CONTENT BASED HYPERSPECTRAL IMAGE RETRIEVAL USING BAG OF ENDMEMBERS IMAGE DESCRIPTORS

Fatih Ömrüuzun[1], Begüm Demir[2], Lorenzo Bruzzone[2], Yasemin Yardımcı Çetin[1*]*

[1]Information Systems Department, Middle East Technical University, Ankara, Turkey
{omruuzun, yyardim}@metu.edu.tr
[2]RSLab, Information Engineering and Computer Science Department, University of Trento, Trento, Italy
{begum.demir, lorenzo.bruzzone}@unitn.it

ABSTRACT

This paper proposes a novel system for fast and accurate content based retrieval of hyperspectral images. The proposed system aims at retrieving hyperspectral images that have both similar spectral characteristics associated with specific materials and fractional abundances to the query image. It consists of two modules. The first module characterizes the query and the target hyperspectral images in the archive by two descriptors: 1) a binary spectral descriptor representing spectral characteristics of distinct materials 2) an abundance descriptor that contains the normalized cumulative fractional abundance information of the corresponding materials. Both descriptors are obtained by a novel bag of endmembers based strategy. The second module aims at retrieving hyperspectral images from the archive that are most similar to query image based on a hierarchical strategy which evaluates the spectral and abundance descriptors similarity. Experiments carried out on a benchmark archive of hyperspectral images demonstrated the effectiveness of the proposed system in terms of retrieval accuracy and time.

***Index Terms*—** hyperspectral imaging, content based image retrieval, feature extraction, unmixing, bag of endmembers, remote sensing

1. INTRODUCTION

Due to the recent advances in hyperspectral imaging technologies, the volume of hyperspectral image archives has increased significantly. Accordingly, one emerging application is the accurate and fast retrieval of hyperspectral images from an archive with respect to their contents. Hyperspectral images have dense spectral information that makes the detection and discrimination of materials possible. The term content in hyperspectral image retrieval can be interpreted diversely depending on the application scenarios. We can define four different categories of problems related to the quantity of distinct materials being queried and the significance of abundance (see Table 1). The first and second categories correspond to cases where the system retrieves images from the archive that contains a specific material represented by a query image. These scenarios are related to *target detection* where presence of even one pixel in the image that is defined by the given query is important. In the Scenario 2, unlike the Scenario 1, a definite quantity of abundance of a material represented by the query is also considered as a key element of the query. In the third and fourth scenarios, the content based hyperspectral image retrieval (CBHIR) system aims at retrieving images that contain multiple materials present in the query image. In the Scenario 3, abundances are not considered, whereas they are included in the fourth scenario.

Table 1. Categorization of CBHIR Scenarios

	Query	
Scenario	**Quantity of Materials**	**Abundance**
1	Single Material	Insignificant
2	Single Material	Significant
3	Multiple Material	Insignificant
4	Multiple Material	Significant

Any CBHIR system essentially consists of two modules: 1) image descriptor module that extracts a set of descriptors that model the distinct materials and their abundances (if considered), 2) retrieval module that retrieves hyperspectral images similar to the query images based on the considered scenario. In this paper, we focus our attention on the last two scenarios, where the aim is retrieving hyperspectral images that include multiple materials contained in the query image. The CBHIR systems proposed in the literature associated to these scenarios rely on the descriptors related to the spectral signatures of distinct materials that are modeled by endmembers and related abundances present in the images. Different retrieval strategies are used. For example in [1] Spectral Angular Distance (SAD) is estimated between each query and target hyperspectral image endmember pairs to quantify simi-

*This research is supported by TUBİTAK 2214/A Research Fellowship Program for PhD Students and METU Scientific Research Projects Grant No: 07-04-2015-005.

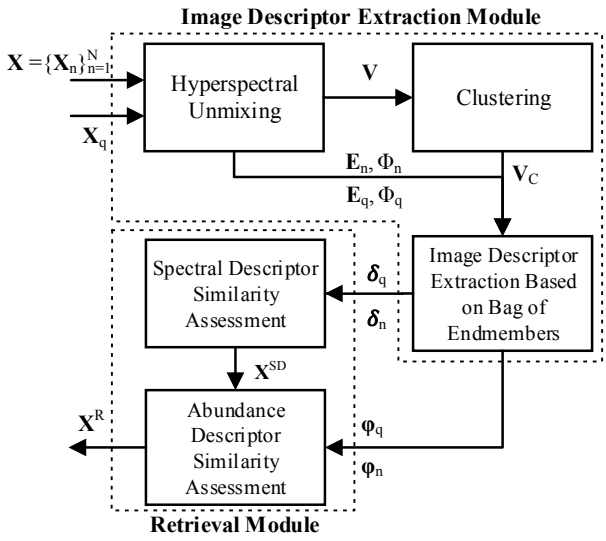

Fig. 1. Block diagram of the proposed CBHIR system.

larity between images. In [2], the retrieval module of the system is built on *Grana Distance* that calculates the similarity between two hyperspectral images as the cumulative magnitude of row and column minimum vectors of the query and target endmember distance matrix. The retrieval module of the CBHIR system proposed in [3] considers the abundance of endmembers in the hyperspectral images as well as pairwise endmember similarities between query and target images. All these systems have two limitations: 1) the storage of all the endmembers in the auxiliary archive is challenging as they are often represented in a very high dimensional feature space; 2) their computational cost increases proportionally with respect to number of endmembers contained in the images. To overcome these problems, in this paper, we propose a system that aims to represent each individual hyperspectral image with only two descriptors obtained via a novel bag of endmembers based feature extraction method.

2. PROPOSED CBHIR SYSTEM

The proposed content based hyperspectral image retrieval system consists of two modules: 1) extraction of bag of endmembers image descriptor, and 2) two-step hierarchical retrieval, based on the similarity of query and target hyperspectral images in terms of spectral and abundance image descriptors. The block diagram of the proposed CBHIR system is illustrated in Fig. 1, and each step is explained in detail in the following subsections.

2.1. Extraction of Bag of Endmembers Image Descriptors

Let $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N\}$ be an archive consisting of N hyperspectral images, where $\mathbf{X}_n, n = 1, \dots N$, is the n^{th} hyperspectral image that is composed of l bands. $\mathbf{X}_n$ consists of S pixels, i.e., $\mathbf{X}_n = \left\{\mathbf{x}_n^1, \mathbf{x}_n^2, ..., \mathbf{x}_n^S\right\}$, $\mathbf{x}_n^s \in \mathbb{R}^l$, $s = 1, \dots, S$. The proposed CBHIR system aims to model each image by two descriptors: 1) a spectral descriptor $\boldsymbol{\delta}$ representing the spectral characteristics of distinct materials contained in the image 2) an abundance descriptor $\boldsymbol{\varphi}$ holding normalized cumulative fractional abundance information of each material. These descriptors are obtained based on a novel bag of endmembers approach. Let $\mathbf{E}_n = \left\{\mathbf{e}_n^1, \mathbf{e}_n^2, ..., \mathbf{e}_n^M\right\}$ be the set of M endmembers extracted from the hyperspectral image $\mathbf{X}_n$, where $\mathbf{e}_n^m \in \mathbb{R}^l$ and $m = 1, \dots, M$. Let $\Phi_s = \left\{\phi_s^1, \phi_s^2, ..., \phi_s^M\right\}$ be the corresponding abundance of each endmember in pixel $\mathbf{x}_n^s$, where $0 \leq \phi_s^m \leq 1$. According to linear unmixing model, each pixel x_n^s is a linear combination of $\mathbf{E}_n$, Φ_s and an additive noise η such that $\mathbf{x}_n^s = \sum_{m=1}^{M} \mathbf{e}_n^m \cdot \phi_s^m + \eta$. In light of this information, normalized fractional abundance of each endmember in $\mathbf{X}_n$ can be defined as $\bar{\Phi}_n = \left\{\bar{\phi}_n^1, \bar{\phi}_n^2, ..., \bar{\phi}_n^M\right\}$, where $\bar{\phi}_n^m = 1/S \sum_{s=1}^{S} \left(\phi_s^m / \sum_{m=1}^{M} \phi_s^m\right)$.

To compute $\boldsymbol{\delta}_n$ and $\boldsymbol{\varphi}_n$ for each hyperspectral image $\mathbf{X}_n$, an endmember vocabulary $\mathbf{V}$ including endmember signatures obtained from $Z \leq N$ images in the archive is generated, i.e. $\mathbf{V} = \bigcup_{z=1}^{Z} \mathbf{E}_z = \mathbf{E}_1 \cup \mathbf{E}_2 ... \cup \mathbf{E}_Z$. Then, endmembers vocabulary $\mathbf{V}$ is clustered into K clusters, and the cluster centers $\mathbf{V}_C = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K\}$ representing a cluster of distinct materials present in endmembers vocabulary $\mathbf{V}$, are estimated. In this way, endmember signatures representing similar materials in the hyperspectral images from the archive are grouped into the same clusters. These cluster centers are used in computation of spectral and abundance descriptors $\boldsymbol{\delta}_n$ and $\boldsymbol{\varphi}_n$ with respect to presence of similar endmembers to the centers and related abundances, respectively. To this end, a spectral distance matrix $\mathbf{D}_{\mathbf{E}_n, \mathbf{V}_C} = [d_{i,j}; i = 1, ..., M; j = 1, ..., K]$ is initially constructed for each hyperspectral image; where $d_{i,j}$ is the distance between the i^{th} endmember and j^{th} cluster center defined by any spectral distance measure. Then, minimum element in each row of the distance matrix is set to 1, and other elements are set to 0. In this way, the distance matrix $\mathbf{D}_{\mathbf{E}_n, \mathbf{V}_C}$ is converted into a binary matrix indicating similarity of image endmembers to cluster centers $\mathbf{V}_C$. Then, the j^{th} component of $\boldsymbol{\delta}_n$ indicating existence of a similar material in hyperspectral image $\mathbf{X}_n$ to the i^{th} cluster is calculated as follows:

$$\delta_n^j = \begin{cases} 1, & \text{if } \sum_{i=1}^{M} \mathbf{D}_{\mathbf{E}_n, \mathbf{V}_C}[i, j] \geq 1 \\ 0, & otherwise \end{cases} \quad (1)$$

According to (1), $\boldsymbol{\delta}_n = \left\{\delta_n^j\right\}_{j=1}^{K}$ is a K dimensional binary spectral descriptor where each element represents existence of a material in hyperspectral image represented by the j^{th} cluster. Abundance descriptor φ_n is also calculated as shown in (2).

$$\boldsymbol{\varphi}_n = \bar{\Phi}_n^T \cdot \mathbf{D}_{\mathbf{E}_n, \mathbf{V}_C} \quad (2)$$

According to (2), $\boldsymbol{\varphi}_n = \left\{\varphi_n^j\right\}_{j=1}^K$ is a K dimensional abundance descriptor where each element quantifies normalized cumulative amount abundance of the similar material in hyperspectral image $\mathbf{X}_n$ to the j^{th} cluster. Due to our strategy on bag of endmembers, the proposed spectral and abundance descriptors are independent from the number of pixels in the hyperspectral images. It is worth noting that the effectiveness of the proposed descriptors depends on the number K of clusters and also the number Z of the images being clustered. The number of clusters should be equal or higher than that of the number of distinct materials present in the archive images in the archive. The number Z of images being clustered should include the endmembers of the materials considered in the image archive.

2.2. Hierarchical Strategy Based Hyperspectral Image Retrieval

The proposed hyperspectral image retrieval module exploits a two-steps hierarchical strategy based on the evaluation of spectral and abundance descriptors $\boldsymbol{\delta}$ and $\boldsymbol{\varphi}$. In the first step, the similarity between query and target hyperspectral images in the archive is computed by the Hamming distance with respect to the binary spectral descriptor $\boldsymbol{\delta}$. The Hamming distance is defined as the sum of bitwise XOR operator on given two binary spectral descriptor vectors $\boldsymbol{\delta}_q$ and $\boldsymbol{\delta}_n, n = 1, \ldots N$. Then, the set $\mathbf{X}^{SD}$ of μ images that have the lowest Hamming distance is chosen. In the second step of the retrieval module, pairwise Euclidean distance between abundance descriptors $\boldsymbol{\varphi}$ of query image and those of hyperspectral images in $\mathbf{X}^{SD}$ are computed. Then, final set of images $\mathbf{X}^R \subset \mathbf{X}^{SD}$, $h \leq \mu$ that are the most similar to the query in terms of the abundance similarity are retrieved. It is worth noting that if the Scenario 3 (given in Table 1) is considered, the second step can be avoided.

3. EXPERIMENTAL RESULTS

To generate a benchmark hyperspectral image archive useful for validation purposes, we selected an image acquired by EO-1 Hyperion sensor over surroundings of Ankara, Turkey. The image consists of 3461x256 pixels with a spatial resolution of 30m. The image was divided into tiles of 63x63 pixels resulting in archive of 216 hyperspectral images. After removing spectral bands those with low signal-to-noise ratio, 119 spectral bands were used. Each hyperspectral image in the archive is annotated with multiple category labels. The list of category labels and number of hyperspectral images associated with each label in the archive are given in Table 2. The number of category labels associated with each image varies between 3 and 17.

Table 2. The category labels and number of hyperspectral images annotated with each label in the benchmark archive.

Category Label	Number of Images	Category Label	Number of Images
Grass Covered Soil	215	Blue Roofing	25
Bare Soil	216	Yellow Roofing	13
Arid Soil	10	Membrane Roofing	57
Rocky	31	Concrete Roofing	55
Tree	174	White Tent	6
Reeds	5	Unpaved Road	106
Crop (Type-A)	47	Asphalt Pavement	183
Crop (Type-B)	38	Highway Pavement	12
Crop (Type-C)	2	Grass (Type-A)	215
Crop (Type-D)	55	Grass (Type-B)	12
Crop (Type-E)	4	Grass (Type-C)	25
Red Roofing	131	Lake	11
Metal Roofing	130	Pool	23
White Roofing	122	Cloud	6
Green Roofing	41		

Since each hyperspectral image in the archive is associated with more than one category label, we used multi-label performance metrics to assess the performance of the proposed system. The considered metrics are accuracy, precision, recall and Hamming loss, which are defined for multi-label retrieval problems (see Table 3). Accuracy measures the performance based on the ratio of total number of intersecting labels in query and retrieved images to the total number of unique labels associated to those images. Precision is the average ratio of the total number of intersecting labels in query and retrieved images to number of labels associated with the retrieved images. On the other hand, recall considers the average ratio of the total number of intersecting labels in query and retrieved images to number of labels associated with the query image. Hamming loss measures the ratio of average number of symmetric difference between label sets associated with query and the retrieved images to number of category labels used for annotating images in the archive.

The performance of proposed CBHIR system was compared with the state of the art CBHIR system introduced in [3]. Both CBHIR systems extract hyperspectral image descriptors based on endmembers and thus the same hyperspectral unmixing setup was used during the experiments. In the unmixing process, the number of principal components having %99 of the total variance in a hyperspectral image was given as input to the Vertex Component Analysis (VCA) [4] algorithm as the number of endmembers. Then, Fully Constrained Least Squares (FCLS) [5] algorithm was used to obtain abundance of endmembers in each individual pixel to ensure linear unmixing constraints. In order to calculate the the distance matrix required by the retrieval module of the system proposed in [3], Spectral Information Distance (SID) [6] was used. The experiments were conducted in MATLAB environment installed on a PC which has 64 GB RAM and Intel i7 processor. In the experiments, considering that we have a benchmark archive, the endmember dictionary $\mathbf{V}$ was gener-

Table 3. Performance metrics used in the experiments. (L_q, $L_{X_r^R}$ and L_X are the category labels associated with query image $\mathbf{X}_q$, retrieved image $\mathbf{X}_r^R$ and set of category labels associated to archive $\mathbf{X}$, respectively.)

Metric	Definition
Accuracy	AC=$\frac{1}{\lvert\mathbf{X}^R\rvert}\sum_{r=1}^{\lvert\mathbf{X}^R\rvert}\lvert L_q \bigcap L_{X_r^R}\rvert / \lvert L_q \bigcup L_{X_r^R}\rvert$
Precision	PR=$\frac{1}{\lvert\mathbf{X}^R\rvert}\sum_{r=1}^{\lvert\mathbf{X}^R\rvert}\lvert L_q \bigcap L_{X_r^R}\rvert / \lvert L_{X_r^R}\rvert$
Recall	RC=$\frac{1}{\lvert\mathbf{X}^R\rvert}\sum_{r=1}^{\lvert\mathbf{X}^R\rvert}\lvert L_q \bigcap L_{X_r^R}\rvert / \lvert L_q\rvert$
Hamming Loss	HL=$\frac{1}{\lvert\mathbf{X}^R\rvert}\sum_{r=1}^{\lvert\mathbf{X}^R\rvert}\lvert L_q \Delta L_{X_r^R}\rvert / \lvert L_X\rvert$

ated from endmembers extracted from all the images in the archive, i.e., $Z = N$. In order to cluster endmembers, kernel k-means algorithm [7] was used during the experiments with the radial basis function (RBF) kernel. The gamma parameter of the RBF kernel was selected to reciprocal of mean squared Euclidean distance between all endmembers. In the experiments, the value of h is set to 15, whereas that of μ is fixed as $2h$. From our initial experiments, we have observed that selecting higher values for μ reduces the retrieval performance. In the experiments, the number K of clusters is varied between [2-100] with step size increment of 1. From the results we observed that selecting very small number of K provides poor retrieval performances. The value of K can be selected according to the number of different materials present in the archive images (which requires a pre-knowledge on the considered archive). Table 4 shows the results obtained by the two systems. The results of the proposed system were obtained by considering $K = 30$. All the results given in Table 4 are associated to the average results obtained by considering each image in the archive as the query image. In this way, overall performance of the system for different query images was observed. In the table, the computational time required for the retrieval phase of each system is also provided. By analyzing the Table 4, one can observe that the proposed method provides higher performance in less computational time. As an example, when accuracy metric is used, the proposed system provides an accuracy of 61.09% in 0.05 milliseconds, whereas the system introduced in [3] yields an accuracy of 57.12% in 31 milliseconds. As an other example, the proposed system provides 3% higher recall performance than the system presented in [3]. These findings prove the efficiency of proposed bag of endmembers based image descriptor extraction strategy in image retrieval.

4. CONCLUSION

This paper proposes a novel bag of endmembers based content based hyperspectral image retrieval system with two modules. While the first module represents hyperspectral images with a binary spectral and abundance descriptors, the second module calculates the similarity between query and hyperspectral images based on hierarchical retrieval strategy. Due to the introduced hyperspectral image descriptors, the proposed system can store efficiently the material information associated to each image and allows to compute similarity between images in a small number of calculations. Experimental results obtained on a benchmark archive of hyperspectral images show that the proposed system allow one to significantly reduce image retrieval time with a high retrieval accuracy compared to the state of the art method. As a future development of this work, we plan to: i) work on multi-label classification algorithms for the retrieval stage; and ii) to develop a strategy to automatically detect the number of clusters required for the image descriptor module.

Table 4. Results obtained by the CBHIR systems.

System	Performance Metric				
	AC (%)	**PR** (%)	**RC** (%)	**HL**	**Time** (ms)
[3]	57.12	75.16	71.90	5.24	31
Proposed	**61.09**	**77.54**	**74.70**	**4.67**	**0.05**

5. REFERENCES

[1] A. J. Plaza, "Content-based hyperspectral image retrieval using spectral unmixing," in *SPIE Remote Sensing*, 2011.

[2] M. Graña and M. A. Veganzones, "An endmember-based distance for content based hyperspectral image retrieval," *Pattern Recognition*, vol. 45, no. 9, pp. 3472 – 3489, 2012.

[3] M. A. Veganzones and M. Graña, "A spectral/spatial cbir system for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 488–500, 2012.

[4] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[5] D. Heinz, C. I. Chang, and M. L. G. Althouse, "Fully constrained least-squares based linear unmixing," in *IEEE International Geoscience and Remote Sensing Symposium*, 1999.

[6] C.I. Chang, "Spectral information divergence for hyperspectral image analysis," in *IEEE International Geoscience and Remote Sensing Symposium*, 1999.

[7] I. S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means: spectral clustering and normalized cuts," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.