COMPARISON OF MERGING ORDERS AND PRUNING STRATEGIES FOR BINARY PARTITION TREE IN HYPERSPECTRAL DATA

Silvia Valero[1 2], *Philippe Salembier*[1], *Jocelyn Chanussot*[2]

[1]Technical University of Catalonia (UPC), Barcelona, Catalonia, Spain
[2]GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology, Grenoble, France.

ABSTRACT

Hyperspectral imaging segmentation has been an active research area over the past few years. Despite the growing interest, some factors such as high spectrum variability are still significant issues. In this work, we propose to deal with segmentation through the use of Binary Partition Trees (BPTs). BPTs are suggested as a new representation of hyperspectral data representation generated by a merging process. Different hyperspectral region models and similarity metrics defining the merging orders are presented and analyzed. The resulting merging sequence is stored in a BPT structure which enables image regions to be represented at different resolution levels. The segmentation is performed through an intelligent pruning of the BPT, that selects regions to form the final partition. Experimental results on two hyperspectral data sets have allowed us to compare different merging orders and pruning strategies demonstrating the encouraging performances of BPT-based representation.

Index Terms— Binary Partition Tree, Hyperspectral data Segmentation, Merging orders, Pruning strategies

1. INTRODUCTION

A hyperspectral image is set of hundred of images $I_\lambda(p)$ where $\lambda \in \{1, , N\}$ indexes the spectral band wavelength and p indicates the spatial pixel position. This whole set of images can be seen as a three dimensional data cube where each pixel is represented by the spectrum related to the light absorption and/or scattering properties of the spatial region that it represents.

During the last years, the segmentation of these images in meaningful regions has remained a challenging problem due to the huge amount of information contained in their high dimensionality. Thus, new segmentation techniques for hyperspectral data based on morphological profiles [1], Bayesian segmentation [2] and hierarchical segmentation [3] have been published in the recent years. These approaches compute a partition directly from a pixel-based representation of the image which is generally not robust for two following main reasons.

First, the pixels may be corrupted by noise and second, the observation scale is too low to have a good estimation of local as well as global properties. Therefore, we propose to firstly develop a multiscale and region-based representation of the image and, in a second step, to derive the final partition from this representation. Note that this representation is based on regions which will allow a robust estimation of the parameters representing the properties of the data. Moreover, the representation is multiscale: this provides both a local and a global view on the data.

For such a representation, we propose the Binary Partition Tree (BPT)[4], which stores hierarchically a region-based representation in a tree structure.

The BPT construction is often based on an iterative region merging algorithm. Therefore, a good similarity metric is needed to establish the merging order between regions. Working with hyperspectral data, the definition of a spectral similarity distance is not straighforward. Theoretically, the best classical distances between spectra are based on the overall shape of the reflectance curve.

In order to compute a distance between regions, a model for each region must firstly be defined. The simplest solution is to use a first order model: the mean. Hence, classical spectral distances between the mean of each region can be used. Unfortunately, some limitations may come from the poor modeling based on the mean due to the high spectral variability of the pixels belonging to the same class. Therefore, this paper investigates alternative models. We propose to use the histogram as a non parametric model of each region. This requires the definition of a robust distance between histograms.

Once the BPT is constructed, the final partition can be extracted by a pruning strategy which should select the most interesting regions in the tree branches according to a specific criterion. The proposed strategy significantly differs from the classical segmentation by region merging where it is assumed that the final partition can be simply obtained from the iterative merging of regions.

The contributions of this work are as follows: 1) we present a comparison between different region models and similarity metrics to construct a robust hyperspectral BPT in such a way that the most interesting or useful regions are represented, 2) we introduce a new hyperspectral data segmentation approach using pruning strategies on the BPT. The advantages of this approach is demonstrated over the classical iterative merging approach where a simple stopping criterion is used to stop the merging and define the final partition.

The paper is organized as follows. Section 2 reviews the BPT construction and introduces the different similarity measures which have been compared. Section 3 proposes an intelligent BPT pruning strategy in order to perform the segmentation of the hyperspectral data. Experimental results are presented in Section 4, and Section 5 concludes the paper.

2. BPT REGION MODELS AND MERGING ORDER CRITERIA

The BPT is a tree structure which represents an image containing *n* pixels by 2*n*-1 nodes. Leaves nodes represent the original pixels of the image. Furthermore, the remaining tree nodes represent the image regions formed by the merging of their two child nodes corresponding to two neighboring regions. The BPT is generally constructed by an iterative region merging algorithm. As a result, the BPT construction relies on two important choices: a region model M_R and a merging order criterion $O(R_i, R_j)$.

On the one hand, the merging criterion corresponds to the similarity

between neighboring regions which determines the order in which regions are merged and create new nodes in the tree. On the other hand, the region model specifies how an hyperspectral region is represented and how to model the union of two regions.
In our case, two types of region models are studied. The first one consists in a first order statistical model, assuming that all pixels belonging to one region have approximatively a constant spectrum. For the second type, the proposed region model is based on higher order statistics represented by probability density functions (pdfs) [6].
The definition of both models allows us to study two different effects in hyperspectral region merging criterion: first, the importance of the shape of reflectance curves, and, second, the influence of the large spectral variability of the pixels belonging to the same class.

2.1. Constant Region Model

The constant region model M_R is defined as a vector with N components which corresponds to the average of the values of all pixels $p \in R$ in each band λ_k.

$$M_R(\lambda_k) = \frac{1}{N_R} \sum_{p \in R} I_{\lambda_k}(p) \qquad k \in [1, ..., N] \tag{1}$$

with N_R the number of pixels contained in region R.
Using this model, M_R can be modeled as a random variable in the λ dimension by the vector $P_R(\lambda) = (p_{\lambda_1},\ p_{\lambda_2},\ ...,\ p_{\lambda_N})$ where

$$p_{\lambda_k} = M_R(\lambda_k) \;/\; \sum_{k=1}^{N} M_R(\lambda_k) \tag{2}$$

Note that the definition of $P_R(\lambda)$ allows us to propose for $O(R_i, R_j)$ a spectral similarity measure taking into account the overall shape of the reflectance curves. For instance, we propose to use the Spectral Information Divergence [5].

2.1.1. *Spectral Information Divergence*

The Spectral Information Divergence computes the distance between two probability distributions $P_{R_i}(\lambda)$ and $P_{R_j}(\lambda)$. In our case, we use this measure to define the merging criterion by:

$$O_{SID}(R_i, R_j) = \underset{Ri,Rj}{\operatorname{argmin}} \; \{\; D(R_i, R_j) \;+\; D(R_j, R_i) \;\} \tag{3}$$

with $D(R_i, R_j)$ the Kullback Leibler divergence between two probability distributions:

$$D(R_i, R_j) = \sum_{k=1}^{N} P_{R_i(\lambda_k)} \log \frac{P_{R_i(\lambda_k)}}{P_{R_j(\lambda_k)}} \tag{4}$$

2.2. Statistical Region Model

The statistical region model does not assume the homogeneity of the spectral values in a region R. It is defined as a set of N histograms

$$M_R = \{H_R^{\lambda_1}, H_R^{\lambda_2}, ..., H_R^{\lambda_N}\} \tag{5}$$

where each $H_R^{\lambda_k}$ is the empirical spatial distribution of the region R in the band λ_k, formed by:

$$H_R^{\lambda_k} = \{H_R^{\lambda_k}(a_1) H_R^{\lambda_k}(a_2), ...H_R^{\lambda_k}(a_{|\chi|})\} \tag{6}$$

with a_i the possible values of the pixels in each band λ_k.
We must remark that this region model can also be defined when tree leaves are single pixels by exploiting the image self-similarity. The probability density function for individual pixels can actually be estimated and the precise modeling of the pixel's pdf is important in order to get very precise region contours [7].
Following the statistical analysis, two different similarity metrics between histograms are proposed as merging criterion. The first metric corresponds to the classical bin-to-bin Battacharya Coefficient metric [6], which assumes that the histograms are already aligned. The second one corresponds to the cross-bin diffusion distance [8] and is known to be less sensitive to quantization, noise effect and histogram misalignment.

2.2.1. *Battacharya Coefficient*

The Bhattacharyya coefficient between two adjacent regions R_i and R_j of a given band λ_k is defined by

$$BC(H_{R_i}^{\lambda_k}, H_{R_j}^{\lambda_k}) = -\log(\sum_{s=1}^{|\chi|} H_{R_i}^{\lambda_k}(a_s)^{\frac{1}{2}} H_{R_j}^{\lambda_k}(a_s)^{\frac{1}{2}}) \tag{7}$$

If the two pdfs overlap perfectly, the Bhattacharyya coefficient is 0. Consequently, a merging criterion of a pair of adjacent regions can be defined as the sum of the N dissimilarity measures obtained for the different bands:

$$O_{BC}(R_i, R_j) = \underset{Ri,Rj}{\operatorname{argmin}} \; \sum_{k=1}^{N} BC(H_{R_i}^{\lambda_k}, H_{R_j}^{\lambda_k}) \tag{8}$$

2.2.2. *Diffusion Distance*

The diffusion distance [8] is computed by constructing a Gaussian pyramid diffusion process. This process consists in convolving a Gaussian filter $\phi(x, \sigma)$ with the histogram difference $d_l(x)$, where $x \in \mathbb{R}^m$ is a vector.

$$d_0(x) = H_{R_i}^{\lambda_k}(x) - H_{R_j}^{\lambda_k}(x) \tag{9}$$

$$d_l(x) = [d_{l-1}(x) * \phi(x, \sigma)] \downarrow_2 \quad l \in [1, ..., L] \tag{10}$$

The notation $\downarrow_2$ denotes downsampling by a factor of two. L is the number of pyramid layers and σ is the constant standard deviation for the Gaussian filter ϕ.
From the Gaussian pyramid, a distance K between the histograms can be computed summing up the L1 norms of the various levels:

$$K(H_{R_i}^{\lambda_k}, H_{R_j}^{\lambda_k}) = \sum_{l=0}^{L} |d_l(x)| \tag{11}$$

Consequently, the proposed merging criterion using the diffusion distance defined in previous equations is derived as:

$$O_{DD}(R_i, R_j) = \underset{Ri,Rj}{\operatorname{argmin}} \; \sum_{k=1}^{N} K(H_{R_i}^{\lambda_k}, H_{R_j}^{\lambda_k}) \tag{12}$$

To conclude this section, we would like to remark that the area of the regions is not included in any proposed merging order. Thus, these approaches may suffer from small and meaningless regions into the generated partition. To overcome this limitation, we propose to give priority to the fusion between small regions.The approach consists in forcing the merging of regions that have an area smaller than a given percentage (typically 15%) of the average size of the regions created by the merging process [6].

3. BPT PRUNING

In a BPT segmentation framework, a pruning strategy can be interpreted as an image filtering tool that aims to achieve an optimal segmentation inside a set of hierarchical partitions. This corresponds to a BPT simplification technique where the resultant tree contains fewer nodes than the original. The task consists in evaluating some regions (or nodes) cost by performing an analysis running from the leaves to the tree root. Moving along tree branches, the pruning analyzes the cumulative cost e_C associated to each BPT node. Setting a threshold δ_T, the pruning then removes all nodes having a cost lower than δ_T. In the following, we compare two different pruning strategies in order to corroborate the efficiency of the BPT representation for the analysis of hyperspectral images.

3.1. Number of Regions as Pruning Criterion

A classical pruning criterion is the number of regions in the BPT following the merging sequence. In this case, the $e_C(N_k)$ associated to a node N_k is the number of non merged nodes and δ_T corresponds to the desired final number of regions.

This technique implies that a segmentation result of an arbitrary number of regions R can be found by undoing the last $R-1$ mergings over the initial partition. Note that using this pruning strategy, the partition result is always a partition that was obtained following the region merging sequence. Thus, this strategy follows the classical approach of iterative region merging segmentation that assumes that the best partition can be obtained by stopping the iterative merging at some point. In practice, this assumption is rarely true. Consequently, we propose a more intelligent pruning algorithm which is not relying so much on the merging order.

3.2. Region homogeneity as Pruning Criterion

The next pruning strategy is based on maximizing the homogeneity of the regions given a BPT branch. A classical criterion to quantify the homogeneity of a cluster is the intra and inter variance of the region. The proposed strategy consists in evaluating first the homogeneity of a node N_k in relation with its sibling N_s:

$$Hom(N_k) = \sum_{p \in N_k} \|N_k(p) - \overline{N_k}\| + \sum_{p \in N_k} \|N_k(p) - \overline{N_s}\| \quad (13)$$

with N_s the sibling of N_k, and $\overline{N_k}$, $\overline{N_s}$ the mean of all the pixels belonging to these regions. Using $Hom(N_k)$, we define a cumulative error cost e_C taking into account previous errors:

$$e_C(N_k) = \sum_{i=1}^{K} Hom(N_i) \quad (14)$$

Given $e_C(N_k)$, the pruning strategy consists in detecting the rate of change at each node along the branch [10]. To obtain the rate of change, one of the most popular choices is to calculate the second derivatives. It is estimated by the following equation:

$$e''_C(N_k) = e_C(N_{k-}) + e_C(N_{k+}) - 2 * e_C(N_k) \quad (15)$$

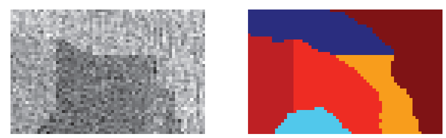
where $e_C(N_{k-})$ is the previous cost along the branch and $e_C(N_{k+})$ is the next one (corresponding to parent node cost).

Hence, the pruning removes a node N_k if $e''_C(N_k) < \delta_T$ and if all its ancestors also satisfy this condition. The segmentation is thus constructed by selecting the leaf nodes of the resulting pruned tree.

4. EXPERIMENTAL RESULTS

In this section we provide a complete evaluation of the merging order criteria of Section 2. First experiments have been performed using a hyperspectral image created manually. Using the partition shown in Fig.1(b), we have filled each region with a set of spectra belonging the same class material. The used spectra correspond to the ground truth of Indian Pines AVIRIS hyperspectral data, containing 200 spectral bands. Fig.1(a) illustrates one of these bands, reflecting the wide variability between spectra of the same class.

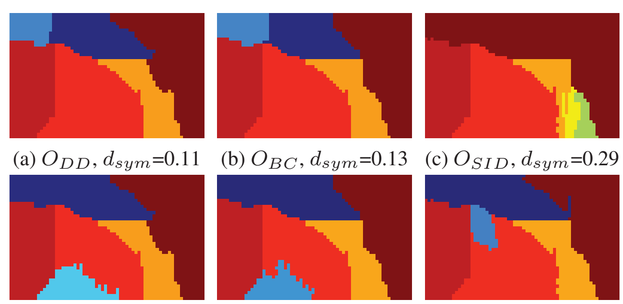
Firstly, three BPT concerning Fig.1 are constructed using the proposed merging criteria presented in Section 2. In statistical region models, the number of bins is set to 200. This value changes for image bands having a lower range difference. In this case, the *minimum* range difference corresponds to the N_{bins}. On the other hand, the *maximum* number of bins is set to $N_{bins} = 200$. Regarding the diffusion distance, we set the gaussian kernel deviation $\sigma = 0.5$ and use a window size 3. After the construction, the two pruning strategies of section 3 provide different segmentation results for each BPT. Hence, the final partitions obtained by BPT pruning are used to evaluate the performances of the merging orders and of the pruning strategies.



(a) Band 2 (b) Ideal Partition

Fig. 1: Manually created test image, with real sets of spectra

To evaluate the resulting partitions, the symmetric distance d_{sym} [9] is proposed as a quality evaluation. Having a partition P and a ground truth GT, the d_{sym} corresponds to the minimum number of pixels whose labels should be changed between regions in P to achieve a perfect matching with GT, normalized by the total number of pixels in the image. Fig.2 shows the segmentation results obtained by the three merging criteria of Section 2 and the two pruning strategies described in Section 3.



(a) O_{DD}, d_{sym}=0.11 (b) O_{BC}, d_{sym}=0.13 (c) O_{SID}, d_{sym}=0.29

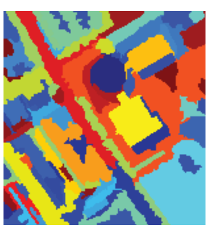
(d) O_{DD}, d_{sym}=0.02 (e) O_{BC}, d_{sym}=0.03 (f) O_{SID}, d_{sym}=0.13

Fig. 2: First row shows the results obtained for the number of regions pruning criterion (δ_T= 6 regions) when different models are used for the construction. Second row shows the results for the region homogeneity pruning criterion, with δ_T=6.10^6

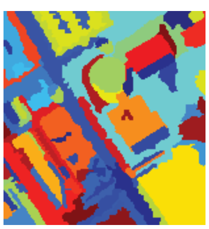
Comparing the results, we observe that the smallest d_{sym} are obtained using the region homogeneity as pruning criterion. This can be corroborated looking at the blue round region located on the bottom, which is only detected in the second row. Note that this result also highlights the importance of using a pruning strategy on the BPT instead of simply stopping an iterative merging algorithm. Regarding merging orders, Fig.2 shows that the statistical region models yield to better results than constant ones. This limitation is due to the importance of the spectrum variability, which makes this last model insufficient. Another important remark is that comparing both statistical region models, O_{DD} is less sensitive to the chosen number of bins and is hence more robust.
As a second experiment, the top-down corner of the ROSIS-03 optical sensor over the University of Pavia is proposed. This image is formed by 103 denoised channels, Fig.3 illustrates a RGB combination of three of them.

Fig. 3: RGB combination of Pavia Rosis data set

For this second image, we have repeated the same procedure obtaining the results of Fig.4. Both pruning results suggest that our merging orders achieve satisfactory results. However, it can be observed that the region homogeneity pruning criterion provides more realistics results. For instance, the first row results do not obtain either the shape of the building located at the right-bottom corner or the metal objects on the roof of the central rounded-squared building.

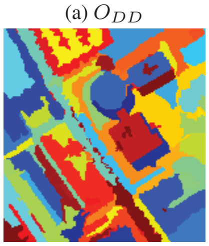


(a) O_{DD}

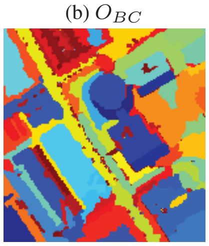


(b) O_{BC}

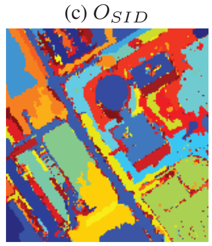
(c) O_{SID}



(d) O_{DD}



(e) O_{BC}



(f) O_{SID}

Fig. 4: First row shows the results obtained for the number of regions pruning criterion (δ_T= 85 regions) when different models are used for the construction. Second row shows the results for the region homogeneity pruning criterion

5. CONCLUSIONS

In this paper, BPTs have been proposed as a new type of hierarchical representation for hyperspectral imagery. These BPTs represent a set of regions structured by inclusion in a tree, which can be pruned for efficient hyperspectral image segmentation. BPT construction for hyperspectral data has been introduced comparing different merging orders. Being the hyperspectral data clustering straightforward, the use of statistical region models has demonstrated a high degree of reliability. Regarding pruning strategies, results have shown that the optimal segmentation can be rarely found simply following the merging sequence. For this reason, the region homogeneity pruning criterion clearly outperforms the straightforward pruning criterion based on the number of regions. Hence, BPT shows that the tree structure allows a robust, generic and also reliable segmentation.The performances of the proposed method were assessed using a quantitative evaluation based on a synthetic image using real spectra measured by the AVIRIS sensor, and a qualitative evaluation based on a real hyperspectral image acquired by the sensor ROSIS.

6. REFERENCES

[1] A. Plaza, J.A.Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, J. Tilton and G. Trianni.“Advanced Processing of Hyperspectral Images”, Remote Sensing of Environment, DOI: 10.1016/j.rse.2007.07.028, 2009.

[2] J. Li, J.M. Bioucas-Dias and A. Plaza, “Semi-supervised hyperspectral image segmentation”, IEEE GRSS Workshop on Hyperspectral Image-WHISPERS’2009, Grenoble 2009.

[3] J. A. Gualtieri and J.C. Tilton,“ Hierarchical Segmentation of Hyperspectral Data”, 2002 AVIRIS Earth Science and Applications Workshop Proceedings, 2002, pp 58.

[4] P. Salembier and L. Garrido, “ Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval” IEEE Trans. Image Processing, 2000, vol. 9, no. 4, pp. 561576.

[5] C.-I Chang, “Spectral information divergence for hyperspectral image analysis,” in Int. Geoscience and Remote Sensing Symp.99, Hamburg, Germany, June 28, 1999.

[6] F. Calderero and F. Marques,“ Region Merging Techniques Using Information Theory Statistical Measures”, IEEE Transactions on Image Processing, vol. 19, n 6, pp. 1567 - 1586, june 2010.

[7] M. Dimiccoli and P. Salembier,“ Hierarchical region-based representation for segmentation and filtering with depth in single images”, in ICIP 2009, November 2009, Cairo, EGYPT.

[8] H. Ling and K. Okada“ Diffusion distance for histogram comparison”. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006.

[9] J. Cardoso and L. Corte-Real, “Toward a generic evaluation of image segmentation,” IEEE Trans. Image Process., vol. 14, no. 11, pp. 17731782, Nov. 2005.

[10] H. Lu, J.C. Woods, and M. Ghanbari, “Image segmentation by binary partition tree,” Electronics Letters , vol.42, no.17, pp. 966-967, August 17, 2006