Binary Partition Tree for hyperspectral remote sensing images

S. Valero$^{1,2}$, P. Salembier$^2$ and J. Chanussot$^1$

$^1$GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology, Grenoble, France.
$^2$Signal Theory and Communications Department, Technical University of Catalonia, Spain
Hyperspectral Imagery

Hyperspectral data cubes contain hundreds of contiguous waveband images.

Hyperspectral image processing is still a difficult endeavor due to the huge amount of data involved.

From the pixel representation, different analysis techniques have been proposed in the literature.

\[ I_\lambda = (I_{\lambda_1}, I_{\lambda_2}, \ldots, I_{\lambda_B}) \]
Hyperspectral Imagery

- The initial pixel-based representation is a very low level and unstructured representation.
- This implies that classification, segmentation or detection techniques are not very robust.

- We propose **Binary Partition Trees (BPTs)** as a new region-based hierarchical representation for hyperspectral images.
- BPTs can be interpreted as a set of hierarchical regions stored in a tree structure
- A multiscale tree representation allows us to cover a wide range of applications.
The BPT is a hierarchical tree structure representing an image.

The tree leaves correspond to the initial pixel level partition.

The remaining tree nodes represent regions formed by the merging of two children regions.

The tree construction is performed by keeping track of merging steps of an iterative region merging algorithm.
The creation of BPT implies two important notions:

- **Region model** $M_{R_i}$
  The region model specifies how an hyperspectral region is represented and how to model the union of two regions.

- **Merging criterion** $0(R_i, R_j)$
  The similarity between neighboring regions determining the merging order.
How to represent hyperspectral image regions?

Which similarity measure defines a good merging order?
From the tree representation, many partitions can be extracted for various applications.

The processing of BPT will then involve an application dependant pruning strategy.
Outline

1. Introduction

2. BPT Construction
   - Region Model
   - Merging Criterion

3. BPT Pruning
   - Classification
   - Segmentation
   - Object Detection

4. Conclusion
How to represent hyperspectral image regions?

Which similarity measure defines a good merging order?
The simplest solution is modelling each region by a first order model such as the spectra mean of the region

\[ M_R(\lambda_k) = \frac{1}{N_R} \sum_{p \in R} l_{\lambda_k}(p) \quad k \in [1, ..., N] \]

Assuming that all pixels belonging to one region have approximately a constant spectrum, classical spectral distances between the mean of each region can be used:

Spectral Angle Mapper, Spectral Information Divergence, Correlation,..
A strong interclass spectra variability can be found in each image band due to noise and mixed pixels.

The homogeneity of the spectral values cannot be assumed in a region.
A strong interclass spectra variability can be found in each image band due to noise and mixed pixels.

The homogeneity of the spectral values cannot be assumed in a region.
We propose to use a non-parametric statistical region model which is represented by a set of $N$ probability density functions

$$M_{Ri} = \{P_{Ri}^1, P_{Ri}^2, ..., P_{Ri}^N\}$$

$$P_{Ri}^k = \{P_{Ri}^k(a_1), P_{Ri}^k(a_2), ....P_{Ri}^k(a_{|\chi|})\}$$

being $a_i$ the possible values of the pixels in each band $k$. 
Statistical Region Model

\[ M_{R_i \cup R_j} = \{ P^1_{R_i \cup R_j}, P^2_{R_i \cup R_j}, \ldots, P^N_{R_i \cup R_j} \} \]

where \( k \) probability function of the union of two regions can be defined as

\[ P^k_{R_i \cup R_j} = \frac{N_i}{N_i + N_j} P^k_{R_i} + \frac{N_j}{N_i + N_j} P^k_{R_j} \]

being \( N_i \) and \( N_j \) the area of the regions \( R_i \) and \( R_j \).
How to represent hyperspectral image regions?

Which similarity measure defines a good merging order?

Silvia Valero et al.
Silvia.Valero-Valbuena@gipsa-lab.grenoble-inp.fr
Following the statistical analysis, we need a similarity measure which allows us to compare probability distributions. An example of measure is the Bhattacharyya Coefficient which measures the overlap between two probability distributions.

\[ O(R_i, R_j) = \arg\min_{R_i, R_j} \sum_{k=1}^{N} BC(P^k_{R_i}, P^k_{R_j}) \]
In order to exploit spectral information, a covariance matrix $\Sigma_{M_{Ri}}$ can be estimated considering $M_{Ri}$ as a set of relative frequencies $P_{ij}$.
Introduction
BPT Construction
BPT Pruning
Conclusion

Region Model
Merging Criterion

Pavia University

50 regions using BC
120 regions using BC

50 regions using MAHA
120 regions using MAHA

Silvia Valero et al.
Silvia.Valero-Valbuena@gipsa-lab.grenoble-inp.fr
Outline

1 Introduction

2 BPT Construction
   - Region Model
   - Merging Criterion

3 BPT Pruning
   - Classification
   - Segmentation
   - Object Detection

4 Conclusion
It can be interpreted as a filtering tool that aims to achieve a goal inside a set of hierarchical partition.

BPT nodes can be evaluated measuring a specific region descriptors for each node.
In a classification context, region descriptors can correspond to a classifier output. It consists in removing subtrees composed of nodes belonging to the same class. An increasing iterative cost measuring the node impurity can be studied along the branches.
## Classification Pruning

<table>
<thead>
<tr>
<th>Class</th>
<th>Simple SVM</th>
<th>Pruned BPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86.11</td>
<td>94.44</td>
</tr>
<tr>
<td>2</td>
<td>88.39</td>
<td>93.41</td>
</tr>
<tr>
<td>3</td>
<td>83.45</td>
<td>89.03</td>
</tr>
<tr>
<td>4</td>
<td>77.56</td>
<td>80.77</td>
</tr>
<tr>
<td>5</td>
<td>95.18</td>
<td>92.77</td>
</tr>
<tr>
<td>6</td>
<td>97.39</td>
<td>98.39</td>
</tr>
<tr>
<td>7</td>
<td>83.33</td>
<td>88.89</td>
</tr>
<tr>
<td>8</td>
<td>97.85</td>
<td>99.08</td>
</tr>
<tr>
<td>9</td>
<td>64.29</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>84.98</td>
<td>88.70</td>
</tr>
<tr>
<td>11</td>
<td>91.19</td>
<td>96.72</td>
</tr>
<tr>
<td>12</td>
<td>92.93</td>
<td>93.66</td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td>98.59</td>
</tr>
<tr>
<td>14</td>
<td>96.99</td>
<td>99.42</td>
</tr>
<tr>
<td>15</td>
<td>67.72</td>
<td>98.43</td>
</tr>
<tr>
<td>16</td>
<td>95.30</td>
<td>100</td>
</tr>
<tr>
<td>Overall</td>
<td>87.67</td>
<td>94.52</td>
</tr>
</tbody>
</table>

Silvia Valero et al.  
Silvia.Valero-Valbuena@gipsa-lab.grenoble-inp.fr
In a segmentation context, the pruning goal is to detect homogeneous and different regions.

A region descriptor can correspond to the total variance of a node:

\[ V_t(R) = V_{\text{intra}}(R) + V_{\text{inter}}(R) \]

A pruning based on minimizing \( V_{\text{total}}(R) \) along BPT branches can be defined.
In a segmentation context, the pruning goal is to detect homogeneous and different regions.
The BPT has been created in such a way that the most meaningful regions are represented in its nodes.
The BPT has been created in such a way that the most meaningful regions are represented in its nodes.
Spatial and spectral descriptors can be computed in nodes in order to search known objects.
Outline

1. Introduction

2. BPT Construction
   - Region Model
   - Merging Criterion

3. BPT Pruning
   - Classification
   - Segmentation
   - Object Detection

4. Conclusion
BPTs have been proposed as a new region-based and hierarchical representation of the hyperspectral data.

Being a generic representation, many tree processing techniques can be formulated as pruning strategies for many applications.

A solution for the problem of the spectra variability for clustering hyperspectral data has been proposed using statistical region models.