

Automated identification of natural and cultivated vegetation based on LiDAR-derived image texture

Randall (Sky) Jones

Faculty Adviser: Dr. Henrique Momm

## **Introduction**

Natural vegetation is fundamental to the health of any watershed; it filters sediment and nutrients from runoff, stabilizes stream banks, controls sediment erosion, shades and cools water, and provides habitat and food for a diverse array of terrestrial organisms. On the other hand, cultivated vegetation can be a poor soil stabilizer, provides limited habitat and implies the application of potentially detrimental pesticides, herbicides and fertilizers. Because of this, rapid and accurate quantification of both the extent and character of vegetation in a watershed is of great interest to hydrologists, watershed planners, agronomists and other stakeholders.

The advent of aerial and satellite imagery has been crucial in allowing rapid characterization of vegetation strictly based on spectral information. However, the limited spectral characteristics of sensors with high spatial resolution have suboptimal ability to distinguish between natural and cultivated vegetation, which may have similar spectral signals despite having obviously different spatial arrangements. Some more comprehensive techniques use zonal properties to quantify “textural” (locally aggregated) information, allowing investigators to use incorporate spatial statistics into their land use predictions.

In this study, I propose a new vector-based method to distinguish natural and cultivated vegetation in a semi-automated fashion. By evaluating the linearity of clusters of vegetation, an “orderedness index” can be calculated and therefore used to classify whether vegetation is natural or cultivated. The purpose of this project is to explore the optimal implementation of this technique, its effectiveness compared to other techniques, and its general limitations. While natural vegetation is highly irregular, cultivated vegetation is expected to show a high degree of linearity due to the regular planting

patterns preferred by producers. This strategy has the advantage of requiring only LiDAR datasets as input.

## **Background**

Spectral methods have dominated landcover classification schemes ever since aerial and satellite imagery became widely available, and the bulk of published schemes are spatially and temporally agnostic. That is, the schemes consider only the pixel-based spectral signals received and ignore the variations surrounding those pixels. More recently, investigators have made attempts to integrate spatial variation (Momm, 2009), time variation (Bargiel, 2011) and both spatial and temporal variation (Zhai, 2018) in order to improve landcover classification, particularly vegetation classification. These techniques greatly improve landcover classification, but generally require high resolution multispectral imagery and training of supervised machine learning algorithms, which limits its general applicability beyond the study site.

While spectral methods are based on multi-channel raster grids, LiDAR data consists of three-dimensional point clouds usually collected from an airplane. Each point is described by a series of attributes in addition to spatial coordinates. LiDAR-based methods have been used for landcover classification (Helmer, 2008), and LiDAR-spectral composite methods have been proposed as well (Sturari, 2017). Like spectral methods, LiDAR methods typically require training a supervised machine learning algorithm.

Both LiDAR and spectral methods have difficulty in explicitly capturing spatial variation, particularly spatial entropy. Even image-based textural methods that create aggregate statistics based on neighboring pixels are agnostic to the arrangement of the pixels, and most contemporary measure of image entropy/disorder do not accurately capture spatial disorder (Razlighi, 2009). This presents a problem when attempting to differentiate natural and cultivated vegetation, which sometimes varies primarily based on spatial arrangement.

Because of this, a spatially-capable algorithm would be of great use for hydrologists, watershed planners, agronomists and anyone else interested in quantifying crop and vegetation impact in a watershed. Such an algorithm would only require inputting the locations of vegetation, rather than a multilevel array of data derivatives, allowing accurate classification of locations where high-resolution spectral data is unavailable or too costly to acquire. General vegetation identification is a well-studied problem, and herbaceous vegetation and even individual trees can be identified with simple color imagery or LiDAR (Chang, 2013).

It is interesting to note that at least some exploration of land use classification using lacunarity measures has been done (Myint, 2006). Lacunarity is variously described as a measure of “gappiness”, rotational invariance or heterogeneity. Though lacunarity-only classification has mixed accuracy, its ability to quantify heterogeneity (order) may be of interest to this project. It has not been applied directly to differentiating natural and cultivated vegetation, and the measure is not spatially explicit, but it may be explored further in this project since measures of lacunarity are unopinionated about pattern shapes, while the proposed method is.

I have an extensive background with environmental modeling, with projects that range from a fuzzy-logic based sedimentation model to a Python package for analyzing and designing stream restoration projects. I also worked for three years at an environmental consulting firm where I designed and implemented a machine learning models to predict locations and qualities of potential stream and wetland restoration projects.

## **Purpose**

The goal of this project is to develop a set of instructions (an algorithm) that can be used to distinguish natural vegetation from cultivated vegetation purely on the basis of its spatial relationship to other vegetation. Thus, the project will entail testing different aspects of the proposed methods, quantifying its predictive power, and comparing the results to spectral-based techniques. If possible, it

would be desirable to produce a Python package that implements this algorithm, making it available to a wide audience.

## **Methods**

Identifying vegetation using LiDAR-derived digital height models (DHMs) is well-studied, and not the focus of this project. Rather, this project is focused on using the relative spatial arrangement of vegetation to classify it as natural or cultivated. To explore this, a study area in middle Tennessee has been selected. The study area encompasses an apple orchard, a natural forest, and an area of cultivated coniferous trees. By using a digital height model derived from a LiDAR dataset, individual tree canopies can be identified.

Once tree locations are identified, the proposed method to identify a given tree as natural or cultivated is as follows: find all other trees within a certain radius. Then, use a clustering algorithm to identify natural groups of trees. Finally, find the line that best fits each cluster, and use each tree's deviation from its cluster's line to determine if the cluster is irregular or linear. Linear clusters imply the tree is cultivated (rows of trees in an order), while irregular clusters imply that the tree is natural. Techniques such as hierarchical clustering and stochastic optimization may be used within the model and to adjust model parameters.

The proposed method is noteworthy because it takes into explicit consideration the spatial arrangement of vegetation, in contrast to the non-spatial spectral methods that currently dominate landcover classification schemes. If this method proves successful in the proposed study area, then it might also be applied to herbaceous vegetation, which shows a similar spatial dichotomy between row crops and natural herbaceous vegetation. Other methods such as spatial entropy and lacunarity-based analyses may also be explored as they are currently understudied for the purpose of vegetation classification.

## **Mentor Collaboration**

Dr. Henrique Momm will be my mentor for this project. We are already collaborating together on a project that is focused on the identification of riparian vegetation using LiDAR-derived datasets. While working on this project, we realized that our model was unable to differentiate natural and cultivated vegetation which limited the model's effectiveness. This proposed project, though separate from our ongoing work, would be able to address this limitation. Dr. Momm and I already have ongoing weekly meetings where we discuss the status of our work. Because Dr. Momm had a critical role in developing the texture-base methods discussed elsewhere in this proposal, he is well-suited to advise on the proposed project.

## **Previous URECA Projects**

This proposal differs significantly from past URECA projects. No previous URECA projects have involved the analysis of remotely sensed data, and projects in the earth and environmental science have been scarce in past years. Of the past earth and environmental science projects, none investigated land use classification schemes or spatially explicit algorithms.

## Citations

- Bargiel, Damian, and Sylvia Herrmann. "Multi-Temporal Land-Cover Classification of Agricultural Areas in Two European Regions with High Resolution Spotlight TerraSAR-X Data." *Remote Sensing*, vol. 3, no. 5, 2011, pp. 859–877., doi:10.3390/rs3050859.
- Chang, Anjin, et al. "Identification of Individual Tree Crowns from LiDAR Data Using a Circle Fitting Algorithm with Local Maxima and Minima Filtering." *Remote Sensing Letters*, vol. 4, no. 1, 2013, pp. 29–37., doi:10.1080/2150704x.2012.684362.
- Helmer, Eileen H. "Mapping Land Cover and Estimating Forest Structure Using Satellite Imagery and Coarse Resolution Lidar in the Virgin Islands." *Journal of Applied Remote Sensing*, vol. 2, no. 1, Jan. 2008, doi:10.1117/1.3063939.
- Momm, H.g., et al. "Evaluation of the Use of Spectral and Textural Information by an Evolutionary Algorithm for Multi-Spectral Imagery Classification." *Computers, Environment and Urban Systems*, vol. 33, no. 6, 2009, pp. 463–471., doi:10.1016/j.compenvurbsys.2009.07.007.
- Myint, Soe W., et al. "Urban Textural Analysis from Remote Sensor Data: Lacunarity Measurements Based on the Differential Box Counting Method." *Geographical Analysis*, vol. 38, no. 4, 2006, pp. 371–390., doi:10.1111/j.1538-4632.2006.00691.x.
- Razlighi, Q. R., and N. Kehtarnavaz. "A Comparison Study of Image Spatial Entropy." *Visual Communications and Image Processing 2009*, 2009, doi:10.1117/12.814439.
- Sturari, Mirco, et al. "Integrating Elevation Data and Multispectral High-Resolution Images for an Improved Hybrid Land Use/Land Cover Mapping." *European Journal of Remote Sensing*, vol. 50, no. 1, 2017, pp. 1–17., doi:10.1080/22797254.2017.1274572.
- Zhai, Yongguang, et al. "Land Cover Classification Using Integrated Spectral, Temporal, and Spatial Features Derived from Remotely Sensed Images." *Remote Sensing*, vol. 10, no. 3, Jan. 2018, p. 383., doi:10.3390/rs10030383.

## **Timeline**

9/9/19: Download study area data, generate canopy model, vectorize tree locations and type

9/16/19: Build prototype model, explore effectiveness

10/14/19: Explore alternate methods (lacunarity)

11/11/19: Compare new models against existing models

1/13/20: Formal validation, begin working on paper for publication and Python package

2/13/20: Explore integration of new method with existing (spectral, LiDAR) methods

3/30/20: Publish Python package, send paper for review

5/15/20: Project finished, including URECA documentation