UNSUPERVISED NONPARAMETRIC CLASSIFICATION OF POLARIMETRIC SAR DATA USING THE K-NEAREST NEIGHBOR GRAPH

Ashlin Richardson[1,2], David G. Goodenough[2,3], Hao Chen[2], Belaid Moa[3], Geordie Hobart[2,3], and Wendy Myrvold[3]

[1]Department of Mathematics and Statistics, University of Victoria
[2]Pacific Forestry Centre, Natural Resources Canada
[3]Department of Computer Science, University of Victoria
SSM A425, 3800 Finnerty Road, Victoria, Canada, V8P 5C2
ashy@uvic.ca 250-853-3293

ABSTRACT

Polarimetric SAR classifications are often based on assumptions about the shape of clusters in the data space. Such a scheme will fail for nonlinear structures in the feature space, unless the classification algorithm has the capacity to describe cluster shapes in sufficient generality. Existing polarimetric SAR classification methods are faced by this exact problem: typically they initialize clusters in the Cloude-Pottier parameter space [1], further optimizing them in the coherency matrix space [2, 3]. Methods using K-means [2] or agglomeration [3] require clusters that are spherical, or compact and well separated, respectively. In the Cloude-Pottier space, these requirements are not met, so initialization in the Cloude-Pottier space cannot be consistent with optimization by K-means or agglomeration. This paper sets out to address this problem, by implementing a new data-driven clustering approach, for arbitrarily shaped clusters. It is applied to quad-polarisation data, demonstrating the new methodology's potential for forest land-cover type discrimination.

Index Terms— Classification, Density Estimation, K-Nearest Neighbor Graph, Polarimetric SAR, Cloude-Pottier Decomposition.

1. INTRODUCTION

Hartigan [4] clarified Wishart's idea of clustering: *the elucidation of data modes, regardless of their shape or variance*. In this pursuit, Hartigan employed the notion of high-density clusters (maximally connected subsets of level sets of density in the feature space). Given a density function p and threshold λ the level set for p and λ is:

$$L(\lambda; p) = \{x | p(x) > \lambda\}. \qquad (1)$$

A hierarchical classification may be produced by determining a set of all critical values $\{\lambda_i\}$ for which the number of maximally connected components of $L(\lambda,p)$ changes. The hierarchical classification is given by a tree, whose nodes represent maximally connected components of level sets of the density; starting at the bottom, $L(0,p)$ is the root node (representing the all-encompassing level set). Suppose we reach a threshold value λ_i (above which a maximally connected component splits into n_i components) the graph branches n_i times. Thus, the result is hierarchical. Stuetzle [5] gives a general data-driven approach using density estimation and graph methods. While such an approach will be required ultimately for image analysis, its great generality comes at the expense of an additional variable; namely: a threshold for the hierarchical splitting. Accordingly, we pursue a simpler approach, facilitating comparison with well-known polarimetric SAR classification algorithms.

2. FINDING NONLINEAR CLUSTERS

The new methodology has two stages. 1) We estimate the density value at each point in the feature space. 2) We climb the density estimate, elucidating arbitrarily shaped "modes" in the density, à la Wishart. The "hill climbing" is performed on the K-Nearest Neighbors (KNN) graph, a directed graph whose nodes correspond to the points of the feature space. From each node of the graph, "K" edges emanate, connecting the point with each of its "K" nearest neighbors. The density estimate at each point is computed using the "K" nearest neighbors – consequently, 1) and 2) are defined to be consistent with one another, ensuring that the scheme is both stable and deterministic. Notably, K-means [2] and agglomerative algorithms [3] are unstable with respect to initial conditions. Furthermore, agglomeration is unstable with respect to the order of merging. In summary, the parameter "K" implicitly defines the number of classes so that the clusters are well formed - both in shape and in quantity - by the estimated density in the feature space.

2.1. Density Estimation

A practical method (2) is estimating the density at a point x as the reciprocal of the average of the distances to the K-nearest neighbors of x:

$$\rho(x) = \frac{1}{\frac{1}{K}\sum_{n \in N} d(x,n)}. \qquad (2)$$

Here, N denotes the K-nearest neighbors of x, and n is an element of N. Note that increasing the value of "K" takes more information about a point into consideration. Setting K=1, the estimate is completely local, which will result in producing a high number of clusters. Larger values of K will yield increasingly global estimates, decreasing the granularity of the result, so that fewer clusters are produced.

2.2. Hill Climbing on the K-Nearest Neighbor Graph

After computing the K-Nearest Neighbors of each point, the K-Nearest Neighbor (KNN) graph is constructed. Following the assignment of an estimated density to each node, we "hill climb" the graph, reaching the highest density points, which are then deemed representative elements of clusters. This is accomplished by a function visiting the graph's nodes recursively: an iteration of the function is performed at each node. Upon visiting a node, if the node's density is greater than that of every K-neighbor, the iteration terminates, whereby the node is deemed to represent an "attractive center" of the data (and is given a new label). Else if the current node is already labeled - the iteration terminates (in either case, upon termination, all other nodes visited during that iteration are assigned the label of the final node). Otherwise the current node must have a neighbor of higher density – accordingly, we visit the highest density neighbor. At each subsequent visit, we consider again the above three possibilities. An "attractive center" along with all points "attracted" to it, through the activity of the recursive function, forms a cluster. A sample KNN graph with four clusters is shown in Figure 1.

2.3. Implementation

A fast visualization suite including the new algorithm was implemented in C++/OPENGL. The suite is interactive so that the user can vary the parameter K on the fly, visualizing the resulting clusters. The user also may interactively choose which subspace of the parameter space to visualize in three dimensions.

In order for the clustering and visualization, including intensive nearest neighbor calculations, to be performed in real time, only a small fraction of the pixels (~10,000) were used to produce the KNN based clustering. After labeling the 10,000 points, a randomized algorithm quickly assigned each of the remaining pixels (~200,000) to the nearest

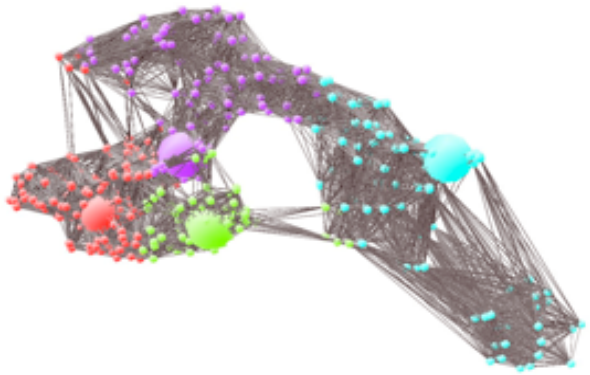

Figure 1: A neighbor graph of randomly generated points in a three dimensional feature space with K=40. Spheres represent the points, colored to indicate cluster membership. Large spheres represent the attractive centers. There are four clusters.

cluster with high probability. Specifically, each point is compared with elements randomly selected from each of the clusters. The new label is given according to the nearest point encountered. Furthermore, a larger proportion of random checks were performed when labeling spatially non-homogeneous pixels, reducing the number of point mis-registrations introduced by the randomized algorithm.

3. DATA AND ANALYSIS

Our study area is located in BC and Alberta, Canada, near the river Chinchaga. The area features mature forest stands, clear-cuts, and a prominent fire scar dating from 2002. ALOS PALSAR quad-polarisation data were collected over the site during the summer of 2009.

3.1. Data Processing

The data were compensated for Faraday rotation according to the Bickel and Bates method [8] and converted to the coherency matrix. A multi-look factor of 5 in the row direction was applied, to approximately square the pixels. A 3x3 box filter was then applied for speckle reduction. Finally, a 5x5 multi-look was applied, reducing the data volume in order to show the full scene. The resulting image size was 737x249. This processing sequence reduces the data volume without excessive smoothing. Three parameters from the Cloude-Pottier decomposition were selected as input for the new algorithm: Cloude Entropy (H), Shannon Entropy (S.E.), and Cloude Alpha (α) – see Figure 2a).

3.2. Unsupervised terrain classification preserving polarimetric scattering characteristics

To illustrate the advantages of the new classification over a topographically challenging region, two algorithms were run for comparison. The first is Lee's 2004 classification [2] featuring a permanent partition into three categories (single bounce, double bounce, volume scattering) based on the

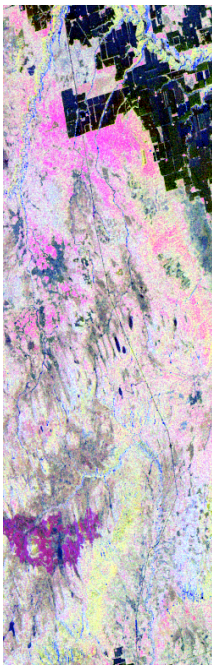

Figure 2a) Red=Entropy, Blue=Alpha, Green=Shannon Entropy.

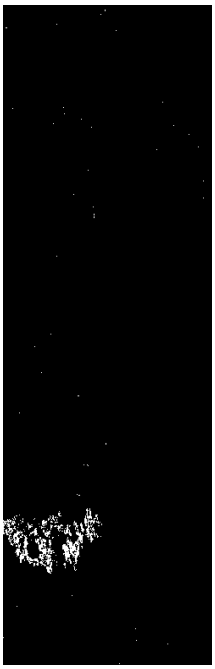

Figure 2b) Lee algorithm – 200 initial segments – 35 final clusters.

Figure 2c) Lee algorithm – 200 initial segments – 40 final clusters.

Figure 3a) Wishart H/A/alpha classifier - 5x5 analysis window (fire scar class).

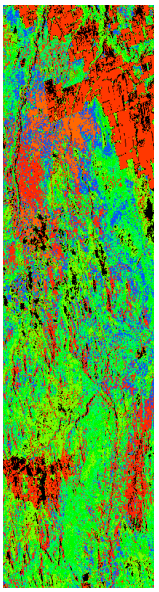

Figure 3b) Wishart H/A/alpha classifier – 1x1 analysis window (16 classes).

Figure 3c) KNN classifier using 10000 points, K=55, 22 clusters.

Figure 3d) KNN classifier using 10000 points, K=35, 51 clusters.

dominant power given by the Freeman decomposition. Next, within the three categories, a large number of small segments are produced, by partitioning the pixels in terms of the power given by the Freeman decomposition. Agglomerative merging is conducted based on the intra-cluster Wishart distance, in order to reach a desired number of clusters. Finally, the clusters are optimized using K-means iteration by means of the Wishart distance.

It is interesting to note that Jong-Sen Lee's algorithm successfully produced the fire scar boundary, using a much greater quantity of initial segments (200 per category) than originally recommended by the authors (30 per category). Since no double bounce pixels were found (leaving only single and volume categories) the use of excessive initial segments somewhat manhandles the algorithm away from its intended function of "preserving scattering characteristics" towards a purely agglomerative scheme (which is arguably better prepared to handle clusters of unusual structure). The binary classification result is shown in Figure 2b) - the fire scar is at the bottom. No additional filtering was used; 35 final clusters were selected. The classification was shown to be unstable: by changing the number of classes to 40, the fire scar boundary was lost in confusion with both urban and natural areas.

3.3. H/A/Alpha-Wishart classifier

The POLSARPRO implementation of the unsupervised classifier: H/A/Alpha-Wishart was also executed for comparison, with 16 classes. Likely owing to the simplistic initialization procedure, detecting the fire scar was not as successful. The fire scar boundary was reasonably produced - see Figure 3a) - when this algorithm was executed with 5x5 additional averaging (although misregistrations occurred in the urban area). The central limit theorem indicates that substantial averaging will cause the clusters to become roughly spherical ("Gaussian") making the clusters compatible with the K-means optimization. Thus when the additional 5x5 averaging was taken away, the fire scar boundary was lost entirely; see Figure 3b) for the classification map, in pseudo-color.

3.4. New Method: KNN clustering

Figure 3c) indicates the fire scar class resulting from the new methodology, which used ~ 10,000 points in the three-dimensional feature space. Figure 3c) was produced using K=55 and the density estimate shown in Equation (2), 22 classes resulted. Figure 3d) was also produced using the same density estimate, however, only K=35 nearest neighbors were used, producing 51 classes. Although a few point mis-registrations were produced, the fire scar class was preserved, demonstrating the algorithm's superior stability with respect to the parameter "K".

4. CONCLUSIONS AND FUTURE WORK

The preliminary results are encouraging, showing the potential and effectiveness of the new methodology in segmenting and classifying polarimetric SAR data for forest applications, especially for fire scar detection. Compared to the two other classifications (based on stricter statistical assumptions and more restrictive initializations) the new classifier showed a superior ability to resolve clusters with non-linear shape. Furthermore, the proposed algorithm provides a vehicle for testing the consistency of feature space initialization followed by matrix space optimization. Experiments addressing this issue using the appropriate vector and matrix distance functions are in progress. In [6] the results will be compared with those of a new coherence based geometrical detector [7]. The validity of the results is being further evaluated using GIS ground reference data and other remote sensing data from a variety of sources.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of Natural Resources Canada (NRCAN), NSERC (D. Goodenough, R. Illner), and the Department of Computer Science and the Department of Mathematics and Statistics of the University of Victoria.

REFERENCES

[1] S. R. Cloude, E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sensing,* vol.35, no.1, pp.68-78, 1997.

[2] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sensing,* vol.42, no.4, pp. 722-731, 2004.

[3] C. Fang, W. Hong, Y. Wu, and E. Pottier, "An Unsupervised Segmentation With an Adaptive Number of Clusters Using the SPAN/H/α/A Space and the Complex Wishart Clustering for Fully Polarimetric SAR Data Analysis," *IEEE Trans. Geosci. Remote Sensing,* vol.45, no.11, pp. 3454 - 3467, 1997.

[4] Hartigan, J. A., *"Clustering Algorithms,"* Wiley, New York, 1975.

[5] W. Stuetzle, R. Nugent, "A Generalized Single Linkage Method For Estimating The Cluster Tree Of A Density", *Journal of Computational and Graphical Statistics,* vol.19, no. 2, pp.397-418, 2009.

[6] S. R. Cloude, A. Marino, D. G. Goodenough, H. Chen, A. Richardson, B. Moa, "Radar Polarimetry for forestry applications: ALOS and RADARSAT-2 studies in Canada", *Proceedings of ForestSat 2010*, Lugo, Spain, Sep. 2010.

[7] A. Marino, S. R. Cloude, "Detecting Depolarizing Targets using a New Geometrical Perturbation Filter", *Proceedings of 8th European SAR Conference,* EUSAR, Aachen, Germany, June 2010.

[8] A. Freeman, "Calibration of Linearly Polarized Polarimetric SAR Data Subject to Faraday Rotation," *IEEE Trans. Geosci. Remote Sensing,* vol. 42, No. 8, pp. 1617–1624, 2004.