Agricultural Research & Technology
Open Access Journal
ISSN: 2471-6774

Juniper Publishers
key to the Researchers

Research Article
Volume 3 Issue 1 – December 2016
DOI: 10.19080/ARTOAJ.2016.03.555602

Agri Res & Tech: Open Access J

Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification

Sudhanshu S Panda[1], Gerrit Hoogenboom[2] and Joel Paz[3]

[1]*Institute of Environmental Spatial Analysis, University of North Georgia, USA*

[2]*Agricultural and Biological Engineering, University of Florida, USA*

Submission: November 23, 2016; **Published:** December 09, 2016

***Corresponding author:** Sudhanshu S Panda, Professor, Institute of Environmental Spatial Analysis, University of North Georgia, 3820 Mundy Mill Road, Oakwood, GA 30566, USA, Tel.: 1-678-717-3594; Email: sudhanshu.panda@ung.edu

Abstract

Blueberry is the second most important fruit and nut crop in Georgia after pecans. The number of blueberry orchards for commercial production has increased in Georgia and other southeastern states during the past few years. Blueberry orchards are generally established in the area that has just been cleared from shrubs and forests. Most of the blueberry orchards in southeast Georgia are surrounded by pine forest. It is, therefore, difficult to distinguish blueberry bushes from other trees and shrubs with low resolution imagery such as 30 m Landsat TM or ETM. Many unmanaged blueberry orchards are generally also covered with tall grasses, making it more difficult to distinguish the bushes with low resolution imageries. The main objective of this study was to use advanced image processing techniques and high resolution multispectral imagery to distinguish blueberries orchards from other land-uses. Very high resolution 1-meter multi-spectral NAIP imagery was used to identify blueberry orchards in three southeastern counties (Bacon, Brantley, and Camden) in Georgia. Advanced image processing techniques, including principal component analysis and self-organizing map (SOM) neural network image segmentation technique, were used in this study. The unsupervised classified images were reclassified to four land-use classes, including 1) mature blueberry bushes/forest, 2) intermediate blueberries bushes/ scattered trees, 3) young blueberries bushes/grassland, and 4) prepared field for blueberry production /bare soil. The classification accuracy assessment was conducted with ground of blueberry orchards in three counties of Georgia. The Bacon County SOM classified image provided 77% producer's accuracy, 92% user's accuracy, 84% overall accuracy, and Kappa statistics of 0.44. The producer's accuracy, user's accuracy, and overall accuracy were 82%, 94%, and 88%, respectively, for the Brantley County SOM classified image along with the Kappa statistics of 0.53. A producer's accuracy of 92%, a user's accuracy of 94%, an overall accuracy of 95%, and Kappa statistics of 0.64 were obtained for the Camden County image classification. In this study we were also able to distinguish blueberry orchards (as a single entity) from the classified NAIP images with an accuracy of 100% for all three counties. This study suggests that different land cover patterns in commercial or large scale blueberry orchards can easily be distinguished with the use of high resolution remotely sensed images and advanced geospatial technology.

Keywords: Blueberry Orchard; NAIP; Principal Component Analysis (PCA); Self-Organizing Map (SOM); Accuracy Assessment; Textural Pattern Analysis; Global Positioning System

Introduction

Precision agriculture is defined as the observation, impact assessment, and the timely strategic response for remedy to minute variation in agricultural production [1]. Precision agriculture is applied in a wide range of agricultural activities such as crop production, dairy farming, horticulture, and forest management. Site-specific crop management (SSCM) is one important component of precision agriculture. The five main processes for a SSCM system are spatial referencing, crop and climate monitoring, attribute mapping, decision support system, and differential action. The process of potential management in crop production is better approached with the first step of data gathering at a spatial scale, i.e., spatial referencing of the crop field. Spatial referencing is performed with the use of geospatial technologies, such as remote sensing, geographic information system (GIS), and global positioning system (GPS).

Site-specific crop management is very common for traditional row crops such as maize, wheat, rice, cotton, and soybean [2-8]. However, SSCM for non-traditional horticulture crops is not very common [9,1]. Horticultural crops like fruit and nut crops are high value crops for which SSCM might have potential for increasing net returns and optimizing resource use. Oranges, peaches, pecans, apples, grapes and blueberries are major component of the agricultural production system in United States (US) and elsewhere in the world [10]. Blueberry is one such high value horticultural crops produced in southeast US and ranks second to pecan that needs the use of advanced geospatial technology for micromanagement.

Remotely sensed imageries are the best possible resource to delineate the spatial variation of crop land use, including horticultural crops. The application of high resolution remote sensing data, e.g., aerial or satellite imaging, along with GPS and GIS is the first step towards the goal of SSCM in fruit and nut crops [11]. Several studies have been conducted over the years to delineate or classify forests or shrubs from satellite and aerial Landsat images [12-17]. Studies have also been conducted to estimate the production of horticultural crops using satellite images classification, but only on a course scale [18,8]. Usha and Singh (2013) in their review article, showed the potential of remote sensing in horticulture crop management decision support. However, few studies have been conducted that distinguishes orchards from mixed forested land-use.

Remotely sensed images are a quick and sound means to detect fruit trees or orchards from other land-uses. conducted a study to distinguish olive tree orchards using remote sensing images by clustering assessment techniques. Scientists from the Space Application Center of the Indian Space Research Organization (ISRO) successfully used low resolution IRS LISS III and IRS AWiFS (23 m and 55 m, respectively) images to characterize apple orchards in India [19]. Shrivastava and Gebelein [20] per formed a study in Florida to classify land-use for the delineation of citrus groves for economic assessment. They were successful with the use of Landsat Enhanced Thematic Mapper Plus imagery and the results showed that there was a significant correlation between citrus production/income with remotely sensed imagery derived citrus area coverage, indicating that remotely-sensed estimates of citrus orchards can be used to forecast on-tree incomes to farmers [20]. O'Connel and Goodwin [21] used remotely-sensed imagery to identify the tree canopy of a peach orchard for yield forecasting and the estimation of crop water requirements. Beek et al. [22] used remote sensing to estimate homogenous pear orchard yield and fruit quality. Aguilar et al. [23] has identified crops in the greenhouse setting using object based image analysis using multi-temporal Landsat imageries. Noori and Panda [8] used ALOS-AVINIR image to create correlation models with olive tree trunk diameter, trunk height, soil plant analysis development, and leaf area index. However, almost all these studies are conducted with fruits and nut crops in homogenous orchards. No studies have yet been specifically conducted to distinguish blueberry orchards from mixed vegetation such as pine or other mixed forest that normally surround blueberry orchards in the southeast. Accurate classification of blueberry orchard land use would lead to its SSCM, i.e., crop yield estimation, growth enhancement, and other management decision support.

Remote characterization of blueberry orchards at a small farm scale and among forested land use is a demanding task. The challenges of distinguishing blueberry orchards in Georgia are enormous as the farms or orchards are very small and managed by mostly smaller producers. Panda et al. [9] demonstrated the efficiency of high resolution remotely sensed data (2.15 m Quick Bird imagery) to distinguish blueberry bushes from mixed vegetation. Therefore, it is hypothesized that high resolution remote sensing images can be analyzed to distinguish blueberry orchards and subsequent management decisions can be taken with the use of other geospatial technology application. Blueberry orchard delineation and spatial analysis using geospatial technology can provide a means for management decision making, such as fruit yield determination, exact and proper fertilizer and irrigation need quantification and scheduling, and diseases treatment. At the same time, it could maximize profits for farmers.

Blueberry is a wax-leaved shrub that grows closer to the ground compared to pine trees and other forest plants in the southeastern USA, but grows above wild grasses [9]. However, the blueberry orchards are row-crop plantations unlike the forests, which surround the orchards in southeast Georgia. In most cases there is bare soil in between the rows of blueberry bushes. Blueberry bushes are known to be strong forward-scatterers, whereas pines and other coniferous plants are both forward and backscatters [24]. Blueberry bushes can be distinguished with 680-700 nm (Red) and 800-9000 nm (NIR) bands of the electromagnetic spectrum [22,25]. Therefore, there is the probability that blueberry shrubs can be distinguished from larger forest trees based on their reflectance nature, as shown by Panda et al. [9]. The row crop nature of the plantation is expected to be advantageous in distinguishing the orchards based on the textural characteristics of the satellite or aerial images.

One must use advanced image-processing techniques and build models for the spectral classification of remotely sensed images to derive pertinent information for decision making [3]. Image processing techniques normally involve classification (clustering) and algebraic manipulation to determine the spatial variability of the image [26]. Clustering, i.e., grouping data with similar characteristics is a data mining technique that is used to reduce the data/digital information complexity in an image [27]. Unsupervised clustering is one of the desirable techniques for image segmentation, because of its associated advantages including a low human involvement and associated error, a high success at finding spectral clusters inherent in the data,

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

and being able to reveal unexpected categories from the image [28]. Unsupervised clustering algorithms that are used in image analyses include fuzzy C-means [29], K-means [30], ISODATA [28], self-organizing map (SOM) unsupervised classification [31], histogram thresholding [32], and region growing techniques [32, 33].

Kohonen's SOM is an unsupervised image clustering technique based on artificial neural networks (ANN) that is used as a data dimensionality reduction method with topological preservation [34]. Unlike most unsupervised clustering techniques that were described previously, the SOM constructs a topology preserving mapping from the high-dimensional space onto map units in a way that relative distances between data points are preserved (Panda, 2007). The map units or neurons, usually form a two-dimensional regular lattice where the location of a map unit carries semantic information. During the SOM clustering method, similarity in input patterns is preserved in the output space during the process of data compression and dimensionality reduction [35]. SOM unsupervised classification techniques have been successfully used for aerial and satellite image clustering [14,33,36-38]. In the early 1990's, Kanellopoulos et al. [39] conducted a study using ANN with SPOT high resolution visible imagery to discriminate land cover into 20 classes for a test site in Ardeche, France. Blueberry was one of the 20 classes that were analyzed. Based on the effectiveness of high resolution imagery and high-end image processing techniques on land-use classification, we hypothesized that blueberry orchards can be distinguished from heterogeneous land cover. The objective of this study was, therefore, to evaluate the potential use of 1-m resolution National Agriculture Imagery Program (NAIP) remotely sensed data along with principal component analysis (PCA) and SOM to distinguish blueberry orchards in mixed vegetation for site-specific crop management.

Methodology

Data acquisition and preparation

This study was conducted in the three highest blueberry producing counties of Georgia. These include Camden, Bacon, and Brantley counties, which are all located in southeast Georgia. Most of the blueberry orchards were located close to and around the main cities of these three counties. The blueberry orchards were relatively small in size compared to other traditional agronomic crops. Many orchards were found close to roads, in patches where the mixed forest cover were cleared, and in patches surrounded by pine and other mixed forest cover. Most of the blueberry orchards that were physically studied were located in and around the city of Alma in Bacon County, the city of Nahunta in Brantley County, and the city of Woodbine in Camden County.

A base map of the County boundaries for Georgia was obtained from the Georgia GIS Clearinghouse (http://www.gis.state.ga.us/). The city map of Georgia was also acquired from the same electronic data base. The high resolution satellite images consisted of the 2007 NAIP imageries of the aforementioned three study counties and were obtained from the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Geospatial Data Gateway (http://datagateway.nrcs.usda.gov/). Figure 1 depicts the counties that were studied and the cities where the blueberry orchards were located.

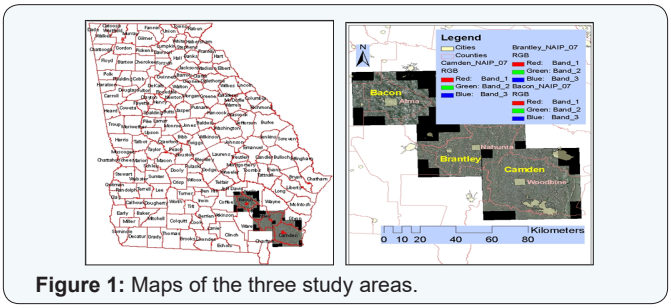


Figure 1: Maps of the three study areas.

NAIP acquires images during the agricultural growing season for the continental US. The images are obtained normally between the middle of April and the middle of September when most of the traditional row crops are grown [40]. These NAIP orthophotographs have a very high resolution and are taken with airplanes. They are later orthorectified with United States Geological Survey (USGS) Digital Ortho Quarter Quadrangles (DOQQs) and mosaiced together for data dissemination on County basis. Until 2007, the resolution of the NAIP imageries was 2 m. Since 2007, the resolution of the NAIP images is 1 m. We obtained the mosaiced County size imageries for our study. Prior to supplying the NAIP images to the public, they are also georeferenced to North American Datum (NAD) 1983 Universe Transverse Mercator (UTM) coordinate systems. Our three County images were georeferenced to the NAD 83 UTM Zone 17N coordinate system [41], which was the same as the other GIS layers that we have used in the study.

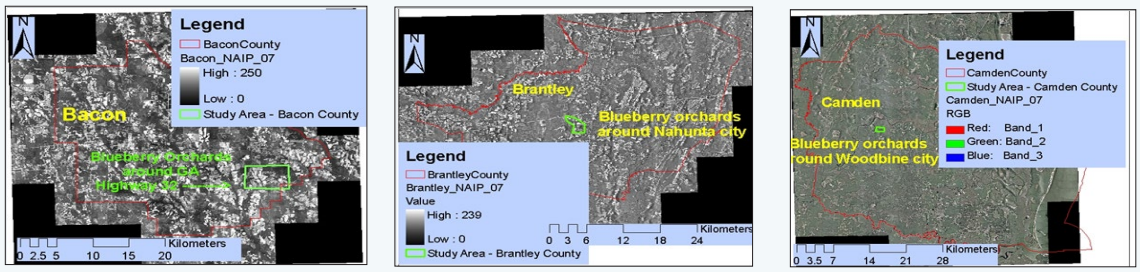
As stated earlier, the images are geometrically corrected by United States Department of Agriculture (USDA). It is to be noted that the NAIP images are also radiometrically corrected by USDA prior to release. The radiometric corrections are conducted with solar correction, i.e., centered on ground position where solar illumination and camera view angle are coincident, dark area subtraction, gain measurement, and mid-tone color alignment [35]. Therefore, these images are suitable for image segmentation and analysis. ArcGIS 10.2 (ESRI, Redlands, CA), ERDAS Imagine 2015 (Leica Geosystems Geospatial Imaging, Norcross, GA), and IDRISI Andes (Clarks Lab, Clarke University, Worcester, MA) software were used in our study for high end processing of end images.

Study area delineation

With a-priori field visit was conducted to obtain first-hand information on the blueberry orchard locations. During this visit, the orchard accessibility for ground truthing was ascertained. This field visit also allowed us to determine the correct size of our

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

images. Due to the image size limitations for image processing by the software, smaller study areas were selected in the three counties. The three study areas in the three counties are shown in Figure 2. The study area in Bacon County was alongside GA highway 32 and few kilometers west of the city of Alma. There were plenty of blueberry orchards in the study area, which were demarcated with GPS instrument polygon mapping procedures. The study area in Brantley County was close to city of Nahunta. The study area in the county included a huge blueberry orchard known as ZBlue Berry Farm. The study area in Camden County was a small area close to the city of Woodbine. It included one large blueberry farm that was surrounded by pine forest. Most of the study areas were selected with an aim to include the weather stations established by Georgia Automated Environmental Monitoring Network (AEMN, www.Georgiaweather.net) in these three counties.



(a) (b) (c)

Figure 2: Study area locations in the three counties: Bacon (a), Brantley (b), and Camden (c).

The study boundary polygon files (Figure 2) were created using the heads-up digitization method in ArcMap with the NAIP images as the background. These polygons were used to extract the study area portion of geometrically and radiometrically corrected 2007 NAIP images of the three counties using the ArcGIS 10.2 software (ESRI TM, Redlands, CA) with the use of Extract by Mask tool. The vector data files (study boundary polygons) were spatially referenced to the NAD 83 UTM Zone 17N coordinate system.

Principal component analysis

SOM unsupervised image segmentation in IDRISI Andes can only be conducted on a single band image with 8-bit unsigned pixel types. The three visible bands, i.e., Red, Green, and Blue, in multi-spectral imagery contain digital information about land use that differ significantly from each other because each band differs with the reflectance pattern and the data from all of the spectral bands involve a certain degree of redundancy [42]. Earlier studies have shown that the digital information of the red band images only is insufficient for the correct representation of the vegetation Panda et al. [9]. Green and blue bands also contain digital signatures that could be helpful in distinguishing vegetation from other land-uses. It is, therefore, essential to use all three bands to extract pertinent information about land use and land cover types from remotely sensed images [43].

Principal component analysis (PCA) is an algorithm that can be applied to help reduce the 24-bit unsigned images (with red, green, and blue bands of the multi-spectral imagery) into an 8-bit unsigned raster without tempering much with the individual band digital records (Byne et al., 1980). PCA is a linear transformation that reorganizes the variance in a multi-band image into a new set of image bands [40]. Each individual band in the output PCA image receives some contribution from all of the input image bands. In our case this was three bands, i.e., the red, green, and blue bands). Therefore, PCA was used to solve the computational problems associated with multi-dimensional digital imagery data as stated above.

Table 1: Principal component analysis results of the National Agricultural Imagery Program multispectral images.

EIGEN MATRIX			
Layer	1	2	3
Bacon County study area :			
1	0.6228439546466815	-0.7819782559394068	0.02398781770124334
2	0.5252457303976895	0.3952399906424137	-0.7535929090005296
3	0.57981232382477	0.4819702864648322	0.6569035790022828
Layer	1	2	3
Brantley County study area:			
1	0.5968627740959972	-0.7878616797049319	-0.1517524383684495
2	0.5666823468494768	0.5478411093771789	-0.6154195614747404
3	0.5680017136132604	0.2813255987797017	0.773453269954698
Layer	1	2	3

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

Camden County study area :			
1	0.586197577598318	-0.8091288623666408	-0.04102297048341577
2	0.5647991250505207	0.4444386719078864	-0.6953245395123529
3	0.5808393481506141	0.3844278228790133	0.7175241463805762
EIGEN VALUES			
Principal Component Layer	1	2	3
Bacon County study area imagery:			
Values	5749.999269961269	45.35597126770778	9.791109067834571
Percentage of variation	99.0%	0.8%	0.2%
Brantley County study area imagery:			
Values	8196.324078405831	27.17866776551659	15.93133197981849
Percentage of variation	99.7%	0.2%	0.1%
Camden County study area imagery:			
Values	5749.999269961269	45.35597126770778	9.791109067834571
Percentage of variation	99.0%	0.8%	0.2%

The PCA was completed in ERDAS Imagine 2015 with the three study images to obtain principal component band rasters. Three principal components for each image were obtained to complement the three bands used in the analysis. The Eigen matrix and Eigen values for each principal component analysis were also determined. Most importantly, the PCA operation was conducted with the option of producing the 8-bit unsigned principal component band images as the output. The PCA data analysis (Table 1) showed that principal component 1 (PC1) for all three study images was the best representative for the R-, G-, and B-band variations as it represented equal to or more than 99% of the total variation. It was, therefore, postulated that further image processing analysis or SOM segmentation of these PC1 images could provide a better result than analyzing the individual R-, G-, and B-bands. The PC1 band images of all three county study areas were imported into IDRISI Andes using the software-specific format (ERDIDRIS) import option. The images had an RST format to make it compatible with the IDRISI Andes software.

Kohonen's SOM neural unsupervised classification of PC band images

General working procedure of SOM in IDRISI Andes

The SOM is a neural network procedure that was used in this study to segment the PC band images. It is closely modeled after the Kohonen [34] procedure that consists of two layers, such as input layer and an output layer that are typically organized as a two-dimensional (typically square) array of neurons (Figure 3). However, the output layer of SOM, which is also known as SOM or Kohonen layer, is commonly a one-dimensional structure These two layers are fully connected to each other's neurons and each input layer neuron has a feed forward connection to each neuron in the Kohonen layer (Figure 3). The synaptic weight vector W is associated with each connection from the input layer to a neural unit, i.e., Kohonen layer (Figure 4). Inputs to the Kohonen layer are calculated with the following equation

$$I_i = W_i \cdot X \ , \qquad (1)$$

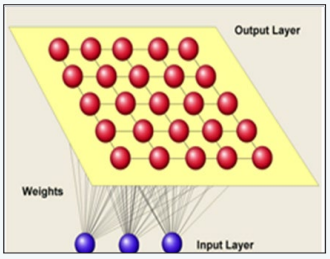
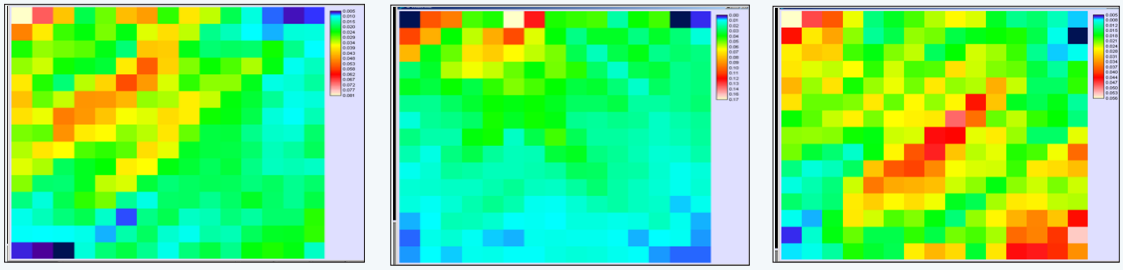


Figure 3: Example of the architecture of a Self-organizing Map with an input layer, consisting of three neurons, and an output layer, consisting of 5 x 5 neurons that are equally spaced (Source: IDRISI Andes software Help resources).



(a) (b) (c)

Figure 4: U-matrix or feature map color coding of three county study area images that shows the possible discernible classes for Bacon (a), Brantley (b), and Camden (c) counties.

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

where W is the weight vector, X is the input vectors, and WiX is a dot product. The SOM works with the approach of winner takes all. Thus, the output winning neuron becomes the neuron with the biggest Ii and the neuron whose weight vector has the minimum Euclidean distance, di, from the input vector X, i.e.,

$$d_i = \|W_i - X\| \quad (2)$$

In the case of the Euclidean distance calculation, weight, W, and input vector, X, are not normalized. In the use of SOM for supervised classification, the process begins with a coarse tuning phase that is effectively a form of unsupervised classification in which competitive learning and lateral interaction lead to a fundamental regional or topological organization of neuron weights that represent the underlying clusters and sub-clusters in the input data (IDRISI Andes Help, 2009). In this study we used the coarse tuning procedure for unsupervised segmentation. The fine tuning stage that was applied later refined intra-class decision boundaries using a Learning Vector Quantization (LVQ) procedure (IDRISI Andes Help, 2009).

In the SOM network, the neurons in the Kohonen layer compete with each other to be the winner when the input neuron is added to the layer. The winning neuron is trained every time with the use of learning rate as shown in equation 3.

$$w_{ji}^{t+1} = w_{ji}^{t} + \alpha^{t}(x_i^{t} - w_{ji}^{t}) \quad \forall\, d_{winner\ j} \in \gamma^2 \qquad (3)$$

$$w_{ji}^{t+1} = w_{ji}^{t} \quad \forall\, d_{winner\ j} \notin \gamma^2$$

where αt is the learning rate at iteration t and dwinner j is the distance between the winner and the other neuron in the output layer (IDRISI Andes Help, 2009). The winner neuron is calculated in the network with the following equation (IDRISI Andes Help, 2009).

$$Winner = \operatorname{argmin}\left[\sqrt{\sum_{j=1}^{t}(x_i^t - w_{ji}^t)^2}\right] \quad (4)$$

where xit is the input neuron i at iteration t, and wtji is the synaptic weight from input neuron I to output neuron j at iteration t (IDRISI Andes Help, 2009). The weights of the winner and its neighbors within a radius r are then altered according to a learning rate αt, which was 0-1 in our case, as shown in equation 3.

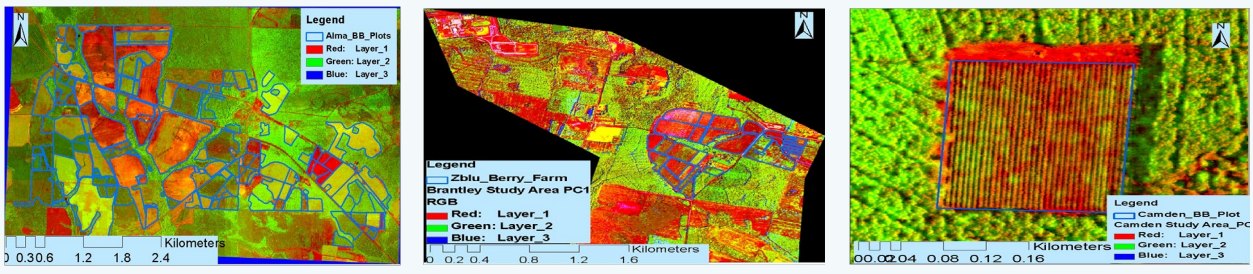
The learning rate used in the SOM network declines over time between its maximum and minimum values based on a time decay function as given in equation 5 in order to determine the correct winner (IDRISI Andes Help, 2009). An identical time-decay function was used to reduce the neighborhood radius r from an initial size that can encompass all the input neurons to the final stage when it encompasses only the winner (IDRISI Andes Help, 2009).

$$\alpha_t = \alpha_{\max}\left(\frac{\alpha_{\min}}{\alpha_{\max}}\right)^{\frac{t}{t_{\max}}} \quad (5)$$

The learning rate is updated from time to time and the neighborhood size is reduced during the course of the learning process (Kangas and Kohonen. 1996). The SOM then determines the output cluster (expected) values of the input vector. With a converged status, i.e., in case of the winner is decided, SOM can characterize the distribution of input samples, and thus generate a two-dimensional map from a multi-dimensional feature space showing the distribution of land-use classes in the input image (Tso and Mather, 2001). The feature map generated in this process is color coded by the information classes and all neurons of a particular class are coded with the same color. However, some of the classes during the initial stages of iteration stay unlabelled.

SOM Image Classification in IDRISI Andes

Kohonen's SOM hard classifier function of IDRISI Andes was used for the three PC1 band (8-bit unsigned) image classification to distinguish blueberry orchards from other land-uses. With an option to find unsuspected classes in the image and at the same time to reduce the human-induced error in the image segmentation process, we used the unsupervised classification option of the Kohonen's SOM method. As we did not have defined weights for the analysis, we employed the training network classification option, in which random weights were generated to train the model for successful winning neuron determination. To fix a neighborhood radius in the classification training process sampling band images of 3 x 3 (means a group of 3 by 3 pixels) was chosen. As mentioned earlier, the learning rate used in this analysis had a minimum value of 0 to a maximum value of 1. Finally, the maximum number of output clusters selected for each study area image was set according to the quality of the images, i.e, the homogeneity or heterogeneity.



(a) (b) (c)

Figure 5: Principal component image of the three study areas in Bacon (a), Brantley (b), and Camden County (c) along with the GPS recorded blueberry farms at different stages.

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

For the images of Bacon and Brantley County, 15 output clusters were selected because the study areas were comparatively larger and contained a lot of heterogeneity, which was studied with visual analysis. For image from Camden County, only 10 output clusters were chosen as the study area was very small and contained only a few blueberry orchards that were in various stages of development and different densities of the f forest covers that included a few bare land patches. The SOM classification model was coarse tuned to provide U-matrix feature maps. The feature maps showing the possible land-use classes are presented in Figure 5. From the feature map color coding analysis, it was confirmed that there were a few unsupervised classes in the image that had a very low coverage, i.e., they covered only less than 1% of the entire study area.. Therefore, image reclassification procedures were applied to merge some classes to their nearest class with close to the same digital mean values of the classes with low coverage in the study area.

As part of the SOM training model process, the root mean square error or quantization error of training was obtained as output. This error provided the efficiency level of the SOM classification method. The Kohonen's SOM unsupervised classification training quantization error for Bacon, Brantley, and Camden County images were 0.0107, 0.0033, and 0.0065, respectively. These errors were significantly low (zero error is the lowest possible error) showing that the SOM classification model trained well to provide a better classification of the images. It should be noted that if the quantization errors would have been larger, the model parameters should be changed until a low quantization error is obtained.

Once the quantization error was found suitable, the classification procedure was performed on the images using the respective training models. The classification results were saved into the specified output folder as RST format. The three SOM classified images were then exported as TIFF format for further analysis in the ArcGIS 10.2 software for image reclassification and subsequent classification accuracy assessment.

Image reclassification

Overall we obtained 15, 15, and 10 unsupervised classes for the study areas in Bacon, Brantley, and Camden counties, respectively. There was, therefore, a need to reclassify the images to reduce the number of land-use classes in each image by merging classes. Based on ground truthing, we found that there were only four to five types of land-use classes present in the study area. These included Mature blueberry bushes,

1. a. Intermediate size blueberry bushes,
2. b. Young blueberry bushes,
3. c. Forest
4. d. Bare land or a field prepared for planting. In one image there was also a school in an urban area. Therefore, using the mean digital values of each class, classes were merged so that our intended blueberry orchards were distinguished well from the classified images.

Accuracy assessment of the classified image

The accuracy assessment was performed based on ground truthing of the sites. Random pixels from the classified images were selected and compared to the corresponding land use classes on the ground to determine the classification accuracy of the unsupervised classification technique to verify the correct identification of the blueberry orchards. As mentioned earlier, the orchards were identified as three different categories based on their age. A Trimble GeoXH handheld GPS unit was used to collect the ground truth points for accuracy assessment of the classification of the image. The error matrix was used to find the accuracy of our image segmentation, which is the most common way of representing accuracy. The error matrix is a very effective way to represent accuracy as it includes both the error of inclusion, e.g., commission error, and the error of exclusion, e.g., omission error [44].

Four different parameters were used to evaluate the performance of our classification process. These included the producer's accuracy, the user's accuracy, the overall accuracy, and the Kappa statistics (See [31] for the detailed algorithms). Instead of selecting stratified random samples from the classified images that covered the entire study area, we concentrated only on the blueberry orchards in different locations of the study areas. Most of the orchards were delineated using the polygon mapping option of the Trimble Geo XH instrument. Ground truth points were recorded with the GPS instrument inside these orchards. Fifty ground truth points were recorded for the Bacon and Brantley County study area and 25 points were recorded for Camden County (woodbine) study area. The recorded GPS ground truth points were differentially corrected and were exported as a point shape file. The ground truth sample points that were collected in the orchards during the visit on August 21, 2008 had a very high average accuracy of 0 to 15 cm. The accuracy assessment was conducted with the ERDAS Imagine 2015 accuracy assessment function. The attribute table of the differentially corrected sample point shapefile was exported as a TXT file in ArcGIS 10.2 and was then imported into ERDAS Imagine to determine the accuracy of our Kohonen's SOM neural network unsupervised classification technique.

Results and Discussion

The three county study area visible band multi-spectral images were combined into principal component images as mentioned in the materials and method section. The PC1 band images for all three sites had more than or equal to 99% variation. Therefore, it was the correct representation of 24-bit data of all three (Red, Green, and Blue) bands spectral information and it was classified to distinguish different land-use, including blueberry shrubs, in the image. Figure 5 shows the PC1 band images of the three study areas and from which

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

the blueberry orchards can visually be easily distinguished. The orchard polygons collected from the field using the Trimble GeoXH and processed in the lab using GPS Pathfinder software was then overlayed on the images. The polygons of the blueberry fields were at different stages, i.e., already matured plants, young or intermediate plants, and fields that were bare and being prepared for planting of your blueberry bushes. Kohonen's SOM neural network unsupervised clustering was applied to the PC1 band images for image segmentation. Each image was segmented to a random high number of land use classes as discussed earlier. Our field physical verification or ground truthing allowed us to understand the classes of the Bacon, Brantley, and Camden SOM classified images. With that experience, the classified images were reclassified to provide a fewer numbers of classes in the field. All three county study images finally contained four differentiable classes as discussed in the materials and method section. However, it was very difficult to distinguish the mature blueberry plants from forest cover, the medium stage blueberry plants from scattered trees or shrubs, and the young blueberry plants from/ grassland as the area had a lot of tall grasses based on the unsupervised classes obtained from SOM classification. However, one important aspect of the remotely sensed imagery helped us to clearly demarcate the blueberry orchards from the other land uses which had very similar reflectance parameters. That textural property of the images helped in distinguishing the blueberry shrubs in the image because of its row-crop characteristics. The blueberry orchards were clearly discerned in the classified image due to the row textural pattern of a row of trees followed by a row of bare soil and a row of blueberry plants (Figure 5c & 8). Thus, the blueberry orchard delineation was a successful with a combination of spectral and textural analysis of the images using the advanced geospatial techniques. Figures 6-8 represent the SOM classified images for the study area images of Bacon, Brantley, and Camden County, respectively.

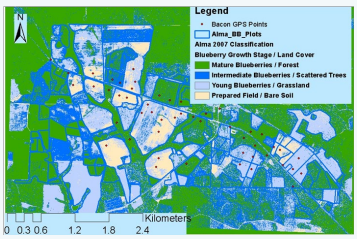


Figure 6: Self-organizing Map classified image of Bacon County study area along with ground truth points (50 points) and GPS recorded blueberry orchards (area) at different stages.

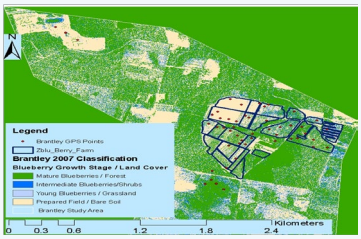


Figure 7: SOM classified image of Brantley County study area along with ground truth points (50 points) and GPS recorded blueberry orchards at different stages.

The ground truth sample points for all three study areas are shown in Figures 6-8. In Bacon County, most of the 50 ground truth points were collected alongside GA highway 32, where the blueberry orchards were in various stages of growth. Many fields were bare and were being prepared for planting. As shown in the Figure 6, the blueberry plots were also recorded with GPS instrument for verification of the classification accuracy. In Brantley County, 50 ground truth points were collected in three major locations. This included the Z Blue Blueberry Farm, a large blueberry orchard along the road to the Z Blue farm, and a house with lot of bare soil (Figure 7). The Z Blue farm has various types of land-uses including matured plants, bare soil, young plants, grass (pasture) land, and surrounding pine forest. The owner of Z Blue Blueberry Farm provided us with the farm layout in hard copy format with the attribute information for each individual field (Figure 7). This was digitized and overlayed on the SOM classified image to verify the classification accuracy. Twenty five ground truth points were recorded with GPS instrument in Z Blue blueberry orchard completely surrounded by pine forest (Figure 8). These ground truth points were taken in blueberry plant locations and the mid-row bare soil. The boundary of the orchard that was recorded with the GPS completely matched with the classified image blueberry orchard as shown in Figure 8.

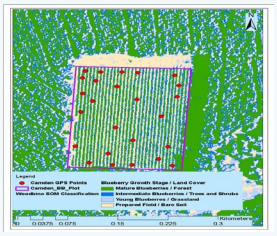


Figure 8: SOM classified image of Camden County study area along with ground truth points (25 points) and GPS recorded blueberry orchards at different stages.

Table 2: The classification accuracy report of Principal Component 1 clustered image (NAIP 2007) of all three study areas.

Study Area	Producer's Accuracy %	User's Accuracy %	Overall Accuracy %	Kappa Statistics
Bacon County	77	92	84	0.44
Brantley County	82	94	88	0.53
Camden County	92	94	95	0.64

Note: 100% accuracy was obtained when blueberry orchards were evaluated with the SOM classified images.

Accuracy assessments were conducted for the three final classified images. The accuracy assessment reports, including the user's, producer's, and overall accuracies and kappa statistics are shown in Table 2. The SOM classified image for Bacon County had a 77% producer's accuracy, a 92% user's accuracy, and an

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

overall accuracy of 84%. The Kappa statistics for the classified image was 0.44 (Table 2). This was an acceptable level of accuracy when compared to the similar reflectance properties of different stage blueberry plants to other land-uses that were abundant in the surrounding. The SOM classified image for Brantley County had a producer's , a user's accuracy, and an overall accuracy of 82%, 94%, and 88%, respectively (Table 2). The obtained Kappa statistics for the classified image was 0.53 (Table 2). The accuracy for this image was well within the range of our expectation as most of the ground truth points were collected from an established blueberry orchard which was perfectly distinguished from the 1-m resolution NAIP imagery. The accuracy assessment of the classified image for the Camden County study area was excellent. Only two points out of 25 ground truth points were misclassified due to an error encountered during the GPS data collection process. The two points were taken on the side of the ridge (blueberry row) that contained blueberry bushes but it was recorded as a bare soil row area. Even a differential correction of the GPS data could not correct it. The producer's accuracy for the classified image was 92% and the user's accuracy was 94%, while the overall accuracy was 95% and the Kappa statistic was 0.64.

Above all, we analyzed the blueberry orchard polygons according as their different growth stages. Because the canopy of the blueberry bushes did not completely cover the entire orchard, we obtained mixed land-use information from all orchards except from the ones that had bare soil in preparation for planting. However, we obtained almost 100 percent perfection in distinguishing the blueberry orchards (not individual bushes) from the SOM neural network classified images due to its textural (row pattern) property of the images (Figures 6-8). Many studies have been able to obtain a very high accuracy for the classification of fruit orchards from other land uses [20,21,43,45]. However, most of these studies were based on large scale orchards that were larger than 100 km2 and used vegetation indices to distinguish these orchards from other land-uses. Similar to our study, Torres et al. [45] found that individual trees are distinguishable from remotely sensed images with their study of a small 2 ha olive orchard in Cordoba, Spain, in 2004 and 2005. They used green, NIR (near-infrared), panchromatic bands and normalized difference vegetation index (NDVI) and ratio vegetation index (RVI) to determine the relation between tree area and olive yield. The correlation coefficients ranged from 0.62 to 0.82 and 0.52 to 0.74 for 2004 and 2005, respectively. Our study results were better compared to their study based on our evaluation statistics. This study also supported the findings of Panda et al. [9] that remotely sensed images and high end image processing technique have the potential to distinguish individual blueberry plants from mixed vegetation.

Summary and Conclusions

The results of our study indicate that high resolution remote sensing imagery along with advanced/high-end geospatial techniques can be used to discern blueberry orchards from mixed vegetation. This study also demonstrated that NAIP images that are obtained during the growing season can be used in concert with geospatial techniques and has the potential to distinguish blueberry orchards from other land-uses, especially mixed forest vegetation. An interesting conclusion was drawn from this study that blueberry plants at different stage of growth have similar spectral reflectance with different vegetation types as discussed in the materials and methods section. The mature plantation has the spectral reflectance closer to forest land cover. The intermediate growth plants have similar spectral reflectance as of other shrubs or sparse vegetation. Therefore, advanced image processing application is essential to distinguish blueberry plants from these land-uses with similar spectral characteristics. The PCA and SOM neural image unsupervised classification combined image segmentation procedure was found to be one appropriate analysis techniques to distinguish blueberry orchards. Based on the results from this study it can be concluded that for a well-maintained farm, mature blueberry bushes can be identified with a very high degree of accuracy because of its unique textural (pattern) characteristics. It is recommended that SSCM can be performed for large scale blueberry orchards with the use of remotely-sensed images and other high-end geospatial technologies [46-51].

Acknowledgement

This work was supported in part by the Georgia Space Grant Program, managed by the Georgia Institute of Technology on behalf of the National Aeronautics and Space Administration, and by State and Federal Funds allocated to Georgia Agricultural Experiment Stations Hatch project GEO1654.

References

1. Panda SS, Panigrahi S, Ames DP (2010a) Crop yield forecasting from remotely sensed aerial images with self-organizing maps. Transactions of ASABE 53(2): 323-338.
2. Casanova D, Epema GF, Goudriaan J (1998) Monitoring rice reflectance at field level for estimating biomass and LAI. Field Crop Research 55(1-2): 83-92.
3. Panda SS (2003) Data mining for precision management of crop. PhD dissertation, North Dakota State University, Fargo, North Dakota, USA.
4. Magri A (2005) Soil test, aerial image and yield data as inputs for site-specific fertility and hybrid management under maize. Precision Agriculture 6(1): 87-110.
5. Gonzalez B, Delia A, Chen PY, Lopez TM, Srinivasan R (2002) Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. Crop Science 42(6): 1943-1949.
6. Lobell DB, Ortiz-Monasterio JI, Asner GP, Naylor RL, Falcon WP (2005) Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agronomy Journal 97: 241-249.
7. Li T, Feng Y, Li X (2009) Predicting crop growth under different cropping and fertilizer management practices. Agricultural and Forest Meteorology 149(6-7): 985-998.
8. Noori O, Panda SS (2016) Site-specific Management of Common Olive: Remote Sensing, Geospatial, and Advanced Image Processing Applications. Computers and Electronics in Agriculture 127: 680 - 689.

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

9. Panda SS, Hoogenboom G, Paz J (2009) Distinguishing blueberry bushes from mixed vegetation land-use using high resolution satellite imagery and geospatial techniques. Computers and Electronics in Agriculture 67 (1-2): 51-59.

10. Panda SS, Hoogenboom G, Paz J (2010b) Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: A review. Remote Sensing 2(8): 1973-1997.

11. Sevier BJ, Lee WS (2005) Precision farming adoption by Florida citrus. Producers: Probit model analysis. University of Florida Cooperative Extension Service, circular 1461, p. 1-6.

12. Holmgren P, Thuresson T (1998) Satellite remote sensing for forestry planning: a review. Scandinavian Journal of Forest Research 13(1): 90-110.

13. Franco-Lopez H, Alan R Ek, Bauer ME (2001) Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbor method. Remote Sensing of Environment 77(3): 253-274.

14. Li D, Di K, Li D (2000) Land use classification of remote sensing image with GIS data based on spatial data mining techniques. International Archives of Photogrammetry and Remote Sensing 33(3): 238 – 245.

15. Haapanen R, Alan R Ek, Bauer ME, Finely AO (2003) Delineation of forest/nonforest land use classes using nearest neighbor method. Remote Sensing of Environment 89(3): 265 - 271.

16. Izzawati ED, Wallington ED, Woodhouse I (2006) Forest height retrieval from commercial X-band SAR products. IEEE Transctions on Geoscience and Remote Sensing 44(4): 863-870.

17. van Aardt JAN, Randolph HW, Scrivani JA (2008) Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. Photogrammetric Engineering & Remote Sensing 74(8): 1033-1044.

18. Rao PPN, Ravishankar HM, Raj U, Nagajothi K (2004) Production estimation of horticultural crops using IRS-1D LISS-III data. Journal of Indian Society of Remote Sensing 32(4): 393-398.

19. Sharma A, Panigrahy S (2007) Apple orchard characterization using remote sensing and GIS in Shimla district of Himachal Pradesh. Remote Sensing and Photogrammetry Annual Conference, p. 11-14.

20. Shrivastava RJ, Gebelein JL (2006) Land cover classification and economic assessment of citrus groves using remote sensing. ISPRS Journal of Photogrammetry & Remote Sensing 61(5): 341-353.

21. O'Connell MG, Goodwin I (2005) Spatial variation of tree cover in peach orchard. Acta Horticulture (ISHS) 694: 203-205.

22. Beek JV, Tits L, Coppin P, Somers B, Deckers T, et al. (2015) Improved Yield and Fruit Quality Estimation in Pear Orchards Using Remote Sensing Time Series. Acta horticulturae.

23. Aguilar MA, Vallario A, Aguilar FJ, Lorca AG, Parente C (2015) Object-Based greenhouse horticultural crop identification from multi-temporal satellite imagery: A case study in Almeria, Spain. Remote Sen 7(6): 7378-7401.

24. Peltoniemi JI, Kaasalainen S, Näränen J, Rautiainen M, Stenberg P, et al. (2005) BRDF measurement of understory vegetation in pine forests: dwarf shrubs, lichen, and moss. Remote Sensing of Environment 94(3): 343-354.

25. Rao P (2007) remote sensing for assessing vegetation dynamics and productivity of a peatland in southern Sweden. International Institute for Geo-Information Science and Earth Observation, Enschede, the Netherlands.

26. Senay GB, Ward AD, Lyon JG, Fausey NR, Nokes SE (1998) Manipulation of high spatial resolution aircraft remote sensing data for use in site specific farming. Transactions of the ASAE 41(2): 489-495.

27. Panda SS, Panigrahi S, Derby N (2001) Crop yield prediction modeling using soil adjusted vegetation index (SAVI). St Joseph MI, USA.

28. Eastman RJ (1999) Guide to GIS and Image Processing, (Vol. 1) Idrisi Production, Clarke University, Worcester, Massachusetts, USA.

29. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plennum Press, Newyork, USA.

30. Pandya AS, Macy RB (1996) Pattern Recognition with Neural Networks in C++. CRC Press, Boca Raton, FL, USA, pp. 432.

31. Zhang X, Li Y (1993) Self organizing map as a new method for clustering and data analysis. Proceedings of 1993 International Joint Conference on Neural Networks, Beijing, China.

32. Kartikeyan B, Sarkar A, Majumdar KL (1998) A segmentation approach to classification of remote sensing imagery. Int J of Remote Sensing 19(9): 1695-1709.

33. Stuckens J, Coppin PR, Bauer ME (2000) Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing Environment 71(3): 282-296.

34. Kohonen T (1995) Self-organizing Maps. Springer-Verlag , New York, USA.

35. Hoffmann R, Molander C, Bohn T (2008) NAIP 2008 summary briefing, Surdex Corporation. NAIP Coordination Meeting.

36. Budiyanto D, Sadananda R, Acharya S (1998) Land use mapping using self-organizing map. Environmental Software and Services.

37. Lindsey CS, Stromberg M (2000) Image classification using the frequencies of simple features. Pattern Recognition Letters 21(3): 265-268.

38. Panda SS, Panigrahi S (2009) Advancing precision agriculture by crop yield forecast with self-organizing map application on remotely sensed aerial images. Transactions of ASAE, USA.

39. Kanellopoulos I, Varfis A, Wilkinson GG, Megier J (1992) Land-cover discrimination in SPOT HRV imagery using an artificial neural network-a 20-class experiment. International Journal of Remote Sensing 13(5): 917 - 924.

40. Byne GF, Crapper PF, Mayo KK (1980) Monitoring landcover change by principal component analysis of multi-temporal Landsat data. Remote Sensing of Environment 10(3): 175-184.

41. http://165.221.201.14/NAIP.html

42. Fung T, LeDrew E (1987) Application of principal component analysis to change detection. Photogrammetric Engineering and Remote Sensing 53(12): 1649-1658.

43. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: A review. ACM Computing Surveys 31(3): 264 - 323.

44. Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37: 35-46.

45. Tremeau A, Borel N (1997) A region growing and merging algorithm to color segmentation. Pattern Recognition 30 (7): 1191-1203.

46. Haykin S (1999) Neural Networks a Comprehensive Foundation. (2nd edn), Prentice Hall Inc., Boston, USA, p. 34.

47. Deering DW (1978) Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. PhD Diss Texas A&M University, College Station, USA, pp. 338.

48. Dobermann A, Ping JL, Adamchuck VI, Simbahan GC, Ferguson RB (2003) Classification of crop yield variability in irrigated production fields. Agronomy Journal 95: 1105-1120.

49. Lin CT, Lee YC, Pu HC (2000) Satellite sensor image classification using cascaded architecture of neural fuzzy network. IEEE Transactions of Geoscience and Remote Sensing 38 (2): 1033-1043.

50. Panda SS (2008) Self Organizing Map (SOM) usage in LULC classification. Encyclopedia of Geographical Information Science. In:

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602

Sekhar S and Xiong H (Eds.), Encyclopedia of GIS. Springer, New York, USA, pp. 1036-1042.

51. Tso B, Mather PM (2001) Classification Methods for Remotely Sensed Data. Taylor and Francis, New York, USA, pp. 195.

How to cite this article: Sudhanshu S P, Gerrit H, Joel P. Blueberry Orchard Delineation with High-Resolution Imagery and Self-Organizing Map Neural Image Classification. Agri Res & Tech: Open Access J. 2016; 3(1): 555602. DOI: 10.19080/ARTOAJ.2016.03.555602