Counting Trees:
Classification of Open Woodland and Forest in the Lower Amazon floodplain

Katy Waechter

University of Denver

Department of Geography

and the Environment

Advanced GIS
Final Project Report

UNIVERSITY of DENVER
NATURAL SCIENCES & MATHEMATICS

Abstract. Mapping vegetation in the Lower Amazon mainstem floodplain is difficult. Only forest can be confidently identified and consistently mapped. Since the Lower Amazon floodplain is less than 30% forested, non-forest land cover classes need to be classified. Using object-based image analysis from high resolution multispectral imagery, a new analytical tool, the Tree-based Classification tool, distinguishes between open woodland and forest land cover. Input data for the Tree-based Classification tool includes polygon object delineations of individual tree crowns as well as spectrally homogeneous neighborhoods (segmented units), generated by eCognition image processing software. The tool is outlined and reviewed throughout the report, but remains to be fully implemented with actual image objects. The Tree-based Classification tool uses existing functions of Arcpy tools to characterize the spatial relationships of tree crowns and segmented units. The tool is not suitable for upland or coniferous forests and may only be used with image objects (no pixel-based classification). The tool also cannot be used to distinguish herbaceous vegetation and land cover surfaces without trees. The development tool is a significant but preliminary step in creating a detailed land cover classification for the Lower Amazon floodplain.

Keywords: floodplain, varzea, Amazon, open woodland, forest, high resolution imagery, tree crown delineation, land cover classification, OBIA

Introduction

The floodplain (varzea) of the main stem Amazon River is one of the largest riverine wetlands in the world, comprised of complex and dynamic landforms in a mosaic of permanent and seasonal lakes, channels, levees, and flats as shown in Figure 1. The varzea houses unique flora and fauna and provides riverside communities with critical ecosystem services. Dominant economic activities in the area include commercial fishing and livestock grazing, the impacts of which are causing habitat degradation (McGrath et al. 2007). In the past forty years, economic development has substantially transformed the floodplain ecosystem. The primary drivers are the growth of commercial fisheries and cattle/water buffalo ranching. Additionally, logging, agroforestry (açai palm, jute), and agricultural conversion are significant drivers of land cover change (Asner et al. 2003). These changes occur at mostly small-holder scales, influenced by individual household economic strategies.

In order to understand long-term impacts of increased economic activity (e.g., increased cattle density, overfishing), particularly habitat degradation and deforestation, a detailed mapping of land cover in the varzea is needed. The main stem Amazon River varzea is covered by a variety of vegetation physiognomies, including terrestrial and aquatic herbaceous, shrub, and trees. The varzea occupy a smaller total area compared to the upland forests, but their contributions to biodiversity and ecosystem services are greater than upland forests (Martinez and LeToan 2007). Forested areas of the lower varzea have decreased 13 percent since the 1970s while other land cover surfaces, such as non-forest vegetation and bare soil, have increased (Réno et al. 2011).

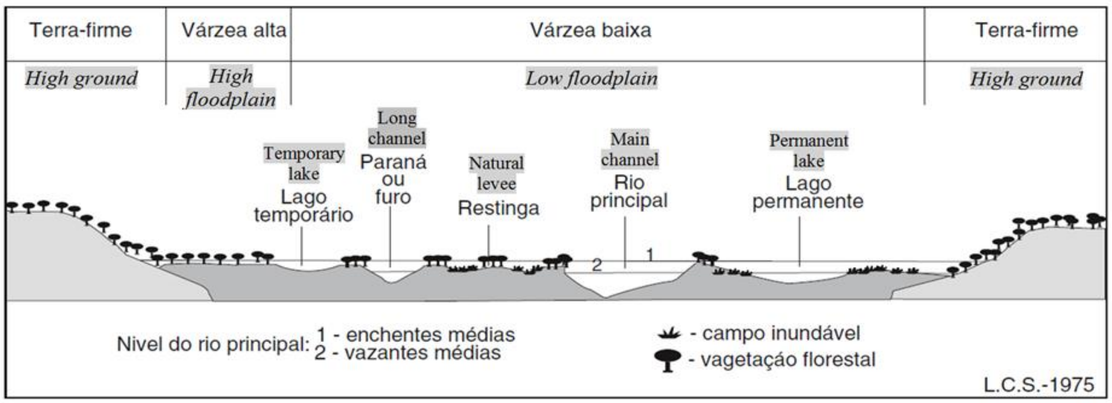

Figure 1. Longitudinal cross-section of the varzea contrasts levees, flats, as well as forested uplands

Land cover surface extent must be known before assessment of impacts of past and current land cover change to communities and biodiversity. Flood extent and vegetation structure in the varzea can be efficiently mapped using synthetic aperture radar (Hess et al. 2003), but only optical imagery is available for multi-decadal range in the varzea. Forests can be consistently mapped using multiple methods with optical imagery, but ability to classify non-forest vegetation in the varzea has been limited by availability of cloud-free scenes and sensor resolution. This project uses object-based image analysis (OBIA) to address the difficulty of classifying detailed land cover surfaces in the varzea. The tool created uses outputs of image segmentation procedures to analyze patterns in tree crowns. From the results of the analyses of tree crowns, a land cover classification is created for each area of the varzea.

Significance

This study and tool, referred to as Tree-based Classification tool, are significant because a detailed classification of treed vegetation (open woodland and forest) is created. All previous studies have generated classifications of forest and non-forest vegetation, which is not sufficient for understanding land cover change through time. Regarding methodology, the Tree-based Classification tool is significant because it allows customization of classification rules and several pieces of information for exploratory spatial data analysis. While OBIA is an available analysis technique with at least one software package (Trimble eCognition), the tools of the existing software do not allow for any calculations to generate classification rules required for this study.

Study Area

When fully implemented, this study will use OBIA to classify woodlands and forest extent in two swaths of the lower varzea in western portion of Pará state (Figure 2). The eastern swath of the study area includes the city of Oriximina and the western swath includes five agro-extractive settlements in its extent. Test data presented in this paper were not generated from these scenes.

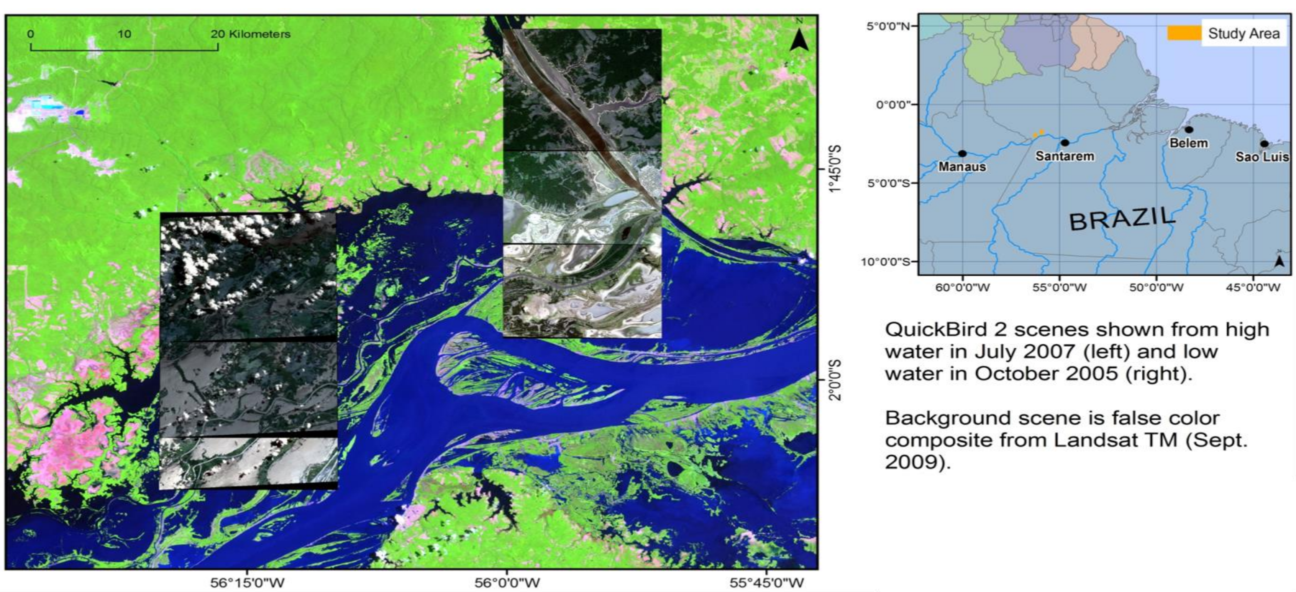

Figure 2. Image map of the study area

Design

The Tree-based Classification tool classifies land cover based on the spatial relationships of image objects. Image objects are pixels that are similar to one another based on a measure of spectral properties, including color, size, shape, texture, as well as context from a neighborhood surrounding the pixels (Burnett and Blaschke 2003). Prior to implementing the Tree-based Classification tool, image objects must be identified. **Individual tree crowns** are the first type of image object that must be segmented to use as input with the Tree-based Classification tool. The segmentation of individual tree crowns relies on a shadow following algorithm to delineate the extent of canopies and is considered robust in distinguishing individual crowns within tree stands. Individual tree crowns (Figure 3, Object Level 3) are smaller extent image objects than segmented units and are used to classify segmented units. **Segmented units** are the second type of image object that must be segmented to use as input with the Tree-based Classification tool. Segmented unit, for lack of a better phrase, are spectrally homogeneous areas based on context of neighboring pixels (Figure 3, Object Level 2).

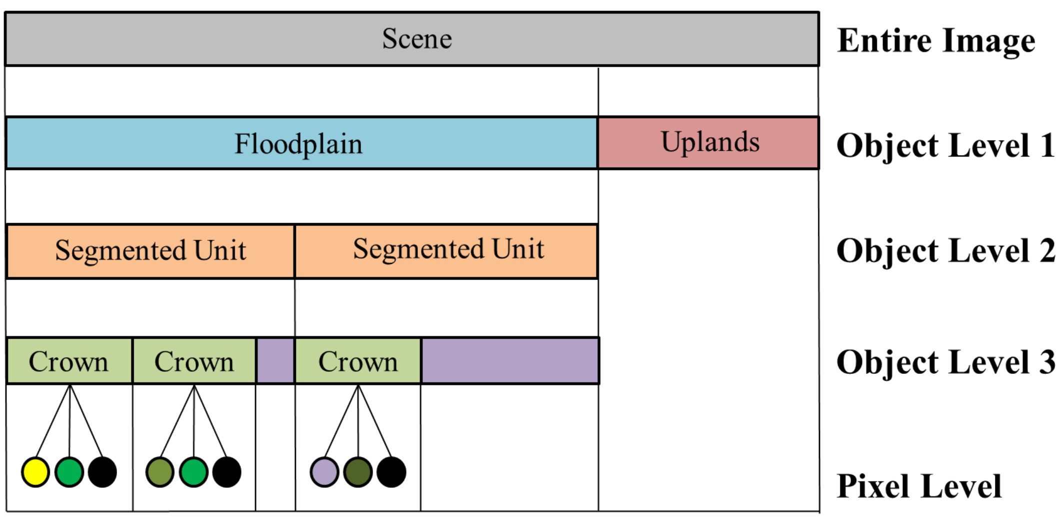

Figure 3. Image object hierarchy for the Tree-based Classification tool

Tree crown image objects and segment unit objects are the two major input datasets for the Tree-based Classification tool. The original high resolution imagery that was used to derive the image objects is a supplementary input, which the tool describes in the tool log for tracking purposes. The Tree-based Classification tool and its results are organized into customized directories based on the original image the image objects were segmented from. The in a base folder in which the input data are placed and other directories are housed, an Output folder that contains all of the analysis results and formatted data, and a Processing folder that is deleted after the tool finishes to minimize extraneous files.

The Tree-based Classification tool uses a mix of built-in functions from Geometry, Analysis, Spatial Analyst, and Conversion tools to describe tree crowns, intra-feature class spatial relationships of tree crowns, inter-feature class spatial relationships of tree crowns to segmented units, and ultimately classify land cover surfaces. The tool creates the following output: (i) polygon feature class of classified segmented units, (ii) polygon feature class of individual tree crowns with additional descriptive fields, (iii) polygon feature class of cluster and outlier analysis of tree crown diameter, (iv) polygon feature class of cluster and outlier analysis of gap-crown ratio, (v) polygon feature class of cluster and outlier analysis of gap-crown pattern, (vi) a table of various tree crown sample statistics for exploratory data analysis, (vii) point feature class of tree crown centroids for visualization, and (viii) text log file that tracks input data, parameters, and any error messages.

Implementation

There are five independent sections of the Tree-based Classification tool: Scene Description, Creating Image Object Geometry Fields, Cluster and Outlier Analyses, Land Cover Classification, and Centroid Creation. There are additional functions like creating and deleting folders, saving copies of files to folders, and check in/out of extensions, which will not be included in discussion of the major functions of the Tree-based Classification tool. The tool can be used in the Python IDLE (run "IDLEFormat.py") or as a tool through ArcGIS 10 GUI (use "OBIA_Am.tbx" and "ArcFormat.py"). The functions of the tool are the same in both versions.

To test the functionality of the Tree-based Classification tool, sample data was created.

Figure 4. Sample data (trees are shown as points to ease visibility)

Sample data was used in lieu of actual derived image objects to expedite tool run time and because the software needed to perform image segmentation was still being installed by Steve Yi at the time of writing this report. The sample data, shown in Figure 4, include a tree crown polygon feature class with seven features and a segmented unit polygon feature class with four features, including a multi-part feature. The sample data includes an object ID number and shape fields. Every other piece of information is generated by the Tree-based Classification tool.

Scene Description

While describing the original scene that image objects were derived from is non-essential to the process of classifying land cover and analyzing tree crowns, it is helpful for tracking and identifying potential problems. If there are not four bands for each input scene, each band has a different spatial resolution from others, the spatial resolution is not 60 centimeters, or the spatial reference is not projected, there are problems with the input data that require addressing. Reviewing the results of this section of the Tree-based Classification tool will be required for quality assurance.

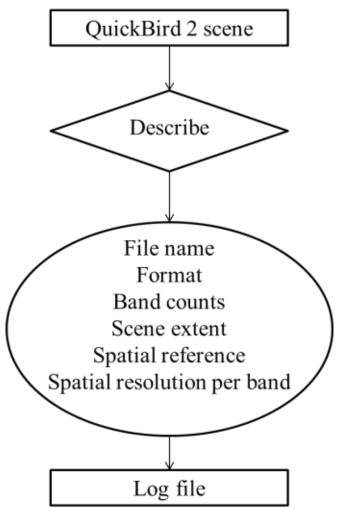

Figure 5. Scene Description section workflow

This section of the Tree-based Classification tool characterizes a multi-band raster (Quickbird 2 scene) with multiple functions of the Describe tool of the Arcpy module, including basename, format, bandCount, extent, spatialReference, meanCellHeight, and meanCellWidth, to retrieve information about the scene and each band. Figure 5 illustrates the workflow. The retrieved information is written to the tool log file, which can be found in the Output folder.

Creating Image Object Geometry Fields

This section of the Tree-based Classification tool is the shortest bit of code (13 lines) and most crucial processing step to classify woodlands and forest based on tree crowns. Figure 6 illustrates the workflow for this section of the tool. The geometry of each input feature class, Segmented Units and Tree Crowns, is calculated using the AddGeometryAttributes tool (“Add Geometry”). For Segmented Units, only area in square meters is calculated. For Tree Crowns, area in square meters and centroid coordinates (inside polygon) are calculated. Additionally, the distance to the nearest tree crown is determined for Tree Crowns using the Near tool (“Near”), which appends the object ID and distance in meters to the nearest tree to each tree crown object’s record. Each function used in this section appends an additional field to every object record: the input feature classes are also the output. No extraneous files are created.

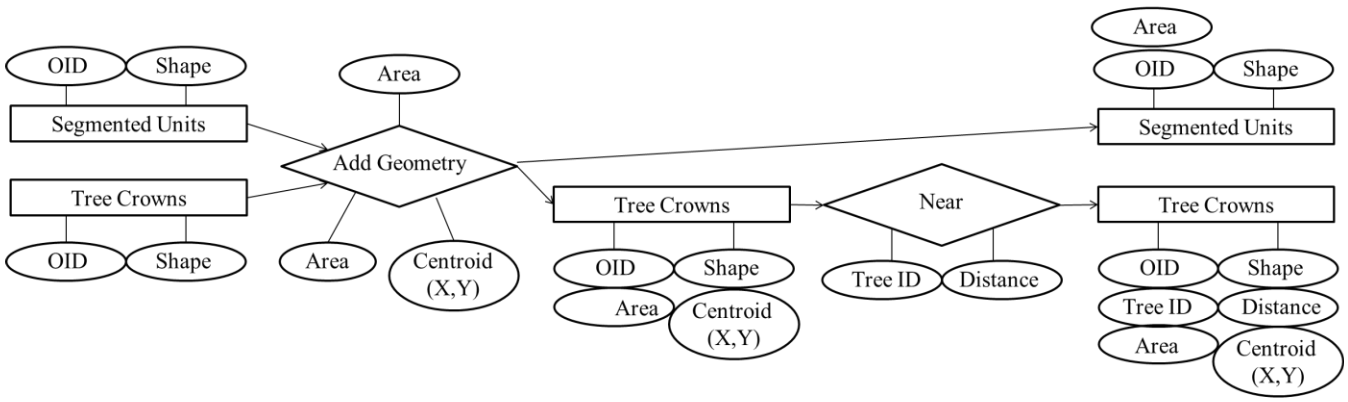

Figure 6. Creating Image Object Geometry Fields section workflow

Cluster and Outlier Analyses

The Cluster and Outlier Analyses section of the Tree-based Classification tool can be divided into two sections: (i) calculation of tree gap-crown ratio fields and (ii) cluster and outlier analyses of those fields. Figure 7 shows the workflow of both sections. The calculations of tree gap-crown ratio fields involve the calculation of three fields. First, an approximate tree crown diameter is calculated from the tree crown area by assuming circular geometry and solving for diameter ($D = \sqrt{\frac{4A}{\pi}}$), which is used to populate the Diameter field. Second, a ratio of gap distance to tree crown diameter is calculated by dividing the nearest tree distance by tree crown diameter, which is used to populate the Gap-Crown Ratio field. Third, a nominative field indicating cluster or dispersion is created and populated based on the gap-crown ratio. A gap-crown ratio of 30.0 and below is considered clustered, ratio of 50.0 and below is considered random, and a ratio above 50.0 is considered dispersed. To supplement these fields, a table including mean tree crown area, standard deviation of tree crown area, mean nearest tree distance, and standard deviation of nearest tree distance is exported for statistical testing (not included in this iteration of the Tree-based Classification tool).

From the three tree gap-crown ratio fields, local spatial statistics are generated using Anselin's LISA ("Cluster and Outlier Analysis"). Each local spatial statistics polygon feature class is saved to the COA folder. The results of these statistics can be used to assess effectiveness of classification rules and ultimate utility of a tree gap-crown distance ratio. Particularly, tree diameters are distributed very narrowly and show high positive spatial autocorrelation follow the pattern of dense, upland forest and should be reviewed per instance.

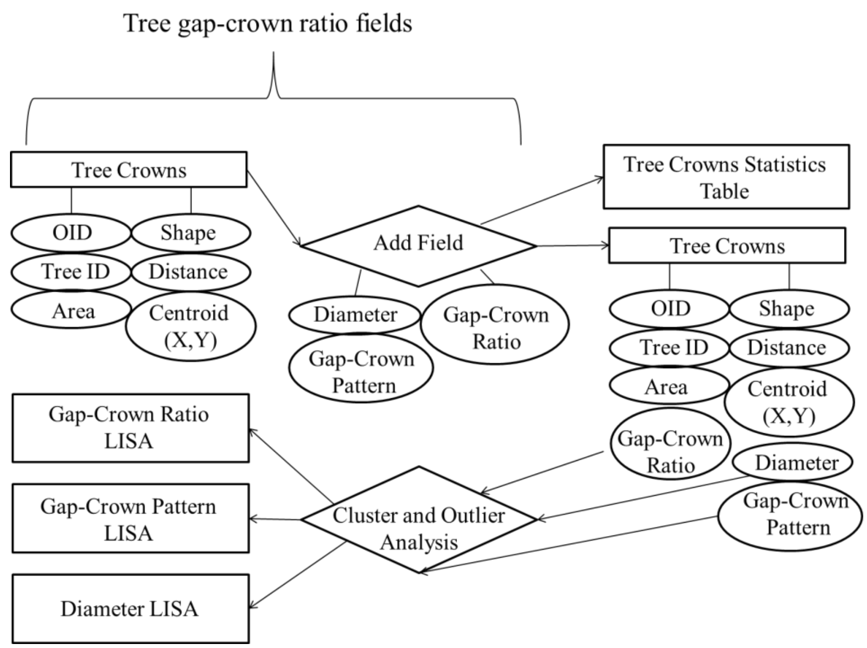

Figure 7. Cluster and Outlier Analyses section workflow

Land Cover Classification

In regards to the goal of this project, e.g. classification of woodlands and forest, this section provides the answer to the question of what areas are woodland and what areas are forest. The rules for the classification are based on the United Nations Food and Agriculture Organization’s Land Cover Classification System (LCCS). According to LCCS, landscape patches (synonymous with Segmented Units for the purpose of this project) can be designated as open woodland if their total area is 30-70% covered by tree and forest if their total area is 70% or greater covered by tree (DiGregorio 2005). Granted, this simple rule does not hold up when shrubs are confused with trees, which is very possible for Tree Crowns feature class in the varzea. A plan to discriminate shrubs from trees has not been incorporated yet into this iteration of the project. The plan may arise from analysis of tree crown diameter cluster and outlier analysis. Additionally, I am relying to an extent on the physiognomic differences between common shrubs and trees to prevent shrubs from being classified as trees.

Figure 8 shows the workflow for the classification section of the Tree-based Classification tool. First, individual tree crowns are identified by the segmented unit they inhabit (“Identity”). From the identified tree crowns, tree crown area is summed by OID by the identified segmented unit (“Sum Crown Area by Segmented Unit”). The summed area field from the identified tree crowns is joined to the segmented units feature class using the segmented unit OID as key (“Join TC Area Field”). A field is added to calculate the percent of each segmented unit covered by tree (“Add Percent Treed”). A nominative field is created to classify segmented units based on the percentage covered by tree (“Add Land Cover Name”). The nominative “LC_Class” is used as the classified land cover surface.

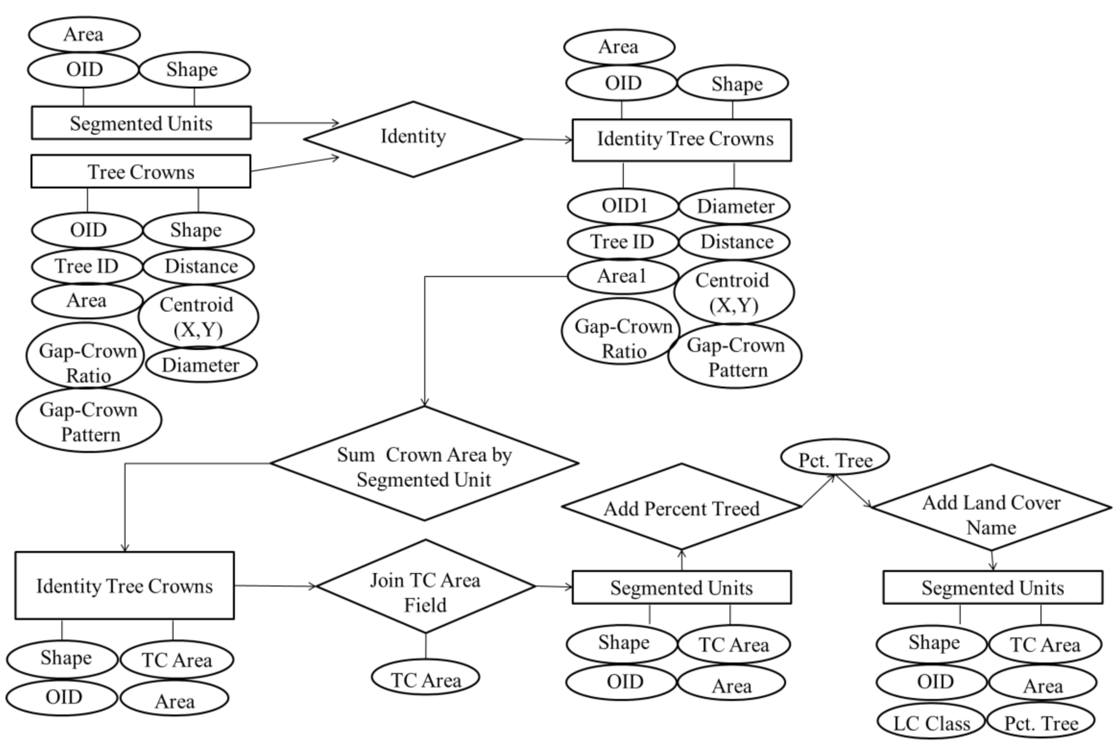

Figure 8. Land Cover Classification workflow diagram

Centroid Creation

Because it's difficult to quickly visualize volumetric attributes of polygon features, particularly if the feature is small compared to the map extent, a point feature class of tree crown centroids is created. As shown in Figure 9, this section uses the database file from the tree crowns polygon feature class to create an XY Event Layer ("XY Event Layer") to save to a layer file ("Save to LYR") with a specified symbology (imported from different layer file; "Apply Symbology"), then saves the layer file to a feature class ("Save to Feature Class"). With the centroid feature class and attributes carried over from the tree crown polygon feature class, a new field is created to designate small, medium, and large-sized trees. This field is helpful for size visualization, especially if there is confusion between trees and shrubs.

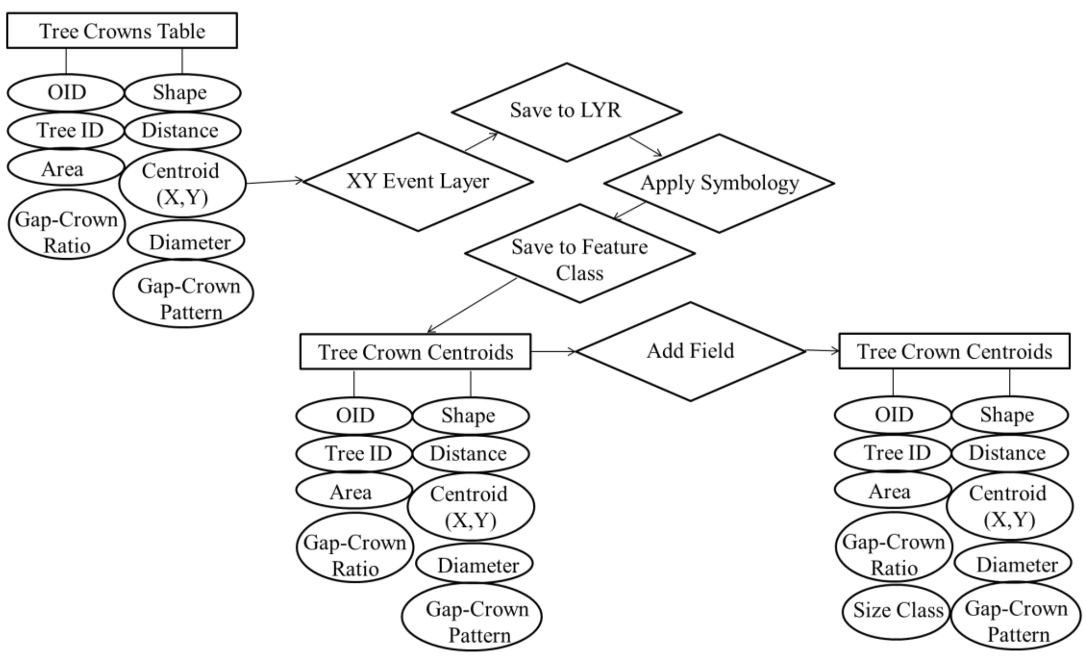

Figure 9. Centroid Creation workflow diagram

Many of the functions and steps of the Tree-based Classification tool are designed to generate information describing the spatial relationships and clustering of tree crowns based on multiple fields. While land cover classification is an important use of the tool, more information about spatial patterns is needed to corroborate and refine the classification rules used to distinguish between woodland and forest.

Discussion

The Tree-based Classification tool has been successfully developed. Results using sample data were meaningful based on the established goals of the project (Tables 1 and 2).

Table 1. Tabular results of Tree-based Classification tool for Tree Crown polygon feature class

OID	POLY_AREA	INSIDE_X	INSIDE_Y	NEAR_FID	NEAR_DIST	DIAMETER	GapCrRatio	GCR_Patter
1	54.84	604603.30	9762327.05	1	1131.85	8.36	135.45	Dispersed
2	144.25	605127.96	9763341.06	0	1131.85	13.55	83.52	Dispersed
3	237.68	604865.16	9765223.35	7	662.88	17.40	38.10	Random
4	129.26	604566.27	9766311.93	4	539.14	12.83	42.03	Random
5	593.22	605011.41	9765970.22	3	539.14	27.48	19.62	Clustered
6	870.37	605576.87	9765858.78	7	503.31	33.29	15.12	Clustered
7	244.80	605879.30	9765068.55	7	411.82	17.65	23.33	Clustered
8	275.71	605537.76	9765333.34	6	411.82	18.74	21.98	Clustered

Table 2. Tabular results of Tree-based Classification tool for the Segmented Unit polygon feature class

OID	POLY_AREA	PCT_TREE	LC_CLASS
1	1619294.37	51.0	Woodland
2	167989.21	15.6	Other
3	121680.11	73.1	Forest

While the sample data included eight tree crown objects and three segmented unit objects, the tool should still perform without error with thousands of tree crown polygons. Outside of tests for functionality, data generated by the Tree-based Classification tool have not been evaluated. However, the availability of additional information in assessing quality of classification and segmentation procedures is extremely helpful.

The utility of the Tree-based Classification tool is limited to tropical floodplains, specifically the Lower Amazon floodplain. It is not suited for vegetation types and land cover units not characterized by tree cover since the tool only analyzes tree crowns. This tool may have potential applications in tropical and temperate floodplains for detailed land cover classification for change detection studies, but will be of limited use for coniferous vegetation unless modified. Even under these restrictions, the Tree-based Classification tool may be shared with remote sensing scientists from INPE (Brazilian space agency). The tool will be shared inside an ArcGIS Toolbox where users will only have to enter input and change classification rules/parameters as needed. I plan on sharing the results of the classification through ArcGIS online, which limits vector datasets to 1,000 objects (unless pay for more).

Conclusion

The Tree-based Classification tool uses many functions of established tools in Arcpy to classify woodlands and forest in the Lower Amazon floodplain as well as to generate supplementary data to evaluate land cover classification. The tool analyzes the spatial relationships of trees and between trees and spectrally homogeneous areas. Taking skeleton spatial data files, the tool makes new fields based on calculations of area, proximity, and overlay to describe tree clustering, tree crown diameter related to tree gaps, and extent of tree cover. The goals of the project were accomplished and desired data can be created using real image objects. This tool is a significant step in preliminary land cover classification and requires evaluation of results.

In the future, I would like to add to the tool or build a supplementary tool to mine the volumes of single-date land cover classifications and compare them through time (object-based change detection). More robust understanding of spatial autocorrelation and clustering of trees in each type of detailed land cover is needed before undertaking a supplementary tool. This tool should be applicable to image objects derived from synthetic aperture radar and LiDAR data as well, which I would like to test. Additionally, results from adjacent scenes for individual dates should be combined to consolidate data and to reduce the impacts of edge effects.

References

Asner, G. P., M. M. C. Bustamante, and A. R. Townsend.(2003). Scale dependence of biophysical structure in deforested areas bordering the Tapajos National Forest, Central Amazon. Remote Sensing of the Environment 87,507-520.

Burnett, C. and T. Blaschke. (2003). A multi-scale segmentation/object relationship modeling methodology for landscape analysis. Ecological Modeling 168, 233-249.

Di Gregorio, A. (2005). Land Cover Classification System (LCCS), version 2: Classification Concepts and User Manual. FAO Environment and Natural Resources Service Series, No. 8 - FAO, Rome.

Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. Remote Sensing of Environment, 87(2003), 404-428.

Leckie, D. G., F. A. Gougeon, N. Walsworth, and D. Paradine. (2003). Stand delineation and composition estimation using semi-automated individual tree crown analysis. Remote Sensing of the Environment 85(3), 355-369.

Martinez, J. M. & Le Toan, T. (2007). Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. Remote Sensing of Environment, 108(2007), 209-223.

McGrath, D. G., Almeida, O. T., & Merry, F. D. (2007). The Influence of Community Management Agreements on Household Economic Strategies: Cattle Grazing and Fishing Agreements of the Lower Amazon Floodplain. International Journal of the Commons, 1(1), 67-87.

Parolin, P., L. V. Ferreira, A. L. K. M. Albernaz, and S. S. Almeida. (2004). Tree Species Distribution in Várzea Forests of Brazilian Amazon. Folia Geobotanica 39, 371-383.

Reno, V. F., Novo, E. M., Suemitsu, C., Renno, C. D., & Silva, T. S. (2011). Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery. Remote Sensing of Environment, 115(12), 3446-3456.