LOSSLESS AND GRADUAL CODING OF HYPERSPECTRAL IMAGES BY LIFTING SCHEME

A. Chourou, A. Benazza-Benyahia

Unité de Recherche en Imagerie Satellitaire et ses Applications (URISA),
Ecole Supérieure des Communications de Tunis (SUP'COM)
Cité Technologique des Communications, 2083 Ariana, Tunisia
email: chourou.asma@gmail.com, benazza.amel@supcom.rnu.tn

ABSTRACT

In this paper, we are interested in compactly coding hyperspectral images for applications requiring a progressive and exact reconstruction. For this purpose, a two-step procedure is envisaged. Firstly, a spectral clustering of the whole data set is carried out. Then, within each spectral cluster, a band-ordering of the images is operated and the spectral and spatial similarities are simultaneously taken into account through an extended version of a lifting-based decomposition scheme. Our contribution relies on coupling these two steps and in optimizing the band-ordering procedure.

1. INTRODUCTION

The hyperspectral imagery consists in capturing the same scene in several narrow and adjacent spectral bands. The Airborne Visible Infrared Imaging Spectrometer (AVIRIS) is one of the famous hyperspectral sensor that delivers images calibrated in 224 contiguous spectral channels corresponding to wavelengths ranging from 400 to 2500 nm [1]. The target applications are identification and measurement of earth and atmosphere constituents. AVIRIS produces 76 Giga Bytes of data per day, such amount of data is a bottleneck in handling image databases. Therefore, compression is required. Generally, lossless (or lossy-to-lossless) coding methods are selected since no artifact due to the compression is allowed in the ground and physical measurements. For instance, predictive coding techniques have been successfully used in order to exploit the spatial and spectral correlation between spectral bands [2, 3]. Vector quantization were also considered in a lossy to lossless procedure [4]. In order to ensure both the progressiveness and the exactness of the reconstruction, a great attention was paid to wavelet-based decompositions. More precisely, three dimensional wavelet transforms were firstly considered [5, 6]. However, the main drawback of such approach is the need of decoding the whole dataset in order to retrieve a given spectral image. In [7], a first solution was proposed to fulfill the random access functionality by adding some markers in the binary stream to enable the 2D decoding without the need to decode all the volume set. Similarly, in [9, 10], an alternative solution was also envisaged with a two-step procedure. The first step consists of a clustering of the components according to their spectral features. The second step amounts to apply to the components within the same cluster an extended version of the integer wavelet decomposition, nicknamed Vector Lifting Scheme (VLS) [8]. In this case, the random access property holds since the decoding of any component in a given cluster only involves some of the components of that cluster. Our work aims at improving such approach by introducing more flexibility. Our contribution is two-folded. More precisely, we firstly propose to adapt the number of resulting cluster instead of setting it to a predetermined value. Secondly, we improve the ordering of the spectral components required by the vector lifting decomposition.

Our paper is organized as follows. Section 2 is devoted to the presentation of the adaptive clustering method we applied. In Section 3, we briefly describe the VLS. In Section 4, the optimal component ordering that ensures the best performances of the VLS within each cluster is described. Finally, in Section 5, we present some experimental results and some conclusions are drawn in Section 6.

2. CLUSTERING

Let $s^{(1)}, \ldots, s^{(B)}$ denote a B-channel hyperspectral image of size $M \times N$. We aim at grouping these B spectral components into C clusters and representing each class by a prototype example. Among the numerous clustering methods, we have preferred to consider clustering techniques that does not fix *a priori* the number of classes. Nowdays, the Competitive Agglomeration Algorithm (CAA) is recognized to be one of the most powerful unsupervised clustering technique [11]. It combines the advantages of both partitional and hierarchical classification techniques. Indeed, the CAA starts by partitioning data into a great number of clusters, then it follows an agglomeration rule (which is the property of hierarchical methods) to merge two or more most appropriate clusters. Nevertheless, CAA outperforms the conventional hierarchical approach because it is not static since points belong to their clusters in a fuzzy way. Furthermore, the search of the optimal number of classes is computationally attractive. Generally, it is used to employ the CAA for image segmentation purposes [11, 12]. However to the best of our knowledge, within the framework of hyperspectral image coding, only the k-means algorithm was considered at the stage of the spectral classification. We would like to introduce a more flexible clustering algorithm (namely the CAA) to improve the performances of this prior stage.

Concerning the features set $\mathscr{F} = \{f^{(1)},\ldots, f^{(B)}\}$ of the B channels, several alternatives have been investigated. In this work, we focus on the spatial average and the most frequent intensity. Preliminary experiments have indicated that they lead to the same clustering result. This is the reason why we have retained as a feature $f^{(b)}$ of a component b its spatial means defined by:

$$f^{(b)} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} s^{(b)}(m,n). \quad (1)$$

Let $\Phi = \{\phi_1,\ldots,\phi_C\}$ represents the set of C prototypes. The CAA aims at minimizing the following objective function:

$$\mathscr{J}(\mathscr{F},\Phi,\mathbf{U};\alpha) = \sum_{c=1}^{C}\sum_{b=1}^{B}(u_{cb})^2 d^2(f^{(b)},\phi_c) - \alpha\sum_{c=1}^{C}(\sum_{b=1}^{B} u_{cb})^2 \quad (2)$$

where α is a positive parameter, $d(f^{(b)},\phi_c)$ is the distance between the feature $f^{(b)}$ and the prototype ϕ_c, u_{cb} is the degree of membership of band b to cluster c and $\mathbf{U} = [u_{cb}]$ is the $C\times B$ constrained fuzzy C-partition matrix. Obviously, the constraint $\sum_{c=1}^{C} u_{cb} = 1$ should be satisfied for any b. Instead of using the Euclidean distance, the Mahanalobis one is used to handle general cluster shapes:

$$d^2(f^{(b)},\phi_c) = \frac{(f^{(b)}-\phi_c)^2}{\sigma_c^2} \quad (3)$$

where σ_c^2 is the variance of group c:

$$\sigma_c^2 = \frac{\sum_{b=1}^{B}(u_{cb})^2(f^{(b)}-\phi_c)^2}{\sum_{b=1}^{B}(u_{cb})^2}. \quad (4)$$

The objective function $\mathscr{J}$ of the CAA combines two terms. Its left term corresponds to the fuzzy C-means objective function [13], it is minimal when the number of clusters C is equal to the number B of features (each cluster contains one component). The right term is the sum of cardinalities, its minimum value is achieved when number $C = 1$. Therefore, the CAA attempts to ensure a tradeoff between the sum of the distances between clusters and the most compact partition. The minimization of the criterion $\mathscr{J}$ is iteratively performed and, an empirical rule is proposed for the evolution of the tradeoff parameter α with the iteration number k.

$$\alpha(k) = \eta_0 \exp(-k/\tau)\frac{\sum_{c=1}^{C}\sum_{b=1}^{B}(u_{cb})^2 d^2(f^{(b)},\phi_c)}{\sum_{c=1}^{C}[\sum_{b=1}^{B} u_{cb}]^2} \quad (5)$$

η_0 and τ are the constants ensuring the exponential decrease of α. In our experiments, we have retained the same empirical values mentioned in [11].

3. VECTOR LIFTING

Once the clusters of spectral bands are produced, the goal is to exploit both the spectral and spatial correlation within a given class through a multiresolution representation. The concept of VLS has been found to be a very efficient multiresolution tool to achieve such goal [8]. The principle is to replace the spatial predictor usually employed in conventional lifting schemes [14] by a hybrid one that operates spatial *and* spectral prediction. For the sake of clarity, a separable decomposition is employed and, a single spectral reference channel b_1 is retained for the spectral prediction of the current component b_2. Therefore, it is enough to describe the VLS for 1D signals (for instance, the row of the components). The idea is to estimate the even samples $s^{(b_2)}(2n)$ by the odd neighboring samples $s^{(b_2)}(2n+n_1)$ and also by samples $s^{(b_1)}(n+n_2)$ of the reference channel (where n_1, $n_2 \in \mathbb{Z}$) . The prediction error is then computed and it is used to update the odd samples $s^{(b_2)}(2n+1)$ to generate a coarse decimated version $s_1^{(b_2)}$ of $s^{(b2)}$. The procedure is recursively repeated through J and a sequence of J ($3J$ in the 2D case) of detail coefficients $\{d_j^{(b_2)}\}_{j=1}^{J}$ is produced along with a coarse version $s_j^{(b_2)}$:

$$d_j^{(b_2)}(n) = s_j^{(b_2)}(2n) - \lfloor(\mathbf{p}_j^{(b_2)})^T\tilde{\mathbf{s}}_j^{(b_2)}(2n)\rfloor \quad (6)$$

where $\mathbf{p}_j^{(b_2)}$ is the vector of prediction weights and $\tilde{\mathbf{s}}_j^{(b_2)}(2n)$ gathers all the reference samples. The classical LS is a special case that considers in $\tilde{\mathbf{s}}_j^{(b_2)}(2n)$ only samples from b_2 such as the well known 5/3 integer wavelet transform [14]:

$$\tilde{\mathbf{s}}_j^{(b_2)}(2n) = \begin{pmatrix} s_j^{(b_2)}(2n-1) \\ s_j^{(b_2)}(2n+1) \end{pmatrix}. \quad (7)$$

Generally, the prediction vector $\mathbf{p}_j^{(b_2)}$ is chosen so as the variance of the detail coefficients $d_j^{(b_2)}$ is minimized. If the rounding operator is omitted, it is straightforward to show that the minimum variance predictor should be solution of normal equations.

Sophisticated VLSs could be designated for instance, involving more than one reference channel but at a price of an additional complexity. In our simulations, we have designed the following 4 VLSs that differ in the considered spectral mask.

• VLS1 is a straightforward extension of the 5/3 transform:

$$\tilde{\mathbf{s}}_j^{(b_2)}(2n) = \begin{pmatrix} s_j^{(b_2)}(2n-1) \\ s_j^{(b_2)}(2n+1) \\ s_j^{(b_1)}(2n) \end{pmatrix}. \quad (8)$$

• VLS2 is defined by an extended spectral mask:

$$\tilde{\mathbf{s}}_j^{(b_2)}(2n) = \begin{pmatrix} s_j^{(b_2)}(2n-1) \\ s_j^{(b_2)}(2n+1) \\ s_j^{(b_1)}(2n) \\ s_j^{(b_1)}(2n-1) \\ s_j^{(b_1)}(2n+1) \end{pmatrix}. \quad (9)$$

• VLS3 assumes some symmetries in the weights of adjacent samples:

$$\tilde{\mathbf{s}}_j^{(b_2)}(2n) = \begin{pmatrix} s_j^{(b_2)}(2n-1) + s_j^{(b_2)}(2n+1) \\ s_j^{(b_1)}(2n) \\ s_j^{(b_1)}(2n-1) + s_j^{(b_1)}(2n+1) \end{pmatrix}. \quad (10)$$

• VLS4 also introduces some symmetries in the prediction coefficients:

$$\tilde{\mathbf{s}}_j^{(b_2)}(2n) = \begin{pmatrix} s_j^{(b_2)}(2n-1) + s_j^{(b_2)}(2n+1) \\ s_j^{(b_1)}(2n) \\ s_j^{(b_1)}(2n-1) + s_j^{(b_1)}(2n+1) \\ s_j^{(b_1)}(2n-2) \end{pmatrix}. \quad (11)$$

For all the VLS, the update operation is computed similarly to the LS one:

$$s_{j+1}^{(b_2)}(n) = s_j^{(b_2)}(2n-1) - \lfloor 0.25(d_j^{(b_2)}(n-1) + d_j^{(b_2)}(n)) \rfloor. \quad (12)$$

It is easy to check that these VLSs are perfectly reversible provided that b_1 does not involve samples from b_2 in its prediction. Consequently, it is mandatory to perform a component ordering to choose what reference channel belonging to the underlying cluster could be used as a reference to a current channel.

4. COMPONENT ORDERING

Component ordering designates the component permutation that specifies the channels that are coded first so as they act as good predictors for the subsequent ones. There are five possible orderings: forward monotonic, reverse monotonic, best forward, best reverse [3] and, the optimal ordering proposed by Tate [15]. The two first methods impose a given coding order whereas the remaining ones are adapted to the content of the components. Indeed, the three last approaches attempt to minimize a coding performance criterion on a set of specified candidates. In our work, we have focused on the Tate's method because it is known as the most flexible one. Furthermore, we have retained the entropy as an optimality criterion. Indeed, the entropy is a suitable measure of the sparsity of the multiresolution representation that is independent of the further coding stage.

In each cluster c of cardinality $|c|$, we calculate the resulting entropies of all coded couples of components. The cost matrix is the squared matrix $|c| \times |c|$ whose generic element $b_1 \times b_2$ is the entropy of band b_2 that is coded using a VLS with band b_1. Therefore, the problem of reaching the minimum entropy reduces to a problem of minimum weighted tree search whose nodes are the spectral channels and whose edges are the entropy values. Once, the cost matrix is calculated, we sort its elements in the ascending order. Then, we take the nodes so as that no cycle is generated. It should be noted that a cycle is a connected and closed chain of spectral features such as a component is always predicted by another one. The procedure is iterated until all the nodes are processed. In our algorithm, we check if spatial coding (the 5/3 transform, for instance) should be preferred to the hybrid one. Indeed, if the entropy of a given channel coded by the 5/3 transform is lower than the one achieved by a VLS, we switch to the intra-channel coding mode. Then, this channel becomes a potential root of the subsequent components. Such flexibility in the choice of the spatial/hybrid mode and the possibiliy of having several references in the cluster makes this new component ordering more flexible and general than the one described in [10]. Besides, the possibility of having more than one root allows the constitution of sub-clusters. It is worth noting that the component-ordering could be viewed as a complementary step of the clustering one that could palliate its limitations due the retained features. Finally, the eventual generated sub-classes could facilitate the access to one component within a sub-cluster. Hence, the random access procedure could be accelerated.

5. EXPERIMENTAL RESULTS

We have employed AVIRIS images Cuprite, Jasper Ridge and Lunar Lake ($B = 224$) [1]. Each component is coded at 16 bpp and has a size of 512×614. We have eliminated in our data set the black components

[1]The test images were downloaded from the NASA database http://aviris.jpl.nasa.gov/html/aviris.freedata.html

as indicated in Table 1. Figure 1 provides the profile of the spatial mean and the most frequent intensity along the spectral axis, in the case of the Cuprite scene. It can be easily deduced that both features have the same evolution and, hence, that corroborates why we preferred to retain the spatial mean as a suitable feature $f^{(b)}$ since it could be easily computed. In Table 2, we indicate the different values of entropies corresponding to the considered reversible decomposition methods. The multiresolution decomposition is applied over 3 stages ($J = 3$). The 3D decomposition method corresponds to the volumetric wavelet decomposition described in [6]. It can be noted that the 3D decomposition leads to the most compact representations. However, for a fast decoding, it is preferred to make use of VLS more than a basic LS. To this respect, the VLS2 ensures the best performances for all the tested images. This is the reason why we have retained in the subsequent experiments the VLS2 which considers the largest hybrid mask. An examination of the evolutions with the spectral number of the original entropies and those resulting from the 5/3 transform and our method in Fig. 2 (in the case of Cuprite scene) corroborates the outperformance of VLS2 w.r.t. the LS 5/3 method. The peaks correspond to the coding of the root pictures that is coded in an intra mode (we have chose the 5/3 transform). It can be noted that hybrid coding is often better than spatial one. Such freedom degree in the choice of the spatial/hybrid mode and the possibility of having several references in the cluster yields to a marked decrease of the entropy w.r.t. the method [9] as shown in Table 3. It is clearly that our approach is more efficient, with a gain around 0.2785 bpp. We illustrate this flexibility by the example of the cluster containing three spectral components $\{s^{(4)}, s^{(5)}, s^{(6)}\}$. The underlying cost matrix is:

$$W = \begin{pmatrix} 6.9607 & 6.9020 & 6.9994 \\ 7.1430 & 6.8597 & 6.9984 \\ 7.2898 & 6.9109 & 7.0010 \end{pmatrix}.$$

Fig. 3 shows the resulting entropies with our component ordering method that allows in the cluster two references and those obtained with the classical one where a single reference is adopted. With our method, the cluster has an average entropy around 6.9396 bpp, whereas the entropy with the conventional one is 7 bpp.

6. CONCLUSIONS

In this paper, we have designed a two-stage coding method for hyperspectral images. The first stage of spectral clustering employs a performant unsupervised algorithm. Then, within each spectral group, a VLS was applied to exploit the spectral and spatial redundancies. We have improved the performances of such second stage by optimizing the channel ordering required by the hybrid predictor involved in the VLS. The experimental results have indicated a substantial gain w.r.t. nonadaptive two-stage coding technique. Several directions can be investigated to extend this work. In particular, it would be interesting to investigate other spectral features, that could both improve the spectral clustering and the subsequent VLS.

REFERENCES

[1] G. Vane Ed., *Airbone visible/infrared imaging spectrometer (AVIRIS). A description of the sensor, ground data processing facility, laboratory calibration and first results*, Jet Propulsion Laboratory, Pasadena, CA, Publ. 87-38, 1987.

[2] J. Mielikainen and P. Toivanen, "Clustered DPCM for the lossless compression of hyperspectral images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2943-2946, December 2003.

[3] A. C. Miguel, R. E. Ladner, E. A. Riskin, S. Hauck, D. K. Barney, A. R. Askew, and A. Chang, "Predictive coding of hyperspectral images," *Proc. of the NASA Earth Science Technology Conf.*, 2003.

[4] G. Motta, F. Rizzo, and J. A. Storer, "Partitionned vector quantization: application to lossless compression of hyperspectral images," *Proc. of the Data Compression Conf.*, March 2003.

[5] A. Bilgin, G. Zweig, and M. W. Marcellin, "Three-dimensional image compression with integer wavelet transforms," *Applied Optics*, vol. 39, no. 11, pp. 1799-1814, April 2000.

[6] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding", *Image and Video Communications and Processing*, vol. 5022, pp. 1037-1047, 2003.

[7] G. Menegaz and J.-P. Thiran, "Three-dimensional encoding/two-dimensional decoding of medical data, *IEEE Trans. on Medical Imaging*, vol. 22, no. 3, pp. 424-440, March 2003.

[8] A. Benazza-Benyahia, J.-C. Pesquet, and M. Hamdi, "Vector-lifting schemes for lossless coding and progressive archival of multispectral images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2011-2024, September 2002.

[9] K. Siala and A. Benazza-Benyahia,"Hyperspectral image compression through spectral clustering," *Proc. of the IEEE Control, Communications and Signal Processing*, pp. 435-438, 2004.

[10] K. Siala and A. Benazza-Benyahia,"Lossless coding of hyperspectral images through components ordering by minimum bidirected spanning tree," Proc. of the *IEEE Internat. on Electronics, Circuits and Systems*, December 2005.

[11] H. Frigui and R. Krishnapuram, "Clustering by competitive agglomeration," *Pattern Recognition*, vol. 30, no. 7, pp. 1109- 1119, 1997.

[12] N. Boujemaa, "On competitive unsupervised clustering," *Proc. of the Internat. Conf. on Pattern Recognition*, vol. 1, pp. 631-634, 2000.

[13] S. Chuai-Aree, C. Lursinsap, P. Sophatsathit, and S. Siripant, "Fuzzy C Mean: a statistical feature classification of text and image segmentation method," *Proc. of Intern. Conf. on Intelligent Technology 2000*, Assumption University Bangkok, Thailand, pp. 279-284, December 2000.

[14] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *J. of Appl. and Comput. Harmonic Analysis*, vol. 5, no. 3, pp332-369, 1998.

[15] S. R.Tate, "Band ordering in lossless compression of multispectral images," *IEEE Trans. on Computer*, vol. 46, no. 4, pp. 477-483, April 1997.

Table 1: Effective data set employed.

Scene	Suppressed components	Effective *B*
Cuprite	1, 2, [108, 111], [155, 166]	206
Jasper ridge	1, 2, 3, [108, 113], [153, 161], 223, 224	204
Lunar Lake	1, 2, 3, [108, 113], [153, 161], 223, 224	204

Table 2: Entropies (in bpp) corresponding to different codings ways.

	Jasper ridge	Cuprite	Lunar lake
Original	9.4675	9.1824	9.4034
3D	4.7401	4.8234	4.6806
5/3	8.4892	7.8055	7.7483
VLS1	7.8955	7.3012	7.6534
VLS2	**7.6153**	**7.1435**	**7.5430**
VLS3	7.6455	7.1992	7.6065
VLS4	7.6589	7.1999	7.6133

Table 3: Entropies (in bpp) of some spectral bands of Jasper Ridge.

Band	Original	[9]	Our method
17	9.1116	6.2688	6.9326
18	9.5547	8.1648	6.9167
19	9.5385	6.1794	6.9562
20	9.5987	8.2512	6.9446
Average	9.4520	7.2160	**6.9375**

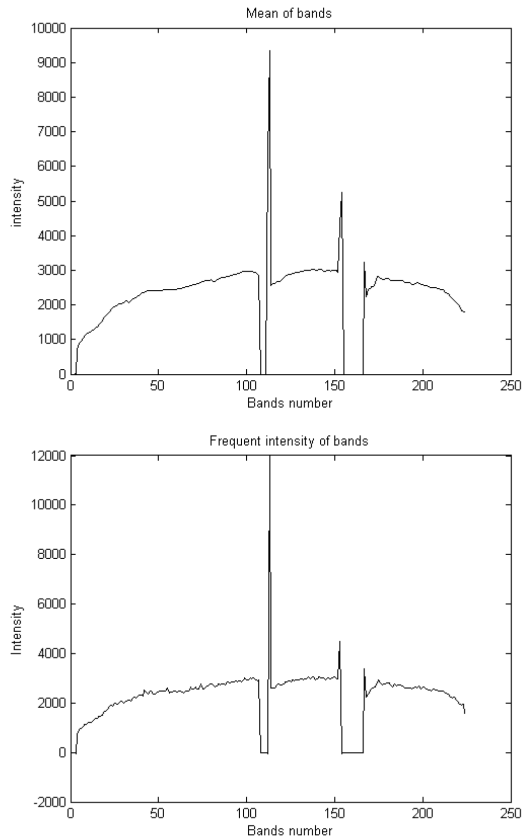

Figure 1: Average and most frequent intensity of Cuprite.

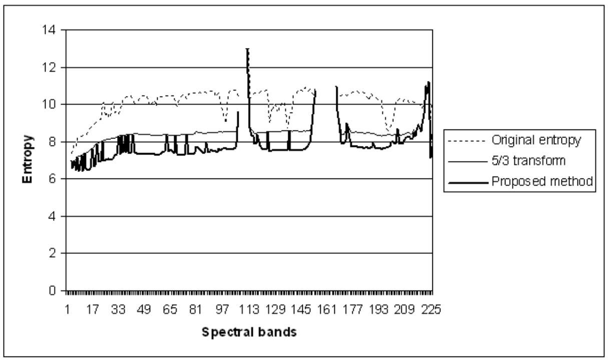

Figure 2: Evolution of the entropy with the spectral component.

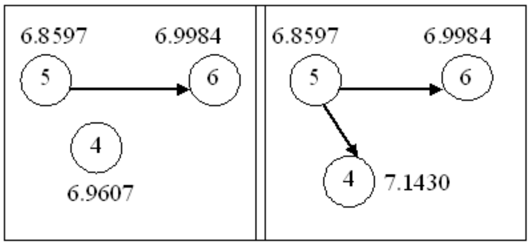

Figure 3: Optimal w.r.t. classic component ordering.