UNSUPERVISED CLASSIFICATION OF AGRICULTURAL LAND COVER USING POLARIMETRIC SYNTHETIC APERTURE RADAR VIA A SPARSE TEXTURE DICTIONARY MODEL

Robert Amelard, Alexander Wong, David A. Clausi

Department of Systems Design Engineering, University of Waterloo, Canada

ABSTRACT

A sparse texture dictionary learning method for unsupervised land cover classification is presented. The method takes the stance that land cover in remote sensing data is best analysed in texture patches rather than localized pixels. To this end, a feature vector is designed that describes local texture information in a spatially coherent manner. This texture model is extracted for each pixel in the scene. A sparse dictionary of global texture models is then learned to characterize the underlying texture distribution of the scene in a simplified manner. An unsupervised classifier is learned using these global texture models for grouping pixels exhibiting high similarity. Being an unsupervised classifier, the class labels that are learned are unbiased toward human interpretation of the scene, and rather are learned according to the texture information. The method is validated using polarimetric SAR data over a Flevoland, Netherlands agriculture scene, but may be generalized to any remote sensing data. Promising experimental results show how the proposed method retains the spatial coherence of crops, and attains higher accuracy than recent unsupervised and supervised classification methods using the same data.

***Index Terms*—** image classification, image texture analysis, land cover classification.

1. INTRODUCTION

Determining the different types of agriculture present in a scene is an important problem in geoscience and remote sensing. It can be used, for example, to analyse land cover change, which is one of the most important factors in determining global ecological shifts [1]. Polarimetric synthetic aperture radar (SAR) data are able to provide information about a scene irrespective of cloud cover, as well as different scene information according to its backscatter properties. However, the image it provides is not intuitive to a human since it operates in non-visible bands of the electromagnetic spectrum, and usually contains much noise. These problems make it difficult to manually interpret a scene. It is therefore desirable to employ an automatic algorithm for SAR scene analysis and classification of the pixels to their respective land cover classes.

Due to continuing advances in remote sensing granularity and image analysis methods, pixel-based methods are being shown to be inadequate for whole scene interpretation [2, 3]. Congruent with these findings, recent classification methods have focused on performing region-based analysis [4, 5]. Unsupervised classification involves learning underlying patterns inherent in the data irrespective of the ground-truth classes. Thus, a label map that is generated using unsupervised classification is unbiased toward any human interpretation as it only operates on the underlying data.

2. METHODOLOGY

The goal of the proposed work was to investigate the effectiveness of employing sparse texture modeling in an unsupervised learning scheme for land cover classification. We learned a sparse dictionary of global texture models for determining a mapping from multi-dimensional remote sensing data (e.g., SAR) to agricultural land cover. Sparse dictionary learning has been successfully used in image processing for learning representative image textures and denoising [6]. It is therefore a promising methodology for determining information from noisy SAR data. In this paper, we use Pauli decomposition [9] of polarimetric SAR, but this method can be generalized to any remote sensing imagery with ground-truth information, as long as the scene being classified was obtained using the same bands and polarizations with which the texture dictionary was learned.

An overview of the proposed approach can be described as follows. First, a set of multi-dimensional spatially-sensitive texture representations were extracted for each pixel in the scene. Then, an N-element sparse dictionary texture model was learned using these global texture representations. Using these representative dictionary elements, an unsupervised classifier was learned and used to assign a class label to each pixel. This approach does not use any manually-defined

Thanks to National Science and Engineering Research Council and the Ministry of Research and Innovation for funding support, GRSS-DFC for providing the Flevoland dataset, and Dr. Anfinsen for the Flevoland ground-truth classes.

labels, as it generates class labels based on image texture characteristics.

3. MULTI-DIMENSIONAL TEXTURE REPRESENTATION

A dictionary of textures is a set of feature representations describing the pixels in a scene. In order to build such a texture dictionary, a feature vector must be designed such that it describes the texture about a pixel in the polarimetric SAR image. We defined a spatially-sensitive multi-dimensional texture representation such that two patches are deemed "similar" if they are both similar in raw texture and spatial position. The spatial constraint is an important and valid constraint in this framework, as crops tend to grow in discrete areas due to soil characteristics and growing techniques. This representation allows us to describe the multi-channel, noisy SAR data in terms of simpler representative texture elements.

Specifically, let f denote an $M \times N$ SAR image with n_c polarization channels with which we wish to perform classification. Then, let $f(\mathbf{x}_i, c)$ represent the SAR data at pixel location $\mathbf{x}_i$ and channel c in image f, where c is one of the n_c polarization channels. Given an r-layer pixel neighborhood centered at pixel location $\mathbf{x}_i$, its spatially-sensitive multi-dimensional texture representation $g(\mathbf{x}_i)$ can be defined as:

$$g(\mathbf{x}_i) = \Big[f(\mathbf{x}_i, c_1) \;\; f(\mathbf{x}_i^{1,1}, c_1) \;\; f(\mathbf{x}_i^{2,1}, c_1) \;\; \dots \;\; f(\mathbf{x}_i^{8r-1,r}, c_{n_c}) \;\; f(\mathbf{x}_i^{8r,r}, c_{n_c}) \;\; \mathbf{x}_i \Big] \quad (1)$$

where $f(\mathbf{x}_i^{j,k}, c)$ denotes the pixel value at the j^{th} position of the k^{th} radial layer around pixel $\mathbf{x}_i$ in channel c (see Fig. 1). The neighborhood can be traversed in any manner as long as the traversal method is consistent across each patch. Note that there are $8k$ pixels at radial layer k. Global spatial context was encoded into this feature vector by appending the pixel's spatial location $\mathbf{x}_i$. That way, patches are deemed "similar" based on both their neighborhood pixel values as well as their relative proximity. This constraint is only held for the sparse coding step, and is not directly embedded into the classification phase. As such, given the SAR image f, a total of $M \times N$ texture representations are extracted, one for each pixel location.

4. DICTIONARY-BASED GLOBAL TEXTURE MODEL

Upon computing the set of spatially-sensitive texture representations $G = \{g(\mathbf{x}_1), g(\mathbf{x}_2), \dots, g(\mathbf{x}_{MN})\}$ for each pixel in f, we wish to learn a sparse dictionary of global texture models for characterizing the underlying texture distribution of f in a concise and simplified manner. To accomplish

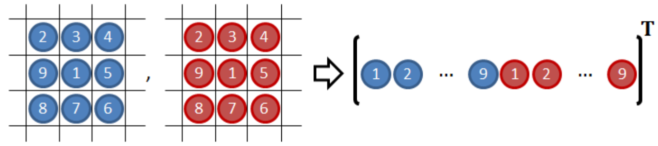


Fig. 1: An example of constructing a texture representation from a pixel neighborhood with one radial layer and two channels (blue and red). The pixels are extracted layer-by-layer to form a feature vector of pixel values. The numbered bubbles represent pixels with numeric labels for visualization purposes. Note that this representation does not include spatial constraints.

this goal, we introduce a dictionary-based strategy for learning such a dictionary, where a polarimetric SAR image f is summarized by a set of n_d texture dictionary elements $D = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{n_d}\}$, where $n_d \ll M \times N$. To learn the set of dictionary elements (i.e., D), we introduce a two-component texture-spatial cost criterion:

$$D = \arg\min_S \sum_{j=1}^{n_d} \sum_{g(\cdot) \in S_j} \Big[C\left(g^t(\mathbf{x}_i), \mathbf{d}_j^t\right) + \beta C\left(g^s(\mathbf{x}_i), \mathbf{d}_j^s\right) \Big] \quad (2)$$

where S_j is the set of spatially-sensitive textural representations being summarized by dictionary element $\mathbf{d}_j$; g^t and g^s are the texture and spatial components of the spatially-sensitive texture representation g, respectively; $\mathbf{d^t}$ and $\mathbf{d^s}$ denote the texture and spatial components of the dictionary element $\mathbf{d}$, respectively; $C(\cdot, \cdot)$ is some cost function describing the discrepancy between two vectors; and β controls the influence of the spatial components relative to the texture component on the learned model. The formulation presented in (2) can be satisfied using the k-means clustering with the Manhattan (i.e., L_1) distance cost function and n_d clusters. The final dictionary D is therefore a set of cluster centroids that are representative of their respective clusters of spatially-sensitive texture representations.

5. CLASSIFIER LEARNING BASED ON GLOBAL TEXTURE MODEL

Given the global texture dictionary D describing image f, we wish to learn an unsupervised classifier that generates the set of $n_l \ll n_d$ land cover labels $L = \{1, 2, ..., n_l\}$ for each pixel based on the pixel's texture description. During this phase, we omitted the spatial components of the dictionary elements in D and therefore used only the texture components D^t. Note that the classifier is never given land cover classes (e.g., stem-beans), but rather is given the desired number of classes. The resulting label map consists of pixels that are grouped according to similar texture descriptions. To generate n_l labels, we used k-means with n_l clusters, which minimizes the withinclass variance of the clusters. Each pixel in f was then assigned a label $l \in L$ based on its representative cluster. These labels are not specific agricultural classes, but rather are arbitrary labels intended to group similar pixels based on their local texture characteristics.

6. EXPERIMENTAL RESULTS

The proposed method was implemented in MATLAB and validated using the real-valued Pauli decomposition of the polarimetric L-band Flevoland, Netherlands dataset provided by NASA/JPL AIRSAR [8]. The Pauli decomposition was computed in the same way as [4]. The image is 862×625 pixels with 15 manually-defined ground-truth land cover classes obtained from [4] denoting the various crops. The Pauli RGB composite of this data and ground-truth classification are shown in Fig. 2.

For robustness and speed, we divided the entire image into distinct partitions, with each of which we employed the dictionary learning stage. The elements learned across all partitions were merged together for the final classification stage. We used 206×206 pixel partitions, 4-layer neighborhoods, $\beta = 50$, and $n_d = 100$ global dictionary elements (which were determined empirically) to generate 18 class labels. We generated more labels than the number of ground-truth classes because there are more visually distinct classes than those identified in the scene [4]. Upon obtaining these labels, we mapped each unsupervised label to the ground-truth class with which it shared the largest amount of pixel coverage so that accuracy metrics can be compared.

The final classification result using the proposed algorithm was first compared against classification using k-means and maximum a posteriori (MAP) with Gaussian Mixture Models (GMM), which are shown in Fig. 3. The proposed method generated a label map that labeled pixels very cohesively compared to the benchmark algorithms, which is consistent with agricultural growth patterns. It attained much higher accuracy (71.8%) than the k-means (46.7%) and MAP (46.6%) methods. We also compared our results to a recent unsupervised classification method (PolarIRGS) [4] and a supervised method [9] that use the same dataset. The proposed method attains slightly higher accuracy than PolarIRGS which uses fully polarimetric data (71.8% *vs.* 69.8%). Lee *et al.* attain higher accuracy (81.63% *vs.* 71.8%) using fully complex polarimetric L-band data, however their classifier is supervised. Also, Lee *et al.* attain a maximum of 60.12% using magnitude data only, the same type of data used by the proposed method to achieve a higher rate of 71.8%.

7. CONCLUSIONS

This paper has proposed a sparse texture dictionary learning scheme for unsupervised land-cover classification on polarimetric SAR imagery. As this approach uses unsupervised learning (i.e., the ground-truth labels were not incorporated into the training phase), the learned pixel-to-label classifier is unbiased toward human interpretation. The resulting label map was shown to be much better suited to the spatial growth patterns of crops than traditional pixel-based methods. Attained accuracy metrics using three-channel polarimetric SAR data were higher than recent unsupervised and supervised methods using the same data [4, 9]. Future work involves automatic parameter determination, comparisons with different dictionary learning methods, and analysis using other data sets.

8. REFERENCES

[1] P. M. Vitousek, “Beyond global warming: ecology and global change,” *Ecology*, vol. 75, no. 7, pp. 1861–1876, Oct. 1994.

[2] T. Blaschke, “Object based image analysis for remote sensing,” *ISPRS J. Photogrammetry and Remote Sens.*, vol. 65, pp. 2–16, Jan. 2010.

[3] S. W. Myint *et al.*, “Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery,” *Remote Sens. of Environment*, vol. 115, no. 5, pp. 1145–1161, May 2011.

[4] P. Yu, A. K. Qin, D. A. Clausi, “Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1302–1317, Apr. 2012.

[5] W. Zhou, G. Huant, A. Troy, M. L. Cadenasso, ”Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study,” *Remote Sens. of Environment*, vol. 113, no. 8, pp. 1769–1777, Aug. 2009.

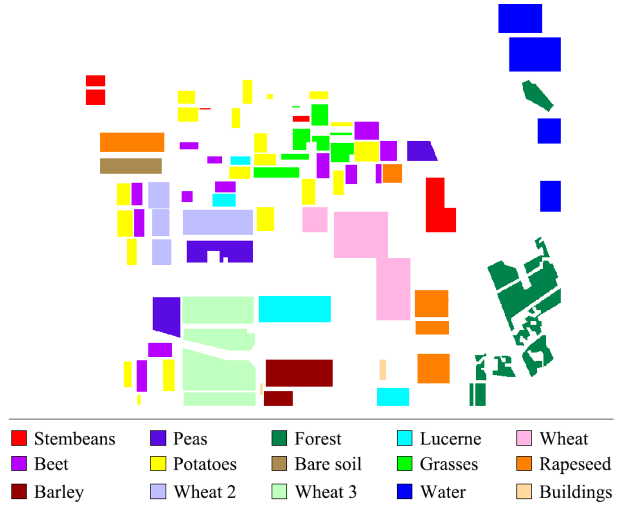
[6] A. M. Bruckstein, D. L. Donoho, M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[7] J.-S. Lee and E. Pottier, *Polarimetric RADAR Imaging: From Basics to Applications*. Boca Raton:CRC Press, 2009.

[8] “IEEE GRSS Data Fusion reference database, data set GRSS_DFC_0004,” 2000. [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/

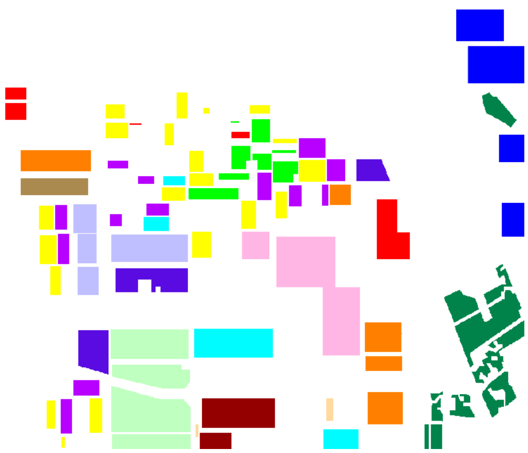
[9] J.-S. Lee, M. R. Grunes, E. Pottier, “Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization SAR,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2343–2351, Nov. 2001.

(a) Flevoland Pauli composite SAR image (L-band)

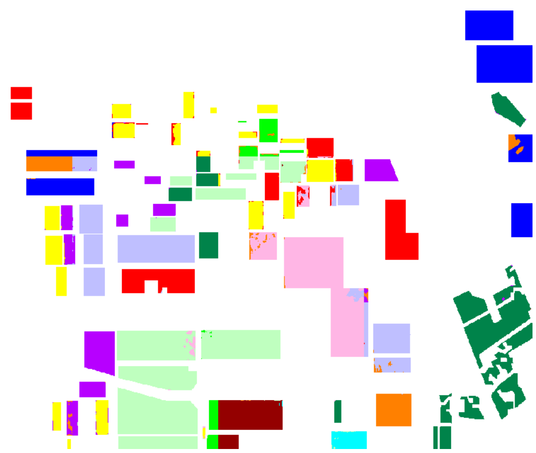


(b) Ground-truth

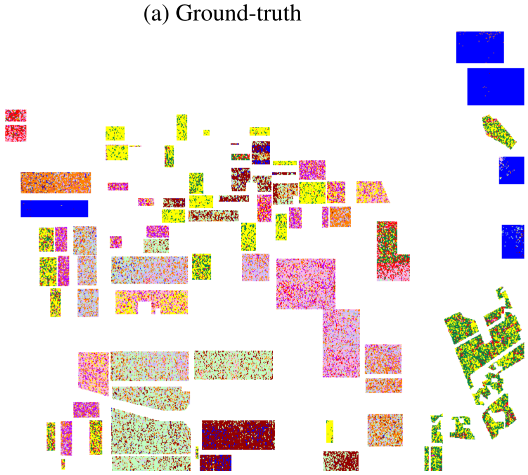
Fig. 2: Flevoland, Netherlands polarimetric SAR data with manually-labeled ground-truth land cover.



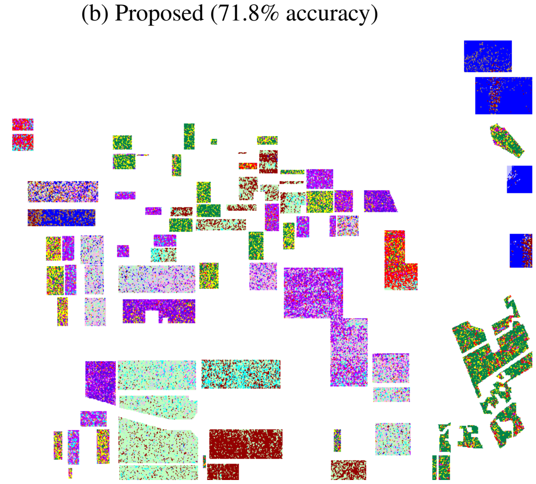
(a) Ground-truth



(b) Proposed (71.8% accuracy)



(c) k-means (46.7% accuracy)



(d) Maximum a posteriori (MAP) (46.6% accuracy)

Fig. 3: Comparing the land cover classification results generated by the proposed algorithm and two benchmark algorithms. Notice how the proposed algorithm groups texture patches into natural coherent units, whereas k-means and MAP generate very localized pixel-by-pixel labels.