IMPROVED BINARY PARTITION TREE CONSTRUCTION FOR HYPERSPECTRAL IMAGES: APPLICATION TO OBJECT DETECTION

Silvia Valero[1][2], Philippe Salembier[1], Jocelyn Chanussot[2], Carles M. Cuadras[3]

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

ABSTRACT

This paper discusses hierarchical region-based representation using Binary Partition Tree in the framework of hyperspectral data. Based on region merging techniques, this region-based representation reduces the number of elementary primitives compared to the pixel-based representation and allows a more robust filtering, segmentation, classification or information retrieval. The work presented here proposes a strategy for merging hyperspectral regions using a new association measure depending on canonical correlations relating principal coordinates. To demonstrate an example of BPT usefulness, a pruning strategy aiming at object detection is discussed. Experimental results demonstrate the good performances of BPT.

Index Terms— Hyperspectral imaging, Binary Partition Tree, canonical correlations, segmentation, object detection

1. INTRODUCTION

Hyperspectral sensors collect multivariate discrete images in series of narrow and contiguous wavelength bands. The resulting datasets enable the characterization of regions based on their spectral properties which provides a rich amount of information. Conventional analysis techniques have traditionally considered these images as an unordered array of spectral measurements. In the last few years, the importance of the spatial information considering, in particular, spatial correlation and connectivity in the image has been proved. This information turns out to be essential to interpret objects in natural scenes. Hence, hyperspectral analysis tools should take into account both the spatial and the spectral spaces. However, the number of wavelengths per spectrum and pixel per image as well as the complexity of handling spatial and spectral correlation explain why this approach is still a largely open research issue. Recently, an abstraction from the pixel-spectrum-based representation has been proposed using Binary Partition trees (BPT) [1]. This representation [2] stores hierarchically a region-based representation in a tree structure. This provides a hierarchy of regions at different levels of resolution to cover a wide range of applications. This generic representation can be based on an iterative region merging algorithm. Hence, a good region similarity metric and a region model are needed to establish the merging order between hyperspectral regions.

Working with hyperspectral data, the definition of both concepts is not straighforward. Regarding the region model, a non parametric statistical model is used (that is a histogram) [3]. This leads us to define a robust distance between histograms taking into account the correlation between bands. The work presented here introduces a new association measure depending on canonical correlations relating principal coordinates.

The BPT is a generic representation, indepently from its construction, it can be used in many different applications such as segmentation [4], classification [1], indexing, filtering, compression or object recognition. This paper focuses on object detection and recognition as this is an important challenge in remote sensing images. The automated selection of results in hierarchical segmentations combining spectral/spatial information has been previously studied [5]. Despite of some interesting results, problems regarding under and over-segmentation remained. Our study also proposes a technique to extract results combining spectral and spatial features. However, the object extraction is based on the study of the descriptors computed for each node of the BPT.

In this paper, experiments are conducted to evaluate the performance of the BPT construction. Moreover, a second set of experiments studies a powerful object detection using the BPT representation The organization of this paper is as follows: Section 2 gives a brief introduction on BPT, explaining the details of its construction and proposes the new merging criterion. The BPT pruning for object recognition is discussed in section 3. Experimental results are shown in section 4. Finally, conclusions are drawn in section 5.

2. CONSTRUCTION OF THE BPT

From an image containing *n* pixels, a BPT generates a tree structure containing *2n-1* nodes. In this tree representation, three types of nodes can be found: Firstly, leaves nodes representing the original image pixels, secondly, the root node representing the entire image support and finally, the remaining tree nodes representing image regions formed by the merging of their two child nodes corresponding to two adjacent regions. A possible way to construct a BPT is to use an iterative region merging algorithm that merges, at each step, the pair of most similar neighboring regions. The BPT is then built by keeping track of the merging steps. Following an iterative region merging algorithm, the most similar adjacent regions are merged at each step. Fig. 1 shows an example of BPT construction starting from an original partition. In the sequel, this initial partition will be the partition of individual pixels.

The creation of BPT relies on two important notions. The first one is the *region model* M_R which specifies how regions are represented and how to model the union of two regions. The second notion is the *merging criterion* $O(R_i, R_j)$, which defines the similarity between neighboring regions and hence determines the order in which regions are merged.
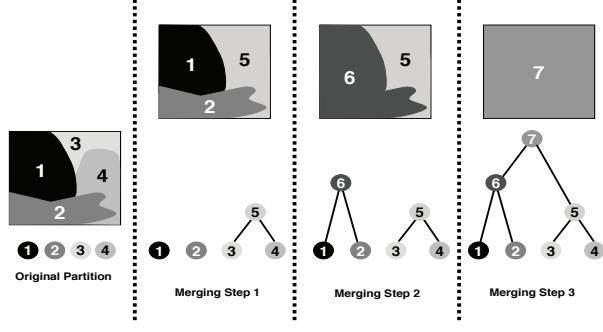
Figure 1: Example of BPT construction using a region merging algorithm

2.1. Region Model: Non-parametric statistical model

This region model M_R assumes that a region is a set of connected pixels with independent identically distributed (i.i.d) spectral values characterized by the corresponding probability distribution [3]. Then, the region model is represented by a set of non parametric probability density functions (pdfs) with no assumptions about the nature of the regions nor the shape of the pdfs. Using this model with an hyperspectral image containing $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ bands, regions are modeled as N arbitrary discrete distributions, directly estimated from the pixel values.

$$M_R = \{H_R^{\lambda_1}, H_R^{\lambda_2}, ..., H_R^{\lambda_N}\} \qquad (1)$$

Fig. 2 shows the non parametric statistical model interpretation. It can be observed how M_R is a matrix where each cell represents the probability of the region pixels to have a radiance value a_s in a specific band λ_k. The region model is formed by the rows of the matrix $H_R^{\lambda_k}$. It corresponds to the empirical spatial distribution (histogram) of the region R in the band λ_k.
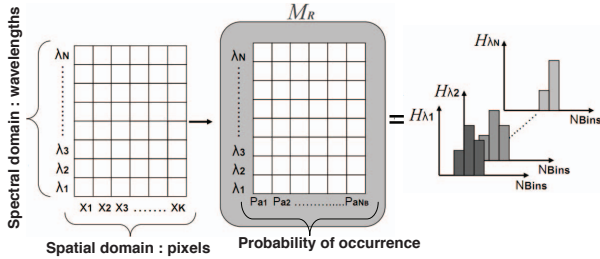

Figure 2: Non parametric statistical model interpretation

Note that this model can also be defined when tree leaves are single pixels by using the image self-similarity [6].

2.2. Merging criterion: Association measure via Multidimensional Scaling

We are interested in defining a measure of association between two non parametric statistical models defined by M_{R_i} and M_{R_j}. The proposed measure is based on the distances between wavebands and canonical correlations[7]. The main idea is to analyze the inter-waveband similarity relationships for each data set M_R via metric scaling and principal coordinates, and then to establish an association measure correlating the principal axis of both data sets obtained via multidimensional scaling. Fig.3 summarizes the proposed procedure.
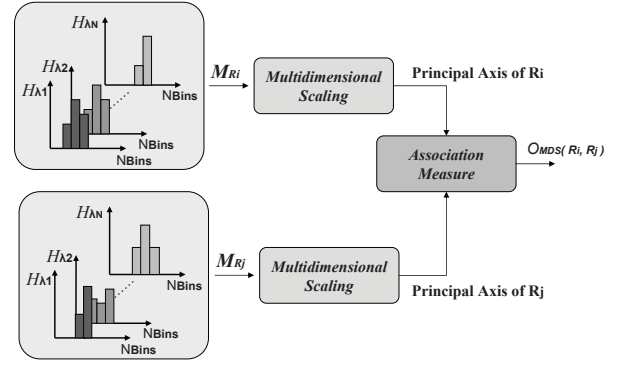

Figure 3: Association measure via Multidimensional Scaling

Multidimensional scaling (MDS) [8] represents a set of objects as a set of points in a map of chosen dimensionality, based on their interpoint distances. The objective is to maximize the agreement between the displayed interpoint distances and the given ones. Thus, MDS attempts to locate n objects as points in Euclidean space E where the geometric differences between pairs of points in E agree, as closely as possible, with the true differences between the n objects.

In our case, the n objects correspond to the N probability distributions of each M_R. Thus, the probability distribution similarities (or dissimilarties) of M_R can be represented by a N x N distance matrix $\Delta_R = (\delta_{kl})$, where $\delta_{kl} = \delta_{lk} \geq 0$ is computed by

$$\delta_{kl} = e^{(K(H_R^{\lambda_k}, H_R^{\lambda_l}))} - 1 \qquad (2)$$

where $K(H_R^{\lambda_k}, H_R^{\lambda_l})$ is the diffusion distance [9] measured between the probability distributions k and l.

Hence, being A the matrix with entries $A = -(\frac{1}{2})\delta_{kl}^2$ and the centering matrix $H = I_n - \frac{1}{n}11'$, the so-called inner product matrix B_R associated to Δ_R can be computed by $B_R = HAH$ for each M_R[8]. The inner product matrix B_R is NxN symmetric matrix which can be spectrally decomposed as $B_R = U_R\Lambda_R^2 U_R'$. Assuming the eigenvalues in Λ_R are arranged in descending order, the matrix U_R contains the standard coordinates of region R where the s first columns are the most representatives coordinates. The aim of MDS is to represent M_R in a reduced dimension, by taking the first standard coordinates.

Given two regions defined by M_{R_i} and M_{R_j}, our interest is to measure the multivariate association between their s first standard coordinates. Therefore, two distance matrices Δ_{R_i} and Δ_{R_j} to find $B_{R_i} = U_{R_i}\Lambda_{R_i}^2 U_{R_i}'$ and $B_{R_j} = U_{R_j}\Lambda_{R_j}^2 U_{R_j}'$ should be computed using the explained procedure.

The number s of dimensions is an important aspect in most multivariate analysis methods. In MDS, the number of dimensions is based on the percentage of variability accounted for by the first dimensions. Here, a criterion which extends a sequence c defined and studied in [10] is used to set the value of s. Firstly, a maximum dimension N_s suggested by the data should be fixed. Then, being u_i and v_i, $i = 1, ..., N_s$, the first N_s columns of U_{R_i} and U_{R_j}, the sequence c_k is defined as

$$c_k = \frac{\sum_{t=1}^{k}\sum_{p=1}^{k} \lambda_{tR_i}^2 (u_t'v_p)^2 \lambda_{tR_j}^2}{\sum_{t=1}^{N_s}\sum_{t=1}^{N_s} \lambda_{tR_i}^2 (u_t'v_p)^2 \lambda_{tR_j}^2} \qquad k \in [1, ..., N_s] \qquad (3)$$

$\lambda^2_{tR_i}$ $\lambda^2_{tR_j}$ are the eigenvalues of B_{R_i} and B_{R_j} which are proportional to the variances of the corresponding principal axes. Here N_s is the minimum dimension for which $\frac{\sum_{t=1}^{N_s} \lambda^2_{tR}}{\sum_{t=1}^{N} \lambda^2_{tR}} \approx 1$ and $(u'_t v_p)^2$ is just the correlation coefficient between the t-th and p-th coordinates. Thus the numerator in c_k is a weighted average of the relationships between principal axes. Clearly $0 \leq c_1 \leq, ... \leq c_s \leq , ... \leq c_{N_s} = 1$. The dimension s is then chosen such that C_s is high, for instance $C_s = 0.9$. At this point, having two regions defined by their standard coordinates U_{R_i} and U_{R_j} whose dimensions are Nxs, the Wilk's criterion W for testing B=0 in a multivariate regresion model is given by:

$$W(R_i, R_j) = det(I - U'_{R_j} U_{R_i} U'_{R_i} U_{R_j}) = \prod_{i=1}^{s} (1 - r_i^2) \quad (4)$$

where det means the determinant and r_i corresponds to the canonical correlation of each axis. Using Eq. 4, an association measure can be defined as:

$$A_W(R_i, R_j) = 1 - W(R_i, R_j) = 1 - \prod_{i=1}^{s} (1 - r_i^2) \quad (5)$$

satisfying $0 \leq A_W(R_i, R_j) \leq 1$ and $A_W(R_i, R_j) = 1$ if R_i is equal to R_j. Thus, this leads us to the definition of the proposed merging criterion:

$$O_{MDS}(R_i, R_j) = \underset{Ri, Rj}{\operatorname{argmin}} \ 1 - A_W(R_i, R_j) \quad (6)$$

3. OBJECT DETECTION WITH BPT

Hyperspectral object detection has been mainly developed in the context of pixel-wise spectral classification. The main problem of this approach is that objects are not only characterized by their spectral signature. Indeed, spatial features such as shape, area, orientation, etc., can also contribute significantly to the detection.
In order to integrate the spatial and spectral information, BPT is proposed as a search space for constructing a robust object identification scheme. The strategy is to analyze the BPT using a set of descriptors computed for each node. The work presented here proposes the analysis of three different descriptors for each node:

$$\mathcal{D} = \{\mathcal{D}_{shape}, \mathcal{D}_{spectral}, \mathcal{D}_{area}\} \quad (7)$$

The proposed shape, spectral and area descriptors are related to the specficic object of interest. Studying $\mathcal{D}$ from the leaves to the root, the approach consists in removing all nodes that significantly differ from the characterization proposed by a reference $\mathcal{D}_{ref}$. Hence, given this reference, the idea consists in considering that the searched object instances are defined by the closest nodes to the root node that have descriptors close to the $\mathcal{D}_{ref}$. model. In order to illustrate the generality of the approach, we describe two detection examples: roads and building in urban scenes.

3.1. Detection of roads

Roads appear as elongated structures having fairly homogeneous radiance values usually corresponding to asphalt. Given their characteristic shape, $\mathcal{D}_{shape}$ is the elongation of the region. In order to compute it, we first define the smallest rectangular bounding box

allowed to have an arbitrary orientation that includes the region corresponding to the BPT node under study. $\mathcal{D}_{shape}$ is defined as the ratio between the width and the height of the bounding box. $\mathcal{D}_{spectral}$ is the correlation coefficient between the mean spectrum of the BPT node and a reference spectrum of the road material. $\mathcal{D}_{area}$ is defined as the area of the region which should be higher than the minimum area required to consider that a region can be a road. This minimum parameter should be set according to the spatial resolution of the image. Computing $\mathcal{D}$ for all BPT nodes, the idea is to select the father nodes closer to the root which have a low elongation, a high correlation between asphalt material and an area higher than a threshold.

3.2. Detection of buildings

Following the same strategy, we consider that the rectangularity of a region is an important characteristic for buildings. Then, a rectangularity descriptor is proposed as $\mathcal{D}_{shape}$. It is computed by the ratio of the area of the BPT node and the area of the smallest bounding box including the region (as in the previous application, the bounding box can have arbitrary orientation). $\mathcal{D}_{shape}$=1 for perfectly rectangular regions. $\mathcal{D}_{spectral}$ also corresponds to the correlation coefficient measured between the mean spectrum of the BPT node and a reference spectrum of the building material. $\mathcal{D}_{area}$ is the required area to consider that a region can be a building.. The procedure of the detection is the same as the roads.

4. EXPERIMENTAL RESULTS

We first provide an evaluation of the BPT construction proposed in Section 2. The experiments have been performed using a portion of a publicy available HYDICE hyperspectral. After removing water absorption and noisy bands, the data contain 167 spectral bands. Fig. 4(a) shows a RGB combination of three of them. The BPT is computed by the procedure described in Section 2. The number of bins to represent the histograms depends of the image range (here $N_{bins} = 256$). The first component dimension found by the sequence c_k is $s = 3$. To evaluate the quality of the BPT construction, we extract a segmentation result involving a given number N_R of regions by undoing the last $N_R - 1$ mergings over the initial partition. In this section, we compare the partition extracted from the BPT with the classical RHSEG [11]. In the case of RHSEG, the similarity criterion used is SAM with spectral clustering weight 0.1[12]. To evaluate the resulting partitions, the symmetric distance d_{sym} [13] is used as a partition quality evaluation. Having a partition P and a ground truth GT, the symmetric distance corresponds to the minimum number of pixels whose labels should be changed in partition P to achieve a perfect matching with GT, normalized by the total number of pixels in the image. The manually created GT is shown in Fig. 4(b).
Fig. 4(c)(d) show the segmentation results obtained with BPT and RHSEG, respectively. In both cases, the resulting partitions involve 63 regions. Comparing both results, the quantitive d_{sym} and the visual evaluation corroborates that the partition extracted from the BPT are much closer to the ground truth than the one computed with RHSEG. This experiment has been done with several other images acquired by different hyperspectral sensors, but, due to space limitations, we can only present one data set. However, the conclusions were the same on the remaining dataset.
A second set of experiments are conducted now for object detection and recognition. We compare the classical pixel-wise classification against the strategy proposed in section 3 for building and road detection. The pixel-wise classification consists in detecting
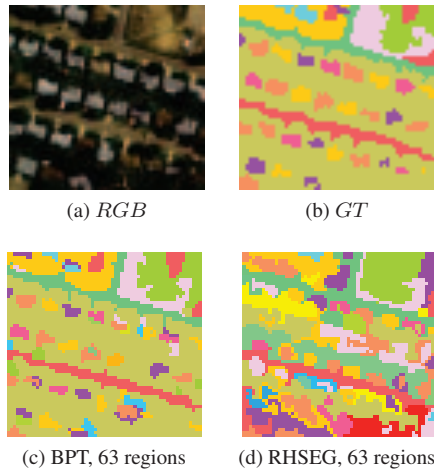
(a) *RGB*   (b) *GT*

(c) BPT, 63 regions   (d) RHSEG, 63 regions

Figure 4: (a) Urban Hydice RGB Composition, (b) Manually created Ground Truth, (c) Partition extracted from BPT leading to d_{sym} =25, (d) Partition computed with RHSEG [11] leading to d_{sym} =70

pixels in the image having a correlation higher than 0.9 with a reference spectrum. The first row of Fig. 5 shows the results of road and building detection having a high correlation coefficient. The second row of Fig. 5 illustrates the detected BPT nodes using the two different $\mathcal{D}_{ref}$. In the case of roads, we are looking for regions having an elongation lower than 0.3, with a correlation coefficient with asphalt higher than 0.9 and an area larger than 15 pixels. Regarding the buildings, we are looking for regions having a rectangularity higher than 0.4, a correlation coefficient between building material higher than 0.9 and an area smaller than 15 pixels. The results obtained with the BPT corroborates the advantage of using this image representation. The use of spectral as well as spatial descriptors of BPT nodes outperforms the classical pixel-wise detection using only spectral information.
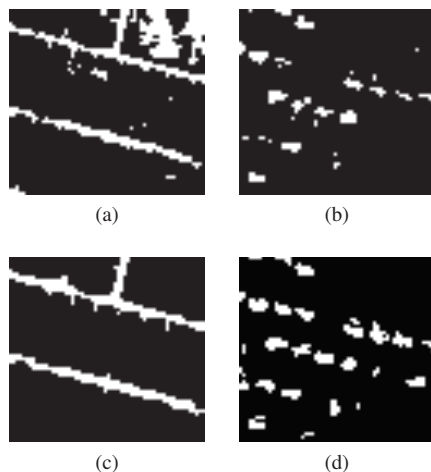

(a)   (b)

(c)   (d)

Figure 5: First row: Road (a) and Building (b) detection using pixel classification. Second row: Road (c) and Building (d) detection based on BPT representation

5. CONCLUSIONS

In the context of hyperspectral images, Binary Partition Tree construction and processing have been discussed in this paper. Con-

cerning the construction, two concepts have been highlighted to define the recursive merging algorithm. The first concept is the use of statistical region models which efficiently deal with the the problem of spectral variability for clustering hyperspectral data. The second one is the use of a new distance-based measure depending on canonical correlations relating principal coordinates. The BPT construction has been tested comparing partitions extracted from the BPT to partitions computed by classical hierarchical segmentation techniques. The results demonstrated that regions contained in the BPT represent regions containing a semantic content of the image. Hence, the BPT enables the definition of a hierarchically structured set of regions to perform object-based recognition. Here, examples of building and road detections on BPT have been presented. The results clearly outperform pixel-wise classification and demonstrate the genericity of the representation.

6. REFERENCES

[1] S. Valero, P. Salembier, and J. Chanussot, "New hyperspectral data representation using binary partition tree," in *Proc.of IGARSS10, Honolulu, Hawaii, USA*, 2010.

[2] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," in *IEEE Trans. Image Processing*, 2000, vol. 9, pp. 561–576.

[3] F. Calderero and F. Marques, "Region merging techniques using information theory statistical measures," IEEE Trans. Image Processing, 2010, vol. 19, pp. 1567–1586.

[4] S. Valero, P. Salembier, and J. Chanussot, "Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data," in *Proc.of ICIP10, Hong Kong*, 2010.

[5] A.J. Plaza and J. Tilton, "Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images," in *Proc.of IGARSS05*, 2005.

[6] M. Dimiccoli and P. Salembier, "Hierarchical region-based representation for segmentation and filtering with depth in single images," in *Proc. of ICIP, Cairo, Egypt*, 2009.

[7] E. Cramer and W.A. Nicewander, "Some symmetric, invariant measures of multivariate association," in *Psychometrika*, 1979, vol. 44, pp. 43–54.

[8] T.F. Cox and M.A.Cox, "Multidimensional scaling," K. Fernandez and A. Morineau (Ed.),Chapman & Hal, 1994.

[9] H. Ling and K. Okada, "Diffusion distance for histogram comparison," Proc. of CVPR, 2006.

[10] C.M. Cuadras, A. Arenas, and J. Fortiana, "Some computational aspects of a distance-based model for prediction," in *Communications in Statistics: Simulation and Computation*, 1996, vol. 25, pp. 593–609.

[11] J. A. Gualtieri and J.C. Tilton, "Hierarchical segmentation of hyperspectral data," AVIRIS Earth Science and Applications Workshop Proceedings, 2002, pp. 5–8.

[12] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "A multiple classifier approach for spectral-spatial classification of hyperspectral data," in *Proc.of IGARSS10, USA*, 2010.

[13] J. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," in *IEEE Trans. Image Processing*, 2005, vol. 14, pp. 1773–1782.