HYPERSPECTRAL IMAGE SEGMENTATION USING BINARY PARTITION TREES

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

The work presented here proposes a new Binary Partition Tree pruning strategy aimed at the segmentation of hyperspectral images. The BPT is a region-based representation of images that involves a reduced number of elementary primitives and therefore allows to design a robust and efficient segmentation algorithm. Here, the regions contained in the BPT branches are studied by recursive spectral graph partitioning. The goal is to remove subtrees composed of nodes which are considered to be similar. To this end, affinity matrices on the tree branches are computed using a new distance-based measure depending on canonical correlations relating principal coordinates. Experimental results have demonstrated the good performances of BPT construction and pruning.

Index Terms— Hyperspectral imaging , Binary Partition Tree, canonical correlations, segmentation, graph partitioning

1. INTRODUCTION

Hyperspectral imaging has enabled the characterization of regions based on their spectral properties. This had led to the use of such images in a growing number of applications, such as remote sensing, food safety, healthcare or medical research. Hence, a great deal of research is invested in the field of hyperspectral image segmentation. The number of wavelengths per spectrum and pixels 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 a hierarchical 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, independently from its construction, can be used in many different applications such as segmentation [3], classification [1], indexing, filtering, compression or object recognition. This paper focuses on the problem of image segmentation by processing an already constructed BPT. The processing of the BPT consists in the analysis of all the different BPT branches and in the pruning of some of these branches. The analysis proposed here is based on the construction of the affinity matrices using the similarity measure used in the BPT construction. It corresponds to a new distance-based measure depending on canonical correlations relating principal coordinates. Thus, a recursive spectral graph partitioning algorithm enabling a minimal cut is proposed as BPT pruning..
The organization of this paper is given as follows: Section 2 gives a brief introduction on BPT, explaining the details of its construction. The BPT pruning for hyperspectral segmentation 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. Fig. 1 shows an example of BPT construction starting from an original partition involving 4 regions. In the following, this initial partition will be the partition of individual pixels.

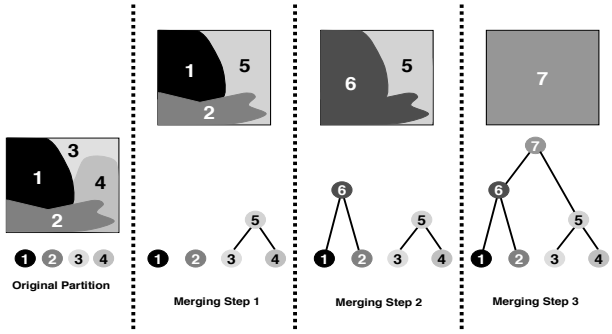


Figure 1: Example of BPT construction

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.

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 [4]. Considering an hyperspectral image containing $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ bands, regions are modeled as N arbitrary discrete distributions $M_R = \{H_R^{\lambda_1}, H_R^{\lambda_2}, ..., H_R^{\lambda_N}\}$, directly estimated from the pixel values. 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 then 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.

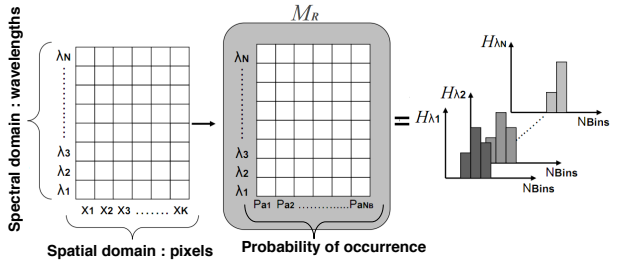


Figure 2: Non parametric statistical model interpretation

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 [5]. 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.

Multidimensional scaling (MDS) [6] 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 dissimiliarties) of M_R can be represented by a N x N distance matrix Δ_R= (δ_{kl}), where $\delta_{kl} = \delta_{lk} \geq 0$ is computed by $\delta_{kl} = e^{(K(H_R^{\lambda_k}, H_R^{\lambda_l}))} - 1$, where $K(H_R^{\lambda_k}, H_R^{\lambda_l})$ is the diffusion distance [7] 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 with Δ_R can be computed by $B_R = HAH$ for each M_R[6]. 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 [8] 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

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

$\lambda_{tR_i}^2$ $\lambda_{tR_j}^2$ 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 $\frac{\sum_{t=1}^{N_s} \lambda_{tR}^2}{\sum_{t=1}^{N} \lambda_{tR}^2} \approx 1$ 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:

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

where det means the determinant and r_i corresponds to the canonical correlation of each axis. Using Eq. 2, an association measure can be defined as:

$$A_W(R_i, R_j) = 1 - W(R_i, R_j) = 1 - \prod_{i=1}^{s} (1 - r_i^2) \qquad (3)$$

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:

$$O_{MDS}(R_i, R_j) = \underset{Ri, Rj}{\operatorname{argmin}} \; 1 - A_W(R_i, R_j) \qquad (4)$$

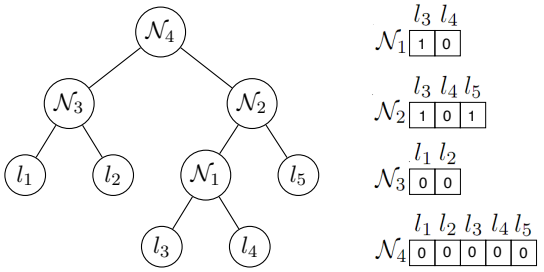
3. PROCESSING OF THE BPT

.

The processing of the tree can be seen as a pruning strategy aiming to remove subtrees composed of nodes which are considered to be homogeneous with respect to some criterion of interest (homogeneity criterion, e.g., intensity or texture). This task can be performed by analyzing a pruning criterion along the tree branches to be able to find the nodes of largest area fulfilling the criterion. This tree analysis may follow a bottom-up or a top-down strategy. Here, the definition of a pruning strategy is presented by a bottom-up analysis of the BPT. Having a BPT such as the example of Fig. 3, the analysis starts by studying all the BPT branches containing different leaves. In the case of Fig. 3, this consists in studying 5 different branches. The purpose is to assign to each node $\mathcal{N}$, a set of $L_\mathcal{N}$ descriptors $\mathcal{N}(l_i)$ indicating whether the leaves l_i of the subtree hanging from $\mathcal{N}$ have to be removed (0) or preserved (1).

Hence, a weight $\mathcal{W}_\mathcal{N} = \sum_{i=0}^{L_\mathcal{N}} \mathcal{N}(l_i)$ can be then defined as the number of leaves contained in $\mathcal{N}$ which consider that the node $\mathcal{N}$ should be removed. This definition leads to define a simple easy greedy algorithm by selecting the highest BPT nodes whose condition $\mathcal{W}_\mathcal{N} > \frac{L_\mathcal{N}}{2}$ is true. According to this criterion, the pruned tree corresponding to Fig. 3 does not contain the branches which hang from $\mathcal{N}_2$. The key of the BPT pruning according to $\mathcal{W}_\mathcal{N}$ is based on how the

$\mathcal{N}(l_i)$ descriptors area assigned. Given a branch $\mathcal{B}$, this should be done by analysing the different partitions obtained by cutting $\mathcal{B}$ at the different BPT levels. In our study, this analysis is done by interpreting each $\mathcal{B}$ as a connected graph. Thus, the optimal cut in the branch can be solved as a spectral graph partitioning problem.

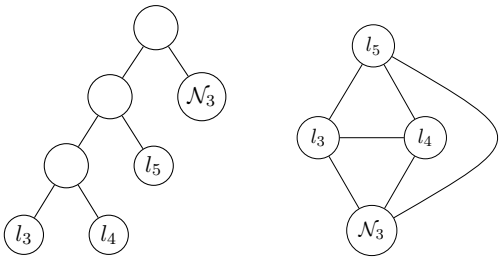


(a) (b)

Figure 3: (a) BPT example. (b) Node decisions

3.1. Tree Branch $\mathcal{B}$ as a connected graph

Let $\mathcal{P}_\mathcal{B}$ be the set of $N_\mathcal{B}$ BPT nodes in the branch $\mathcal{B}$ forming a partition of the image. Given a leaf l_0, a local pruning of $\mathcal{B}$ regarding l_0 consists in deciding which nodes belonging to $\mathcal{P}_\mathcal{B}$ should be removed with l_0. To answer this question, we propose to represent each $\mathcal{P}_\mathcal{B}$ space as a weighted undirected graph G, where each edge is formed between every pair of BPT nodes in $\mathcal{B}$. Fig. 4(a) illustrates the branch example of the leaf l_3 in Fig. 3. For this example, the graph interpretation corresponds to Fig. 4(b). Having a $N_\mathcal{B}$ possible regions to be merged with a leaf, the idea is to study the similarities/dissimilarities between these regions to assure a cut bi-partitioning the set $\mathcal{P}_\mathcal{B}$ into two disjoint non-empty sets $(\mathcal{A}, \mathcal{B})$. In the resulting space, $l_0 \subset \mathcal{A}$ and such that similarity among nodes in $\mathcal{A}$ is high and similarity across $\mathcal{A}$ and $\mathcal{B}$ is low.



(a) BPT branch $\mathcal{B}$ (b) G according $\mathcal{B}$

Figure 4: Example of $\mathcal{P}_\mathcal{B}$ space

The graph G is weighted by w_{ij}, which measures the similarity of an edge linking a pair of nodes $\mathcal{N}_i$ and $\mathcal{N}_j$. These values form a matrix W and are given by Eq. 5, where d is the distance between regions presented in Section 2 and σ controls the size of the neighborhood.

$$w_{ij} = \begin{cases} e^{-\frac{d(\mathcal{N}_i,\mathcal{N}_j)}{\sigma}} & \text{if } i \neq j, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Let D be the diagonal matrix whose values in the diagonal are the total number of connections from each node i to all its neighbor nodes $d_i = \sum w_{ij}$. Given a graph G defined by its D and W, its normalized Laplacian matrix is then computed as $L = D^{\frac{-1}{2}}(D - W)D^{\frac{-1}{2}}$. Using L, a solution for the graph partitioning problem solving the normalized min-cut of G was proposed in Shi and Malik [9]. The method consists in partitioning the graph into two pieces using the eigenvector $\mathcal{E}$ of L associated to the second smallest eigenvalue. This can be done by using the signs of the values of $\mathcal{E}$ which can determine exactly how to bi-partition the graph. In the exemple of Fig. 4, let be $\mathcal{E} = \{l_3, l_4, l_5, \mathcal{N}_3\} = \{+1, +1, +1, -1\}$, this result describes how to cut the graph into two pieces: $\{l_3, l_4, l_5\}$ and $\{\mathcal{N}_3\}$. Having this cut example, nodes will be populated by $\mathcal{N}_1(l_3)$ and $\mathcal{N}_2(l_3)$ to 1, whereas $\mathcal{N}_4(l_3)$ to 0. However, it can not be assumed that the best cut will be performed in a first cut. Thus, we propose a recursive graph partitioning algorithm for each leaf l_i to detect the best cut in the branch. Given a leaf l_i, the method consists in preserving at each iteration the closer BPT neighboring nodes to l_i sharing the same sign of $\mathcal{E}$. For instance, having an $\mathcal{E}$ such as $\mathcal{E} = \{\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_6, \mathcal{N}_7, \mathcal{N}_8\} = \{+1, +1, +1, -1, +1\}$, the method only preserves in this iteration: $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_6$. Computing at each iteration $Ncut$ [9] associated to this cut relating the first sign change. The recursive algorithm stops when this value is higher than a set threshold. The algorithm for each l_i is given by

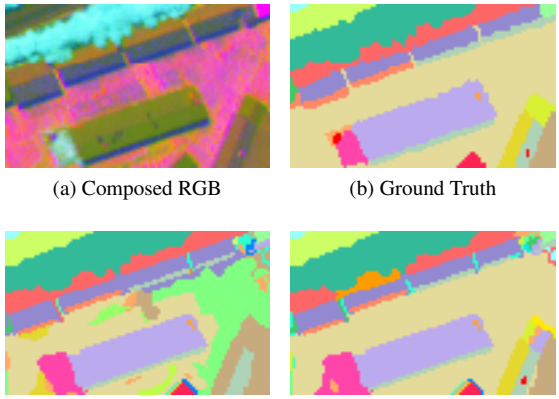
1: $N_\mathcal{B}$ = hanging nodes from l_i until the root
2: **while** $N_\mathcal{B} > 2$ and END is not true **do**
3: Compute the Laplacian matrix relating l_i with all their possible $N_\mathcal{B}$
4: Compute the graph cut level k according to the closer BPT nodes on the branch having the same sign than $\mathcal{E}(l_i)$.
5: Compute $Ncut$ between the first nodes arriving to level k and the remaining $N_\mathcal{B}$
6: **if** $Ncut <$ maximum allowed $Ncut$ **then**
7: L_{Cut} is equal to k
8: **else**
9: END is true
10: **end if**
11: Next Laplacian matrix to study is given by the hanging nodes until level k, then $N_\mathcal{B} = k$
12: **end while**

At the end of the algorithm, the optimal cut on the branch will be retained in L_{Cut}. Thus, this information is used to populate the BPT nodes on the studied branch. Starting from a leaf l_i, all father nodes formed in its branch until the level L_{Cut}, will have a positive weight $\mathcal{N}(l_i) = 1$ (See Fig. 3). Contrarily, fathers nodes formed in a level higher than L_{Cut} will have a zero weight in the decision. Note that once a BPT is pruned, a segmentation result is obtained by selecting the leaf nodes of the pruned BPT.

4. EXPERIMENTAL RESULTS

We first provide an evaluation of the BPT pruning proposed in Section 3. The experiments have been performed using a portion of Pavia Center image from hyperspectral ROSIS sensor, acquired by DLR and provided by Prof. P. Gamba. The data contain 102 spectral bands. Fig. 5(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 on the image range (here $N_{bins} = 100$). The first component dimension found by the sequence c_k is $s = 2$. Once the BPT is constructed, we proceed by pruning the tree using the algorithm presented in Section 3. σ value

of Eq. 5 is set to 0.01 knowing that the range of the distance defined in Eq. 4 varies between 0 and 1. The maximum $Ncut$ value allowed in all the studied branches is 0.3. To evaluate the quality of the BPT pruning, we compare the obtained results against a trivial pruning criterion based on the number of regions in the BPT following the merging sequence [3]. This classical strategy consists in extracting a segmentation result involving a given number N_R of regions. To evaluate the resulting partitions, the symmetric distance d_{sym} [10] is used as a partition quality measure. Having a partition P and a ground truth GT (Fig. 5(b)), 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.
Fig. 5(c)(d) show the segmentation results obtained with the trivial and the PBT pruning proposed in this work, respectively. In both cases, the resulting partitions involve 54 regions. Comparing both results, the quantitive d_{sym} and the visual evaluation corroborate that the partition obtained by the pruning proposed in Section 3 is much closer to the ground truth than the one computed with a simple stopping of the region merging algorithm. On the lower right corner,, Fig. 5(c) illustrates how various small regions can be retrieved by using the proposed BPT pruning.



(a) Composed RGB (b) Ground Truth

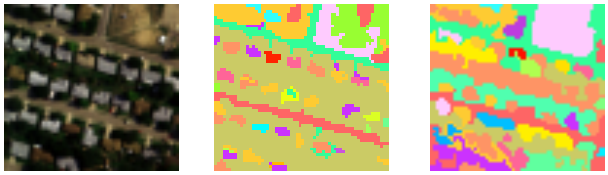
(c) Pruning by number regions (d) Min $Ncut$ Pruning BPT

Figure 5: (a) Pavia Center ROSIS RGB Composition, (b) Manually created Ground Truth, (c) Partition extracted from the trivial pruning leading to d_{sym} =40, (d) Partition computed with the proposed pruning leading to d_{sym} =20

A second experiment is carried out using a portion of a publicly available HYDICE hyperspectral image. After removing water absorption and noisy bands, the data contain 167 spectral bands. A RGB combination of three of them is presented in Fig. 6(a). In this case, the number of bins representing the histograms is $N_{bins} = 256$ and the first component dimension is $s = 3$. For this second data set, the quality of the BPT pruning results are compared with the results obtained by the classical RHSEG [11]. In the case of RHSEG, the similarity criterion used is SAM with spectral clustering weight 0.1. Both partitions contain the same number of regions equal to 57. The results shown in Fig. 6 illustrate the interest of the BPT pruning strategy compared to the state of the art.

5. CONCLUSIONS

In the context of hyperspectral image segmentation, the processing of Binary Partition Tree has been discussed in this paper. A recursive graph partitioning algorithm has been proposed for studying BPT branches to perform optimal cuts. The use of BPT with graph partitioning enables to compute affinity matrices between coherent regions which is more robust than affinity matrices computed using the pixel-based representation. Regarding the BPT construction and pruning, a new distance-based measure depending on canonical correlations relating principal coordinates has been presented in this work. Future works will be conducted to optimize the graph interpretation since the knowledge about the merging order could provide an insightful additional information. The study of affinity matrices introducing feature descriptors will be also studied in the future.



(a) RGB (b) Pruned BPT (c) RHSEG

Figure 6: (a) Urban Hydice RGB Composition, (b) Partition extracted from pruned BPT, (c) Partition computed with RHSEG [11]

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] 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.

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

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

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

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

[8] 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.

[9] J.Shi and J.Malik, "Normalized cuts and image segmentation," Proc. of CVPR, 1997.

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

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