averaged Normalized Rank (ANR)					
	ILSIA		N-FINDR		FIPPI	
Category	ED	SAM	ED	SAM	ED	SAM
Forests	0.146	0.097	0.115	**0.054**	0.169	0.113
Fields	0.194	0.213	0.187	**0.180**	0.219	0.227
Urban Areas	0.338	0.255	0.156	**0.108**	0.253	0.220
Mixed	0.356	**0.340**	0.352	0.348	0.364	0.359
Average	0.259	0.226	0.203	**0.173**	0.251	0.230

Table III
ANR RESULTS FOR HYMAP EXPERIMENT 2.

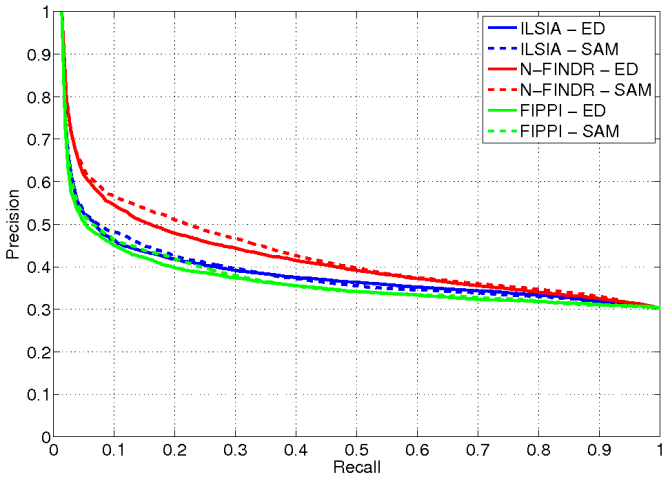

Figure 10. Precision-recall curves for HyMAP experiment 3.

data. The source code for the synthetic data is free and available via the research group's site. To validate our approach we have followed a rigorous methodological framework using both synthetic and real datasets. The results on the synthetic datasets demonstrate the system robustness against noise and changes in the choice of endmember induction algorithm and endmember distances. The results on the real data confirm the usefulness of the proposed system. Further comparison with other CBIR systems based on other features would reinforce this conclusion.

Further research efforts will be addressed to the definition of a convenient user interaction for relevance feedback implementations based on the one-class SVM and/or similar classification approaches.

ACKNOWLEDGMENTS

The authors very much acknowledge the support of Dr. Martin Bachmann from DLR.

	Averaged Normalized Rank (ANR)					
	ILSIA		N-FINDR		FIPPI	
Category	ED	SAM	ED	SAM	ED	SAM
Forests	0.139	0.095	0.109	**0.051**	0.166	0.113
Fields	0.211	0.228	0.197	**0.196**	0.236	0.242
Urban Areas	0.336	0.261	0.155	**0.119**	0.250	0.224
Mixed	0.360	**0.347**	0.352	0.356	0.372	0.368
Other	0.460	0.466	0.455	**0.404**	0.482	0.463
Average	0.301	0.280	0.254	**0.225**	0.301	0.282

Table IV
ANR RESULTS FOR HYMAP EXPERIMENT 3.

Algorithm 2 ILSIA algorithm

1) Initialize the set of lattice sources $X = \{\mathbf{x}_1\}$ with a randomly picked vector in the input dataset Y.
2) Construct the LAAM, W_{XX}, based on the strong lattice independent (SLI) vectors, X.
3) For each data vector $\mathbf{y}_j$, $j = 1, \ldots, N$:
   a) If $\mathbf{y}_j = W_{XX} \boxtimes \mathbf{y}_j$ then $\mathbf{y}_j$ is lattice dependent on the set of lattice sources X, skip further processing.
   b) Test max/min dominance to ensure SLI, consider the enlarged set of lattice sources $X' = X \cup \{\mathbf{y}_j\}$
      i) $\mu_1 = \mu_2 = 0$
      ii) For $i = 1, \ldots, K+1$
      iii) $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{0}$
         A) For $j = 1, \ldots, K+1$ and $j \neq i$
            $\mathbf{d} = \boldsymbol{x}_i - \mathbf{x}_j$; $m_1 = \max(\mathbf{d})$; $m_2 = \min(\mathbf{d})$.
            $\mathbf{s}_1 = \mathbf{s}_1 + (\mathbf{d} == m_1)$, $\mathbf{s}_2 = \mathbf{s}_2 + (\mathbf{d} == m_2)$.
         B) $\mu_1 = \mu_1 + (\max(\mathbf{s}_1) == K)$ or $\mu_2 = \mu_2 + (\max(\mathbf{s}_2) == K)$.
      iv) If $\mu_1 = K+1$ or $\mu_1 = K+1$ then $X' = X \cup \{\mathbf{y}_j\}$ is SLI, go to 2 with the enlarged set of lattice sources and resume exploration from $j+1$.
4) The final set of lattice sources is X.

Miguel A. Veganzones is supported by predoctoral grant BFI_07.225 from the Basque Government.

APPENDIX

A. ILSIA

The Incremental Lattice Source Induction Algorithm (ILSIA) [15] is grounded in the formal results on continuous Lattice Auto-Associative Memories (LAAMs) [20]. The algorithm aims to produce sets of Strong Lattice Independent (SLI) vectors extracted from the input dataset. The resulting sets are affine independent, that is, they define convex polytopes that cover some (most of) the data points in the dataset. Algorithm (2) shows the ILSIA pseudo-code. To ensure that the resulting set of vectors are SLI, we first ensure that they are lattice independent in step 3a of Algorithm. Each new input vector is applied to the LAAM constructed with the already selected lattice sources. If the recall response evoked by the vector is perfect, then it is lattice dependent on the lattice sources, and can be discarded. If not, then the new input vector is a candidate lattice source. In step 3b the min and max dominance of the set of lattice sources enlarged with the new input vector is tested.

B. N-FINDR

Algorithm 3 presents the N-FINDER [16] pseudo-code. The N-FINDER algorithm works by growing a simplex inside the data, beginning with a random set of pixels. The vertexes of the simplex with higher volume are assumed to identify the endmembers. Previously, data dimensionality has to be reduced to

Algorithm 3 N-FINDR algorithm

1) Apply Principal Component Analysis (PCA) to reduce the data dimensionality. Keep the first $p-1$ principal components.
2) Randomly select p vectors from the data to initialize the set of induced endmembers E.
3) Calculate the volume of the simplex $v = V(E)$ (15). $v_{actual} = v$.
4) For each endmember $\mathbf{e}_k$, $k = 1, \dots p$:
   a) For each data vector $\mathbf{x}_i$, $i = 1, .., N$:
      i) Form a new matrix E' by substituting the endmember $\mathbf{e}_k$ by the data vector $\mathbf{x}_i$.
      ii) Calculate the volume of the simplex $v' = V(E')$.
      iii) If $v' > v_{actual}$ then E' becomes E. $v_{actual} = v'$.
5) If $v_{actual} > v$ then $v = v_{actual}$. Go to step 4.

$p-1$ dimensions, being p the number of endmembers searched for.

Let E be the matrix of endmembers augmented with a row of ones

$$E = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{e_1} & \mathbf{e_2} & \dots & \mathbf{e}_p \end{bmatrix}, \tag{14}$$

where $\mathbf{e}_i$ is a column vector containing the spectra of the i-th endmember. The volume of the simplex defined by the endmembers is proportional to the determinant of E

$$V(E) = \frac{abs(\det(E))}{(p-1)!}. \tag{15}$$

The N-FINDER starts by selecting an initial random set of pixels as endmembers. Then, for each pixel and each stored endmember, the endmember is replaced with the spectrum of the pixel and the volume recalculated by equation (15). If the volume of the new simplex increases, the endmember is replaced by the spectrum of the pixel. The procedure ends when no more replacements are done. The N-FINDER is a greedy algorithm, prone to fall in local maxima of the volume function.

C. FIPPI

The Fast Iterative Pixel Purity Index (FIPPI) [17] method is an improved version of the Pixel Purity Index [21] algorithm for endmembers induction. PPI is an heuristic algorithm based on the projection of the dimensionality reduced dataset onto a set of k random unit vectors denoted as skewers, $\{skewer_j\}_{j=1}^{k}$, where k is a sufficiently large positive integer. All the data sample vectors are projected onto each $skewer_j$, selecting the extreme vectors that form an extrema set, denoted by $S(skewer_j)$. The PPI score for each point vector r is calculated by

$$N_{PPI}(r) = \sum_j I_{S(skewer_j)}(r) \tag{16}$$

where $I_{S(skewer_j)}$ is an indicator function of an extrema set defined as

$$I_{S(skewer_j)} = \begin{cases} 1, & if\ r \in S(skewer_j) \\ 0, & if\ r \notin S(skewer_j) \end{cases} \tag{17}$$

After finding the PPI scores, $N_{PPI}(r)$, for all the sample vectors, the vectors with a PPI scores such that $N_{PPI}(r) \geq t$ are selected as endmember candidates. The PPI algorithm involves defining two parameters k and t, being k the number of random generated skewers and t the threshold value to find the endmember candidates.

PPI has several drawbacks, it's not an iterative process and does not guarantee that the generated endmembers are actually true endmembers due to the randomness of generated skewers, a different set of skewers generate a different set of endmembers; it's very sensible to noise, there is no criteria to select the correct values for the k and t parameters, which determine the number of final endmembers, and it requires human intervention to manually select a final set of endmembers. FIPPI algorithm addresses these major drawbacks. FIPPI uses the HFC method [18] to find the virtual dimensionality of the data, being p the number of endmembers required to be generated. Then, FIPPI applies a MNF (or PCA) transform to reduce the data to the first resulting p dimensions. FIPPI also uses the Automated Target Generation Process (ATGP) [22] which is an Endmember Initialization Algorithm to generate the initial set of p skewers, $\left\{skewers_j^{(0)}\right\}_{j=1}^{p}$. After algorithm initialization, FIPPI iterates projecting in each iteration k all the sample vectors onto each skewer, $skewer_j^{(k)}$ to find those which are at its extreme positions to form an extrema set, denoted by $S(skewer_j^{(k)})$, and then to find the sample vectors, $\left\{r_j^{(k)}\right\}$, that produce the largest $N_{PPI}(r_j^{(k)})$ values (16). In each iteration the joint set $\left\{skewer_j^{(k+1)}\right\} = \left\{r_j^{(k)}\right\}_{N_{PPI}(r_j^{(k)})>0} \cup \left\{skewer_j^{(k)}\right\}$ is formed. If $\left\{skewer_j^{(k+1)}\right\} = \left\{skewer_j^{(k)}\right\}$ no new endmembers are added to the skewer set and the algorithm is terminated returning as endmembers those vectors with $N_{PPI}(r_j^{(k+1)}) > 0$.

REFERENCES

[1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, 2006.

[3] Y. Liu, D. Zhang, G. Lu, and W. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, pp. 262–282, Jan. 2007.

[4] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives: system concepts," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, no. 12, pp. 2923–2936, 2003.

[5] H. Daschiel and M. Datcu, "Information mining in remote sensing image archives: system evaluation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 1, pp. 188–199, 2005.

[6] M. Datcu and K. Seidel, "Human-centered concepts for exploration and understanding of earth observation images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 601–609, 2005.

[7] J. Maldonado, D. Vicente, M. Veganzones, and M. Grana, "Spectral indexing for hyperspectral image CBIR," in *Proceedings of the ESA-EUSC 2006 Image Information Mining for Security and Intelligence conference*, ESA-EUSC '06, (Torrejon air base, Madrid (Spain)), ESA Publications Division, 2006.

[8] A. Plaza, J. Plaza, A. Paz, and S. Blazquez, "Parallel cbir system for efficient hyperspectral image retrieval from heterogeneous networks of workstations," in *Symbolic and Numeric Algorithms for Scientific Computing, 2007. SYNASC. International Symposium on*, pp. 285–291, sept. 2007.

[9] M. A. Veganzones, J. O. Maldonado, and M. Grana, "On Content-Based image retrieval systems for hyperspectral remote sensing images," in *Computational Intelligence for Remote Sensing*, vol. 133 of *Studies in Computational Intelligence*, pp. 125–144, Springer Berlin / Heidelberg, 2008.

[10] M. Grana, M. Veganzones, and J. Gallego, "An endmember-based semimetric for content based hyperspectral image retrieval," submitted.

[11] N. Keshava and J. F. Mustard, "Spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 44–57, 2002.

[12] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: integrated region matching for image retrieval," in *Proceedings of the eighth ACM international conference on Multimedia*, MULTIMEDIA '00, (Marina del Rey, California, United States), pp. 147–156, ACM, 2000.

[13] C. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer, 1 ed., July 2003.

[14] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: overview and proposals," *Pattern Recognition Letters*, vol. 22, pp. 593–601, Apr. 2001.

[15] M. Grana, D. Chyzhyk, M. Garcia-Sebastian, and C. Hernandez, "Lattice independent component analysis for functional magnetic resonance imaging," *Information Sciences*, vol. 181, pp. 1910–1928, May 2011.

[16] M. E. Winter, M. R. Descour, and S. S. Shen, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," vol. 3753, (Denver, CO, USA), pp. 266–275, SPIE, Oct. 1999.

[17] C. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 1, pp. 63–67, 2006.

[18] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 3, pp. 608–619, 2004.

[19] C. L. Lawson, *Solving Least Squares Problems*. Prentice Hall, June 1974.

[20] G. X. Ritter, P. Sussner, and D. J. L., "Morphological associative memories," *Neural Networks, IEEE Transactions on*, vol. 9, no. 2, pp. 281–293, 1998.

[21] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of aviris data," *JPL Technical Reports Summary*, vol. JPL TRS 1992+, 1995.

[22] A. Plaza and C. Chang, "Impact of initialization on design of endmember extraction algorithms," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 11, pp. 3397–3407, 2006.