HIERARCHICAL UNSUPERVISED NONPARAMETRIC CLASSIFICATION OF POLARIMETRIC SAR TIME SERIES DATA

Ashlin Richardson [(1)], *David G. Goodenough F.IEEE* [(1)], *and Hao Chen* [(2)]

(1) Department of Computer Science, University of Victoria, Victoria, BC
(2) Pacific Forestry Centre, Canadian Forest Service, Victoria, BC

ABSTRACT

Clustering (and classification) among other approaches of land-cover type discrimination for Polarimetric SAR (Pol-SAR) data often explicitly or implicitly assume a lot about the shape of the clusters (or the classes, in the case of classification). For example, this is an issue for Pol-SAR classification methods [1,2,3] that initialize clusters in decomposition parameter feature spaces [4], subsequently refining the clusters by Wishart moving-means iterations in coherency matrix (T3) space. Indeed, using the means as cluster (or class) representatives can be successful, provided that clusters in the data are compact, well separated, and convex. However, highly nonlinear features and unusually shaped clusters are often obtained when dealing with Pol-SAR data. To address this issue we present a data-driven hierarchical clustering technique. This we demonstrate for forest-type discrimination purposes with a multi-temporal Radarsat-2 sandwich.

Index Terms— Classification, Clustering, Density Estimation, Nonparametric, Hierarchical, Radarsat-2, Time Series

1. INTRODUCTION

1.1. Problem

Data clustering (and classification) approaches for Polarimetric SAR (Pol-SAR) typically assume something about the shape of clusters (or classes, in the classification context). Such is the case for established Pol-SAR classification methods based on clustering [1,2,3] that initialize clusters in decomposition parameter spaces [4], subsequently refining those clusters by k-means Wishart optimization in the coherency matrix (T3) space. Indeed representing a cluster by the mean is successful provided that clusters in the data are compact, well separated, and approximately round. Such assumptions are computationally effective but may yield poor results when highly nonlinear features and unusually shaped clusters are encountered. This is the challenge we address here.

1.2. Method

Pol-SAR classification tools should ideally be sufficiently general in order to represent arbitrarily shaped clusters. Accordingly, an unsupervised nonparametric density based clustering scheme for Pol-SAR data was presented [5] based on "mode seeking", and applied successfully to space-borne Pol-SAR data for automatic isolation of previously burned forest areas. For increasingly flexible, accurate and automatic Pol-SAR data clustering, in this paper we demonstrate improvements to [5] including: improved performance implementation, improved density estimation, and full hierarchical operation. The "mode seeking" approach considered here, unlike typical "mean shift" [6] implementations, is based on K-nearest neighbor density estimation with hill climbing on the K-nearest neighbor graph. Thus, local adaptiveness is conferred in the sense that no distribution is explicitly assumed (effectively, there is a locally varying distribution). Furthermore, the method is non-iterative, and requires no initialization. The modes are sought via a data-driven step size: the displacements between a given point and its K-nearest neighbors are the only such steps permitted. The only parameter is K, the number of neighboring points to consider about each point.

Figure 1: The time-series ensemble with encoding H,S,V=alpha, Shannon entropy, span

2. STUDY SITES AND RADARSAT-2 POLARIMETRIC SAR DATA

2.1. Data Acquisitions

Time series Radarsat-2 image data (FQ-17 mode, ascending, look angle ~37°) were collected over a study area in northern Alberta (center geo-coordinates 57°35'N, 117°45'W) near the Keg River, as in Table 1.

Date	Interval	**Date**	Interval
13-02-18	-	**13-10-16**	24
13-03-14	24	**13-11-09**	24
13-04-07	24	**13-12-03**	24
13-05-01	24	**13-12-27**	24
13-05-25	24	**14-01-20**	24
13-06-18	24	**14-02-13**	24
13-08-05	48	**14-04-02**	48
13-09-22	48	**14-04-26**	24

Table 1: Radarsat-2 acquisitions

Forests (predominantly coniferous) are prominent within the Keg River study area, which features an extensive recorded history of fires (multiple fires in every decade since 1950). The most recent is the Keg River wild fire that burned over 4830 hectares in 2002. This is the primary feature of interest for this study.

2.2. Data Processing

To accomplish speckle reduction and simultaneous data volume reduction, multi-looking was applied to each date (4 in azimuth and 2 in range) resulting in 4x4 covariance (C4) matrices. Subsequently, 5x5 box filtering was applied. Next, correlation techniques were utilized to co-register all other dates to a master date (March 14, 2013). The master was selected according to the minimization (over all possible choices for the master) of the grand sum of the magnitudes of those displacement vectors terminating at each slave image center and originating at the center of the master.

Decomposition parameters were produced for all of the co-registered dates, using C++ code adapted from PolSARPro (please see Figure 2 for an example). Moreover, the same parameters were calculated also for the time series ensemble (the time-averaged covariance matrix). To these parameters, a further 2x2 multi-look was applied, resulting in decomposition products of dimension 670x875 pixels with spatial resolution of 42x30m. Ultimately, each time-varying decomposition product was temporally filtered using C++. This was accomplished by calculating the following six quantities: the minimum, maximum, mean, standard deviation, median, and median absolute deviation (MAD) for each decomposition parameter, with respect to time. Selected parameters were then used as input to the unsupervised classification scheme.

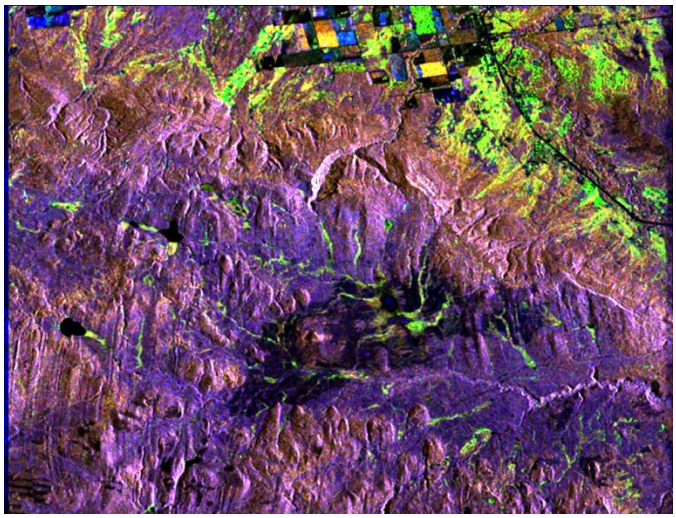

Figure 2: The time-series ensemble with encoding R,G,B=(max(T33,T22,T11))

3. NONPARAMETRIC APPROACH TO FIND ARBITRARILY SHAPED CLUSTERS

3.1. Distance Matrix Evaluation

Vectors in n-dimensional Euclidean space are taken as input to the unsupervised scheme. Each vector represents a pixel, and each dimension corresponds to one of the selected parameters. For convenience we first linearly scale each of the dimensions so that they are normalized to the interval [0,1]. This step is justified since the approach represents estimation of the cluster tree of a density: the cluster tree is topologically invariant under affine transformations of the feature space [7]. Next, we calculate the matrix of (pairwise) distances. We alleviate the computational expense of this operation by using multi-threaded acceleration,

1) a max-heap to retain references to only the KMAX-nearest points to a given point (where KMAX is some relatively small number) and,
2) storing, for each point, references to the KMAX-nearest points, for further use.

The parameter KMAX is taken sufficiently large to permit the user to vary the parameter K, as below.

3.2. Density Estimation

Next, we estimate the density for each point. This step involves the parameter K. Although more sophisticated methods exist [8] a simple and effective way is to take the density estimate at a point x, to be proportional with the reciprocal of the average of the distances to the K-nearest neighbors of x:

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

In the above formula, N denotes the set of K-nearest neighbors of x, to which n belongs. Thus, larger values of K

entail that the density estimate is more global, leading to coarser-grained results (fewer clusters). Accordingly, smaller values of K result in increasingly local estimates, resulting in a greater number of fine-grained clusters.

3.3. Hill Climbing

For the density estimate at each point, a label is assigned according to the following recursive function:

myLabel(**x**):

- If **x** is of highest density (among its K-nearest neighbors) we define myLabel(**x**) to be a new label.
- Else, **x** has a higher density neighbor **y**, and we define myLabel(**x**) to be mLabel(**y**).

3.4. Cluster Merging

The previous step resulted in each pixel being assigned a label, corresponding to the associated cluster. The final result will be a dendrogram (tree) for hierarchical cluster representation (formally, an estimate of the cluster tree of a density [7]). The hierarchical merges are (unlike model-based approaches) motivated by density observations. Indeed, the splits in the tree are ordered by density. We fix the tree according to the following recipe:

1) Find the interfaces between clusters.
2) Perform merging with respect to the density ordering of the interfaces (highest first). For efficiency, a disjoint set-forest data structure (with path flattening) is used.

3.5. Interactive Visualization: the "*" operation

According to the hierarchical clustering result, each cluster C is contained within a greater cluster C* (with the exception of the root of the tree, which represents the totality of pixels). The operation "*" is implemented (in C++/OpenGL) as a real-time user-interactive aspect of the image display, of which pixel selection (hence, cluster selection) and variation of the parameter K are also features. Accordingly, we can select a cluster (a leaf of the tree) and produce binary clustering maps (in this case, for the fire scar).

4. PRELIMINARY RESULTS

The six temporal quantities were generated for the following selected parameters: Entropy, Alpha, Shannon Entropy, and Span, resulting in twenty-four bands. The 670x875x24 image cube was then input to the unsupervised scheme. For the 2002 Keg River fire, a GIS truth polygon from the Alberta Wild-Fire database (created by a human photo-interpreter) was available (see Figure 3). For the result we exhibit binary classification maps due to a cluster representative of the same fire, for different K-settings: for Figures 4 and 5 we have K=150 and K=30, respectively.

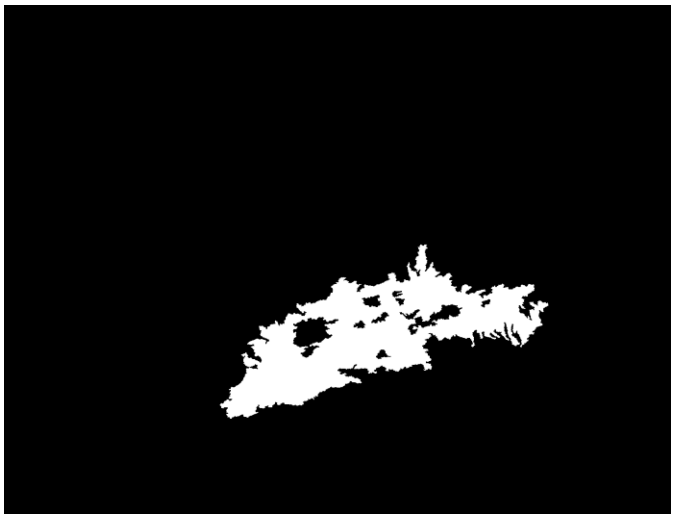

Figure 3: Alberta Wild Fire database polygon (the 2002 fire).

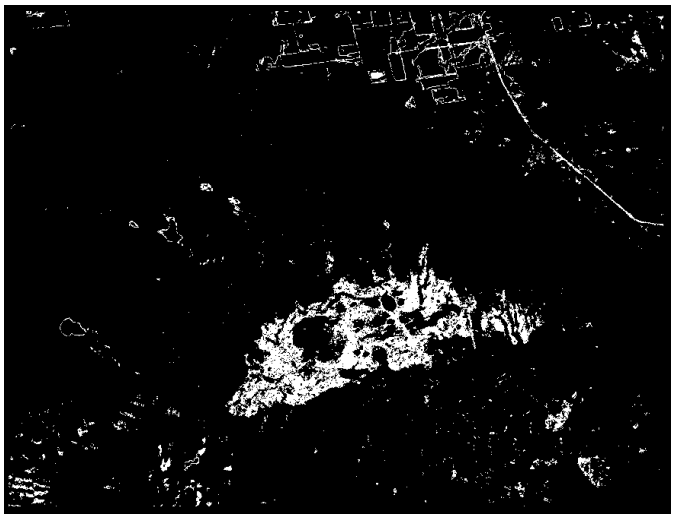

Figure 4: Binary fire scar classification (K=150).

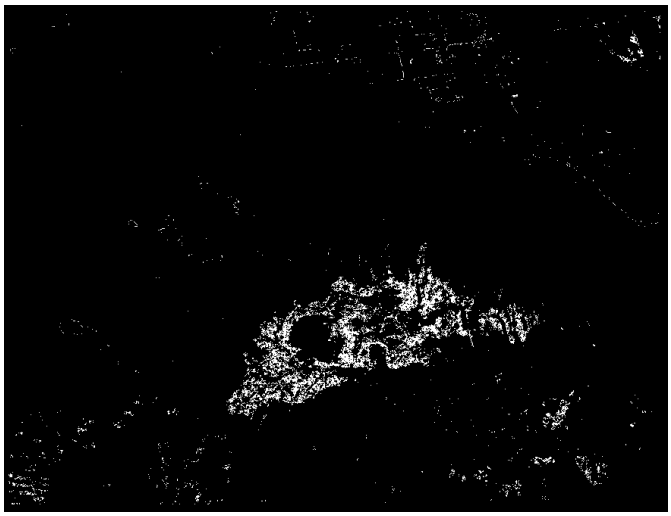

Figure 5: Binary fire classification (K=30) using three applications of "*".

As indicated by the results, larger K-values produce more inclusive (less specific) clusters. Then, when using smaller K-values (for more detailed analyses), applying the "*" operation may be conveniently used to produce a more inclusive category (as was done for the result in Figure 5). Cf. [9], accuracy and precision measures for binary classification are reported in Table 2 (the respective formulae also appear there). For convenience the quantities **TP, TN, FP, FN** are expressed as fractions (rather than as counts) i.e.,

$$\mathbf{N=TP+TN+FP+FN=1}.$$

The Figure 4 (K=150) result is more inclusive, meaning that (relative to the Figure 5 result) both true positives (TP) and false positives (FP) are more frequently observed. Thus for the Figure 5 (K=30) result, we observe less frequently both true positives (TP) and false positives (FP). Both results show 95% binary classification accuracy. For the Figure 5 (K=30) case a substantially greater value is observed for the precision measurement, relative to the Figure 4 case.

	K=150	**K=30**
TP - True Positives (%)	3.20	1.69
TN - True Negatives (%)	92.28	93.49
FP - False Positives (%)	1.70	0.49
FN - False Negatives (%)	2.82	4.33
TP/(TP+FP) - Precision (%)	**65.30**	**77.36**
(TP+TN)/N - Accuracy (%)	**95.48**	**95.18**

Table 2: Accuracy and Precision

5. CONCLUSIONS AND FUTURE WORK

The results indicate the potential and effective utility of the new method for applications in forestry, such as the detection of historically burned areas using Radarsat-2. Moreover, as the cluster boundaries in the feature space are well motivated in terms of statistical density, the method shows good stability with respect to the choice of K. Further work should investigate using the new methodology in the context of other metrics, particularly those that offer a physically meaningful interpretation in terms of polarimetry.

6. REFERENCES

[1] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 5, pp. 2249-2258, 1999.

[2] J.-S. Lee, M. Grunes, E. Pottier, L. Ferro-Famil, "Unsupervised Terrain Classification Preserving Polarimetric Scattering Characteristics," IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no.4, pp. 722-731, 2004.

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

[4] S. Cloude, and E. Pottier, "An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR," IEEE Transactions on Geoscience and Remote Sensing, vol. 35, no.1, pp. 68-78, 1997.

[5] A. Richardson, D.G. Goodenough, H. Chen, B. Moa, G. Hobart, W. Myrvold, "Unsupervised Non-parametric Classification of Polarimetric SAR Data Using The K-nearest Neighbour Graph," in Proc. IGARSS 2010, Honolulu, Hawaii, pp. 1867-1870, 2010.

[6] K Fukanaga, L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," IEEE Transactions on Information Theory, 1975.

[7] W. Stuetzle and 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, 2010.

[8] G. Biau, F. Chazal, D. Cohen-Steiner., L. Devroye., C. Rodriguez, "A weighted k-nearest neighbor density estimate for geometric inference," Electronic Journal of Statistics, vol. 5, pp. 204-237, 2011.

[9] T. Fawcett, "An introduction to ROC analysis", Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006.