Boise Center Aerospace Laboratory Data

Boise State Data Sets

3-17-2016

2013 Morley Nelson Snake River Birds of Prey National Conservation Area RapidEye 7m Landcover Classification

Lucas P. Spaete
Boise State University

Nancy F. Glenn
Boise State University

Charles W. Baun
Idaho Army National Guard

Orchard Combat Training Center and Bureau of Land Management– Boise State University Vegetation Mapping with Multitemporal Satellite Data Report

Lucas Spaete (lucasspaete@boisestate.edu) and Nancy Glenn (nancyglenn@boisestate.edu)
BCAL, Boise State University

Overview

Boise State University conducted an area-wide vegetation classification of the Orchard Combat Training Center (OCTC) for the Idaho National Guard/IDARNG and expanded the classification to cover areas of the Morley Nelson Snake River Birds of Prey National Conservation Area for the Bureau of Land Management. This report documents the field data collection and processing, image acquisition and processing, and image classification. Work was performed between January 2012 – October 2015.

Field Data Collection and Processing

Field data collection methods were designed in consultation with IDARNG in summer 2012. The collection methods were designed to capture the diversity of vegetation types both at the plot and landscape scale for use with the remote sensing data.

Plot Selection

Plots of 20 m x 20 m were chosen to correspond to roughly 3 x 3 pixels in the RapidEye multispectral imagery (7 m pixel size). This design accommodated a 1-pixel georegistration mismatch. The locations were selected using a stratified random sampling approach based on targeted vegetation types identified as the species list from IDARNG. 150 plots were located across vegetation types in and around the OCTC. Additional plots collected in summer of 2015 were used for training to the west of the OCTC for the expanded vegetation classification. Also additional signature plots collected in the fall of 2015 were used to capture additional EXAN Areas.

Collected plots were screened to use only the most homogeneous plots of each cover type. Table 1 shows the number of signature plots used for each vegetation class for the final classification.

Table 1. Training data used for final vegetation classification

Vegetation Class	Number of plots
Agriculture	13
Bare ground/Roads	19
Cheatgrass	19
Cinder ground	7
Crested Wheat Grass/Planted Bunch Grass	3
Exotic Annuals	19
Kochia	12
Playa ground	8
Sandberg’s Bluegrass	13
Rabbitbrush	5
Sagebrush	29
Shadscale Saltbush	11
Water	7
Winterfat	10

Imagery and Pre-processing

Satellite imagery

RapidEye satellite imagery were selected as the primary remote sensing data for the project based on low cost, high spatial resolution, and available spectral resolution (5 bands, including 2 bands in the near-infrared). RapidEye Basic Product Level 1B imagery were selected in lieu of RapidEye Ortho Product Level 3A. Level 3A products are delivered as mosaics and cannot be atmospherically corrected. Level 3A products are also clipped to the study area. Level 1B products are delivered as complete scenes with accompanied information to be atmospherically corrected. Dates of image acquisition were selected in coordination with IDARNG to maximize seasonality differences among vegetation. Apollo Mapping, Boulder CO, was contracted to acquire the images. Five dates of imagery and a total of 10 images were collected over the course of the project (Table 2).

Table 2. RapidEye image collection dates. Images covering entire study area notated with (*); all other images cover portions of OCTC and are mosaicked by season.

Season	Number of images	Dates
Spring 2012	3	6/18/2012, 6/24/2012, 6/25/2012
Fall 2012	2	9/26/2012, 10/4/2012
Winter 2013	1	1/17/2013*
Early Spring 2013	2	4/9/2013, 4/12/2013
Mid-Spring 2013	2	5/20/2013, 5/25/2013*

After data collection, the images were preprocessed, georegistered, and mosaicked. Next, vegetation indices and band transformations were developed. These products, along with the original bands, were used in a classification (described below).

Imagery preprocessing and registration

RapidEye imagery were preprocessed using the ENVI image processing software suite. An automated orthorectification tool available for Rapideye was used along with a 10m NED DEM to orthorectify the images. The resulting 7 m rasters were converted to radiance, and top of atmosphere (TOA) corrections were applied (Figure 1).

Figure 1. Top of atmosphere equation used with RapidEye (BlackBridge V6.0, 2013) http://www.blackbridge.com/rapideye/upload/RE_Product_Specifications_ENG.pdf.

$$\rho = \frac{\pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\theta_s}$$

ρ_λ = Planetary directional TOA reflectance for lambertian surfaces [unitless]
π = Mathematical constant approximately equal to 3.14159 [unitless]
L_λ =Spectral radiance at the sensor's aperture [$W/(m^2\ sr\ \mu m)$]
d = Earth-Sun distance [astronomical units]
$ESUN_\lambda$ = Mean exoatmospheric solar irradiance [$W/(m^2\ \mu m)$]
θ_s = Solar zenith angle [degrees]

The May 25, 2013 image, covering the entire study area, was registered to the 2011 Idaho NAIP imagery using a 1st order polynomial transformation using 20 road intersection pairs in ArcGIS with a resulting RMSE of 2.5m (< .5 pixel). All other images were registered to the May 25 2013 image using the nearest neighbor interpolation method with varying RMSEs using ENVI (Table 3).

Table 3. RapidEye georegistration report.

Image	RMSE(Pixels 7m)	Tie points
6/18/2012	.249	20
6/24/2012	.34	20
6/25/2013	.28	20
6/30/2012	.09	18
9/26/2012	.22	20
10/4/2012	.265	20
1/17/2013	.36	20
4/9/2013	.179	20
4/12/2013	.204	20
5/20/2013	.181	20
5/25/2013		

Images from each season collection were mosaicked to generate one image for each season. Mosaics were then stacked to form a multi-layered RapidEye composite image for use in classification. Vegetation indices and image transforms from the mosaics were used for image classification.

Imagery products

Vegetation Indices: Vegetation indices were calculated for each season mosaic except winter 2013 using the ENVI Vegetation Index Calculator. The indices calculated include NDVI, SR, EVI, ARVI, and ARI1 (Table 4).

Table 4. Vegetation indices calculated from RapidEye and used in classification.

Normalized Difference Vegetation Index (NDVI)	$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}$
Simple Ratio Index (SRI)	$\mathrm{SR} = \frac{\mathrm{NIR}}{\mathrm{RED}}$
Enhanced Vegetation Index (EVI)	$\mathrm{EVI} = 2.5 \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + 6\ \mathrm{RED} - 7.5\ \mathrm{BLUE} + 1}$
Atmospherically Resistant Vegetation Index (ARVI)	$\mathrm{ARVI} = \frac{\mathrm{NIR} - 2\ \mathrm{RED} - \mathrm{BLUE}}{\mathrm{NIR} + 2\ \mathrm{RED} - \mathrm{BLUE}}$
Anthocyanin Reflectance Index (ARI)	$\mathrm{ARI1} = \frac{1}{550} - \frac{1}{700}$

Minimum Noise Fraction (MNF) transforms: Forward MNF transforms were calculated for each season mosaic to minimize noise and maximize the signal. The MNF is similar to a principal components analysis (PCA). Based on eigenvalue scores and visual inspection the first 2 MNF bands were included from each season mosaic.

Image Classification

A RandomForest (RF) classification method was used for image classification. Random Forest is a machine learning algorithm that uses a tree-based classifier method and is iterative in design to address limitations associated with overfitting and instability that can arise when using conventional classification tree-based approaches. Multiple bootstrap samples from the original training dataset are selected (with replacement) to generate multiple classification trees, and data are classified by taking the most popular voted class from all the tree predictors in the forest (Pal, 2005). The selection of variables at each node of the classification tree is based on a measurement of variable importance called the Gini index (Breiman et al., 1984). The Gini index represents a degree of node impurity, computed as the difference between out-of-bag error and the error from a permuted subset of data at each node. The ensemble learning approaches such as RF are preferred for “wide-data” such as in the present study with small sample size but a large number of predictors, as they result in smaller prediction variance and bias and better model performance. The RF includes a built-in robust validation using random subsets of both the data and the predictors bootstrapped several hundreds of times. Such random subsetting of predictor variables prevents problems associated with correlated variables (Brieman 2001). As such, the RF data modeling approach has been successfully applied in a wide range of studies (e.g., Mitchell et al., 2013, Pal 2005, Strobl 2009).

Classification RandomForest method was used to classify the vegetation into vegetation classes.

Model inputs

Imagery: The RapidEye bands, vegetation indices, and MNF data were stacked as individual bands into one file for each season. Each resulting seasonal file was then combined to create a multi-season image stack, hereafter called RapidEye Stack. A separate file was also created by combining the multi-season image stack and the LiDAR dataset, hereafter called LiDAR_RapidEye Stack.

Pre-Processing: Plot boundaries were used to extract pixel level data from the RapidEye Stack using the ENVI ROI tool. The resulting pixel data were combined with the vegetation class assignment for each plot in an Excel comma separated variable file as direct input into the RandomForest software.

Classification RandomForest: Vegetation

Classification RandomForest was used to identify the most important variables from the RapidEye Stack to predict vegetation classes. The resulting error matrix is displayed in Tables 5

Imputation

Vegetation classes were imputed using yaImpute, an R module. YaImpute applies the classification RandomForest models to the imagery using a built-in RandomForest distance matrix to assign each pixel value to the proper cover (species cover) or class (vegetation classification).

Resulting images had a 3x3 majority filter applied to remove noise. A 3x3 majority filter assigns each pixel the majority value based on the surrounding 8 pixels while maintaining the 7 m pixel size.

Class Names	Total Classes	Percent Correct	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Agriculture (1)	13	84.62%	11	0	0	0	0	1	0	0	0	0	0	0	0	1
Bare ground/Roads (2)	19	68.42%	2	13	0	0	0	0	0	2	0	0	0	1	0	1
Cheatgrass (3)	19	57.89%	0	0	11	0	2	0	0	0	2	2	2	0	0	0
Cinder (4)	7	100.00%	0	0	0	7	0	0	0	0	0	0	0	0	0	0
Crested Wheat Grass/Bunch Grass (5)	3	100.00%	0	0	0	0	3	0	0	0	0	0	0	0	0	0
Exotic Annuals (6)	19	57.89%	0	1	0	0	2	11	0	0	0	0	0	3	0	2
Kochia (7)	12	33.33%	1	0	0	0	0	0	4	0	0	2	0	3	0	2
Playa (8)	8	87.50%	0	1	0	0	0	0	0	7	0	0	0	0	0	0
Sandberg's Bluegrass (9)	13	92.31%	0	0	0	0	0	1	0	0	12	0	0	0	0	0
Rabbitbrush (10)	5	80.00%	0	0	0	0	1	0	0	0	0	4	0	0	0	0
Sagebrush (11)	29	65.52%	0	0	0	1	3	0	0	0	1	4	19	1	0	0
Shadscale Saltbush (12)	11	72.73%	0	0	0	0	0	0	2	0	0	0	0	8	0	1
Water (13)	7	100.00%	0	0	0	0	0	0	0	0	0	0	0	0	7	0
Winterfat (14)	10	50.00%	0	0	0	0	1	1	2	0	0	0	0	1	0	5
Total:	175															
Average:		75.02%														
Overall % Correct:		69.71%														

Table 5. Error matrix of vegetation classification

References

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5 − 32.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees (Monterey, Ca: Wadsworth).

Mitchell, J., Moore, C., & Glenn, N. (2013). Single and multi-date landsat classifications of basalt to support Soil Survey efforts, *Remote Sensing*, 5(10), 4857 – 4899.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*, 217 – 222.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application,and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4): 323 – 348.