International Journal of Applied Environmental Sciences
ISSN 0973-6077 Volume 12, Number 10 (2017), pp. 1821-1830

http://www.ripublication.com

Application of Support Vector Machine (SVM) and Quick Unbiased Efficient Statistical Tree (QUEST) Algorithms on Mangrove and Agricultural Resource Mapping using LiDAR Data Sets

Carmel Jean G. Madanguit[a,*], John Paul L. Oñez, Herzon G. Tan, Milce D. Villanueva, Jesson E. Ordaneza, Remie M. Aurelio Jr., Annabelle U. Novero

Phil-LiDAR 2 Project, College of Science and Mathematics, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, Philippines

Abstract

Accurate mapping of mangrove and agricultural areas is necessary for effective planning and management of ecosystems and resources. While expert interpretation has been the typical method of classifying data sets, more efficient, objective, and faster methods of classification are required. This study applied the two classification techniques namely Support Vector Machine (SVM) and Quick Unbiased Efficient Statistical Tree (QUEST) algorithms for mapping mangrove and agricultural resources using LiDAR data. Ten LiDAR data sets were used for mangrove delineation. Each data set had a total of 90 ground-truth samples (30 per class) and 150 training points (50 per class) grouped into three classes: Mangroves, Other Vegetation and Non-Vegetation. Using Lastools software CHM, DSM, DTM, Intensity, Hillshade, Numret and Slope derivatives of the three LiDAR blocks were generated. eCognition software was used to perform classification of mangroves. A paired t-test was done to compare the accuracy of these two algorithms to determine which performed better in classifying mangroves. For agricultural resource mapping, LiDAR data sets for Tagum City and Panabo City were analysed. These areas contain large banana, coconut, and mango plantations. Statistical analyses showed that SVM performed better than QUEST in mangrove delineation. In agricultural resources mapping on the other hand, results showed that SVM and QUEST combined improved the general overall accuracy for Tagum and Panabo Cities to 97% and 96%, respectively. The agricultural land cover extracted could be used for a more accurate and effective resource management and monitoring of the cities' agricultural land. Both SVM and QUEST have a potential to improve the overall accuracy of LiDAR blocks in both mangrove and agricultural areas.

Keywords: SVM, QUEST, LiDAR, Mangrove mapping, Agricultural Resource mapping

1. BACKGROUND

The Philippines is rich in agricultural and mangrove resources especially in the areas of Mindanao, southern Philippines. Panabo and Tagum are known agro-industrial cities in Mindanao which cater to large plantations of banana, mango, and coconut. Mangroves are also important in coastal biodiversity since they are home to many species of fish, crustaceans, and the like. They can also reduce shoreline erosion and provide protection from possible tidal waves (Spalding et. al., 2014).

Accurate mapping of mangrove and agricultural areas is necessary in effective planning and management of ecosystems and resources. While expert interpretation has been the typical method of classifying data sets, more efficient, objective, and faster methods of classification are required (Stephens and Diesing, 2014).

Remote sensing images have been exploited as the latest information to study land cover and land uses. Acquiring remote sensing data has many advantages. It is unobtrusive and large volume of data or area can be collected in a short period of time (Suárez et. al., 2005). It can also acquire data from areas which are inaccessible or inconvenient to survey on site. Furthermore, these datasets can be easily calibrated and fed to the computer for further analysis.

The advent of high technology surveying method such as Light Imaging, Detection, and Ranging (LIDAR) allows for maps with high spatial resolution and accuracy. LiDAR data is commonly used for extracting land cover features. It has spatial and spectral information as it contains accurate 3D elevation data. It also provides intensity or reflectance values which are very helpful in extracting land cover (Antonarakis et.al., 2008). Unlike other satellite imageries, it is not susceptible to shadows and cloud cover since its data acquisition is independent from the sun (Gatziolis and Andersen, 2008).

There are many classification algorithms used for land cover classification applied on remote sensing data. Both non-parametric classifiers, SVM and QUEST can be used with arbitrary distribution and without statistical assumption. The SVM makes use of superior machine learning algorithms to find optimal boundaries for each classes inside the hyperplane (Huang et. al, 2002). On the other hand, QUEST is a binary decision tree primarily used for data classification and data mining (Loh and Shih, 1997) which was developed by Wei-Yin Loh and Yu-Shan Shih. Using decision trees has many advantages which includes faster computations, and data which are represented on differing measurement scales can be handled (Pal and Mather, 2003).

There is much current research in exploring more appropriate algorithms on different remotely sensed data for various types of land cover classification. Examples of these studies are Pal and Mather (2003), Li et al (2014), Song and Lu (2015), and Hasan et al. (2012). These studies utilizes various satellite imageries such as multispectral Landsat ETM and hyperspectral DAIS (Pal and Mather, 2003), and Multi-Beam Sonar (Hasan et al., 2012). However, applying and evaluating the accuracy of classification algorithms using LiDAR data set is rare.

This study applied two classification techniques namely Support Vector Machine (SVM) and Quick Unbiased Efficient Statistical Tree (QUEST) algorithms for mapping mangrove and agricultural resources using LiDAR data.

2. METHODS

2.1 Data sets used

LiDAR imagery acquired by Phil-LiDAR 1 was used for this study. Figure 1 shows the area coverage of the LiDAR data sets used. Shaded green area shows the LiDAR data sets used for agricultural resource mapping while the shaded blue area shows the LiDAR data sets used for mangrove resource mapping.

Validation points were taken from field work while training points were made from both image interpretation and field work. Ten LiDAR data sets were used for mangrove delineation across the provinces of Compostela Valley, Davao del Norte, Davao Oriental, South Cotabato, and Sarangani province. Each data set had a total of 90 ground-truth samples (30 per class) and 150 training points (50 per class) for mangrove resource mapping. It was grouped into three classes: Mangroves, Other Vegetation and Non-Vegetation.

LiDAR data sets covered in Tagum and Panabo Cities, Davao del Norte were used for agricultural resource mapping. Panabo City has around 100 training points and 40 validation points for non-ground features such as Banana, Coconut, Mango, and Non-agricultural trees while ground features such as grassland and fallow have 70 training points and 30 validation points. Tagum City has around 30 ground-truth samples and 70 training points of each classes except Mango and Non-agricultural trees since there is confusion between these features. 40 additional training points were added to lessen the confusion between these two features.

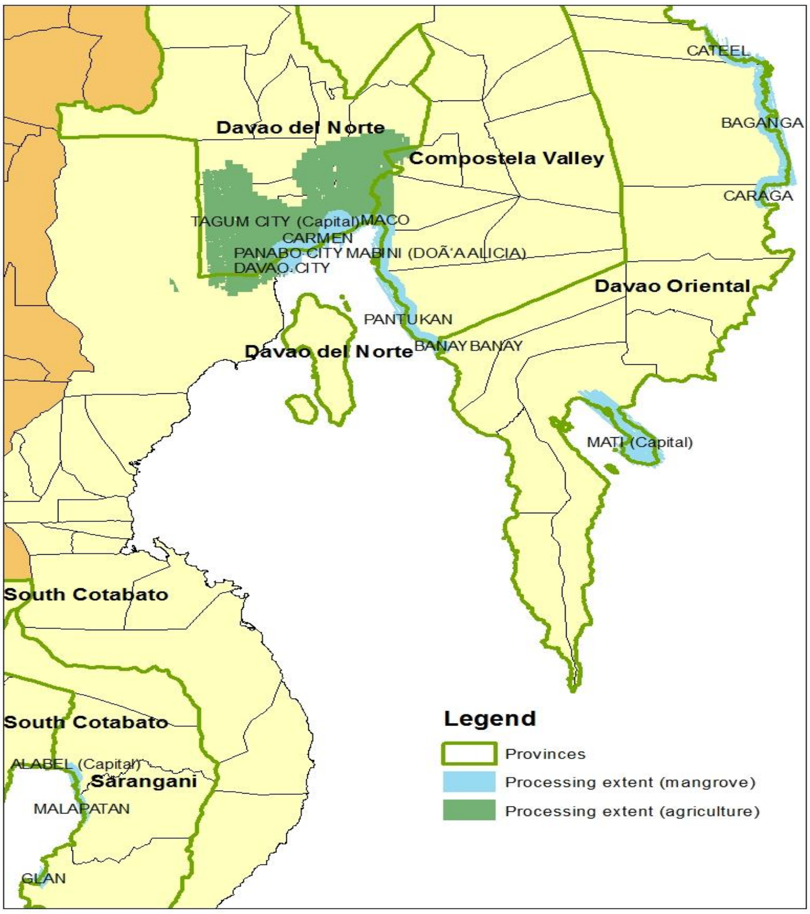

Figure 1: Processing extent of mangrove and agricultural resources mapping.

2.2. Algorithms used

The classification of remotely sensed images is an important phase in determining land cover and land use. There are two general categories of classification techniques- parametric and nonparametric. Parametric models assume some finite set of parameters and that the data for individual classes are distributed normally (Taati et al, 2014). Among the existing parametric techniques, Maximum likelihood classification (MLC) is the most established approach (Jensen, 2005). It creates decision surfaces based on the mean and covariance of each class. In contrast, non parametric techniques makes no assumption of the statistical nature of the data. Support Vector Machine (SVM) classification belongs to this new addition to the existing classification techniques. Decision tree methodologies like Quick, Unbiased, Efficient, Statistical Tree (QUEST) are also non-parametric and can efficiently deal with large, complicated datasets without imposing a complicated parametric structure (Song and Lu, 2015).

2.3 General Workflow

The general workflow of this study is shown in Figure 2. It followed the manual developed by Carranza et. al. (2014) and utilizes Object Based Image Analysis (OBIA) and various software applications to process LiDAR data sets for land cover classification.

First, the LiDAR point cloud were processed by deriving rasters through LAStools. LAStools is a software application developed for the processing of LiDAR data (Isenburg, 2016). A number of rasters were produced as prerequisite to the production of terrain models. Along with training and validation shapefiles, these rasters were then integrated to eCognition. The software eCognition is a software designed for object based image analysis that utilizes classification and feature extraction (Trimble, 2014).

Various rulesets and algorithms were then applied together with the rasters and shapefiles. Multi-threshold segmentation with nDSM layer was executed to distinguish ground and non-ground features. The threshold was set to 1 with Normalized Digital Surface Model (nDSM) or the Canopy Height Model (CHM) as the image layer used for segmentation. Both ground and non-ground objects produced from the previous segmentation were then further divided into finer objects through multiresolution segmentation.

Training points created were then loaded as training samples which were incorporated with the derivatives for classification. For SVM classification, an array was created to store selected statistical image layers that would best identify among the land covers. These statistical image layers were identified through SEaTH (Separability and Threshold) which automates feature extraction through a statistical analysis of training samples (Nussbaum and Menz, 2008) . The classes to be identified are then employed to the SVM classifier algorithm and were normalized. The C and rbf value used is 200 and 0 respectively. In QUEST classification, a univariate binary decision tree is generated where it shows the order of the statistical image layers to be used with its corresponding threshold value to produce land cover classification on LiDAR data. The split point method applied was an exhaustive search using Gini index.

Lastly, validation points are loaded as validation samples to compute an accuracy assessment which shows the overall and per class accuracy.

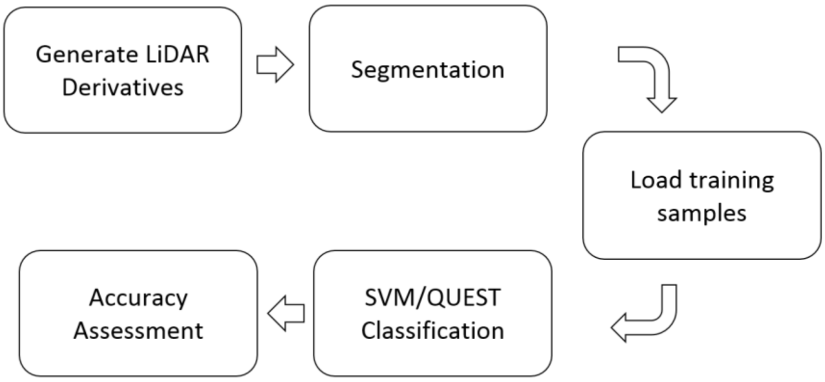

Figure 2: General work flow of the study

3. RESULTS

3.1 Mangrove resource mapping

Table 1 shows the overall accuracy of the LiDAR data sets used for mangrove resource mapping. Paired t-test was done to compare the accuracy of these two algorithms to determine which performed better in classifying mangroves. Shapiro-Wilk was done to test the normality of the result first before applying the paired t-test. Since the result of Shapiro-Wilk shows that it is normalized, one-tail paired t-test was then computed. This statistical test was used to determine which algorithm is better in mapping mangroves in terms of accuracy. The results show that SVM is better in delineating mangroves compared to QUEST.

Table 1: Results of classification using SVM or QUEST for Mangrove resource mapping

LiDAR Block	SVM Accuracy	QUEST Accuracy
79AB_North	92.88%	50.46%
79AB_South	91.00%	91.68%
84A	93.66%	77.68%
85B	90.42%	69.13%
90A	91.46%	82.09%
90E	93.27%	90.57%
Tagum A	94.00%	89.13%
Tagum As	95.40%	96.93%
Tagum As_add	93.26%	91.28%
Tagum B	89.80%	81.99%

Table 2: Results of paired t-test

	SVM	QUEST
Mean	92.52%	82.09%
Variance	3.14	189.55
Observations	10	10
Shapiro-Wilk	0.96	0.86
p-value	0.74	0.08
p(T<=t) one-tail	0.018	

3.2 Agricultural resource mapping

SVM and QUEST were both applied for agricultural resources mapping. Upon many classification iterations, it was noticed that SVM produced good individual accuracies for some classes and it is the same with QUEST. These rulesets were then combined to produce a higher overall accuracy. As shown in Table 3, combining these two algorithms for land cover classification has increased the overall accuracy as much as 2-6%.

Table 3: Results of classification using SVM or QUEST for Mangrove resource mapping

LiDAR Block	SVM Accuracy	QUEST Accuracy	SVM-QUEST Accuracy
Tagum City	94.91%	93.57%	97.00%
Panabo City	92.50%	90.00%	96.64%

These results were then integrated into one map which shows the resources classified with their area coverage as shown in Figure 3. These shows that LiDAR data sets can be used in generating high resolution and more precise maps for resource management and planning.

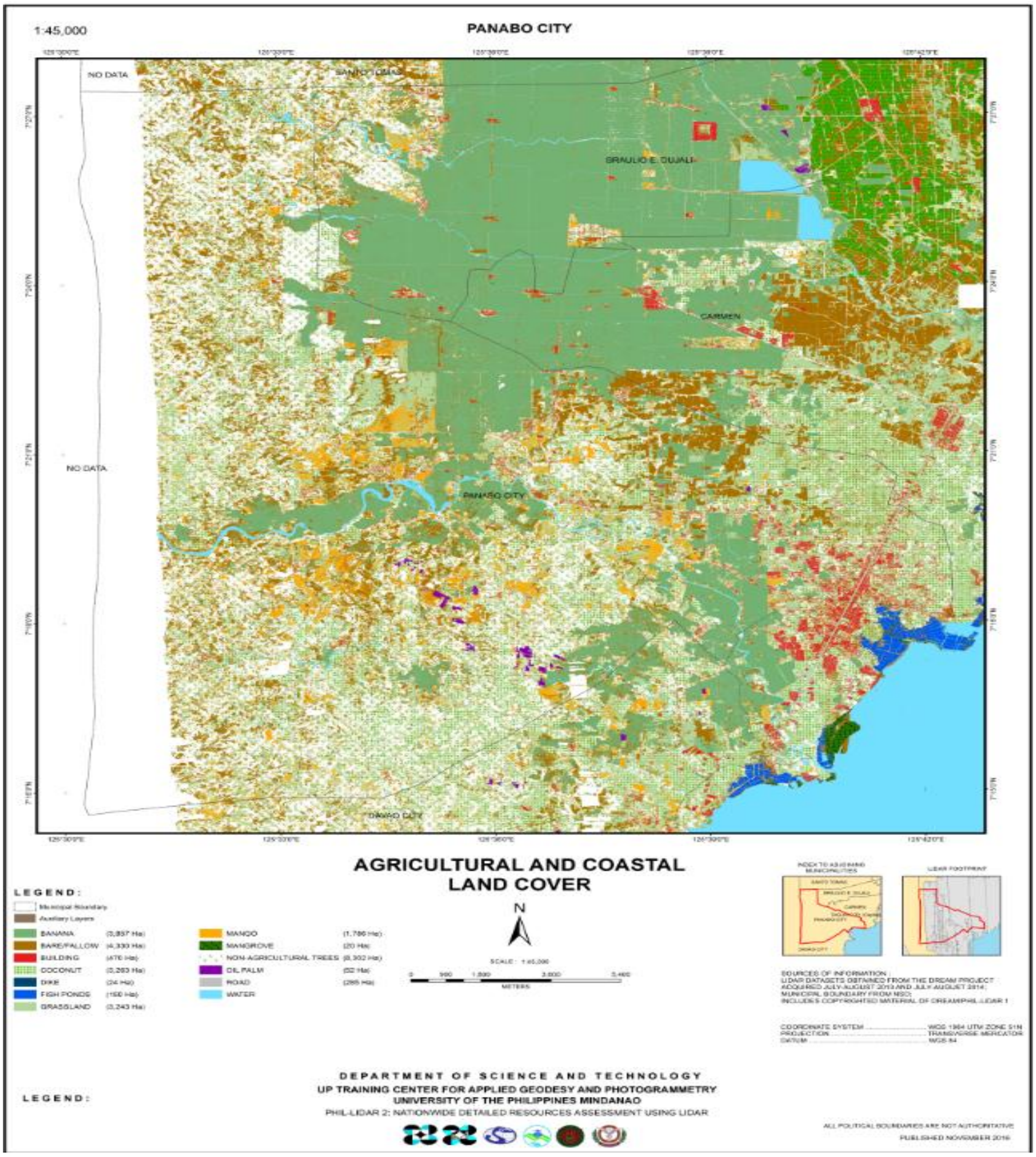

Figure 3: Mangrove and agricultural resources incorporated into one map.

3.3 Acknowledgments and Legal Responsibility

This study was conducted under the Phil-LiDAR 2 research program. It was funded by the Department of Science and Technology (DOST) and monitored by DOST-PCIEERD (Philippine Council for Industry, Energy and Emerging Technology Research and Development). We would also like to acknowledge the Training Center for Applied Geodesy and Photogrammetry (TCAGP) based from UP (University of the Philippines) Diliman for training the personnel and giving us the training manuals on how to perform Object-based Image Analysis (OBIA) applied on LiDAR data sets.

REFERENCES

[1] Antonarakis, A.S., Richards, K.S., & Brasington J. (2007). Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment*, *112*, 2988-2998. doi:10.1016/j.rse.2008.02.004

[2] Carranza, C., Rollan, T., Tañada, E., Guerrero, J., Jerez, M., & Blanco, A. (2014). Phil-LIDAR 2 OBIA Training Session Exercises Manual. Nationwide Detailed Resource Assessment using LIDAR (Phil-LIDAR 2) Program Project 1 Agricultural Resources Extraction from LIDAR Surveys (PARMAP).

[3] Gatziolis, D., Andersen H-E. (2008). *A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest* (Gen. Tech. Rep. PNW-GTR-768). Portland, OR: U.S: Department of Agriculture, Forest Service, Pacific Northwest Research Station.

[4] Hasan, R.C., Ierodiaconou, D., & Monk, J. (2012). Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing 2012*, *4(11)*, 3427-3443. doi:10.3390/rs4113427

[5] Huang C., Davis, L.S., & Townshend, J.R.G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725-749. doi: 10.1080/01431160110040323

[6] Isenburg, M. (2016). LAStools [computer software]. Germany: Gilching.

[7] Jensen, J. (2005). Introductory digital image processing: A remote sensing perspective (3rd ed.). Upper Saddle River, NJ: Prentice Hall. 526 pages.

[8] Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training samples sizes in urban land classification with Landsat Thematic Mapper imagery. *Remote Sensing*, *6*(2), 964-983. doi:10.3390/rs6020964

[9] Loh, W.Y. & Shih, Y.S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7, 815-840. Retrieved from http://www3.stat.sinica.edu.tw/statistica/j7n4/j7n41/j7n41.htm

[10] Nussbaum, S. & Menz, G. (2008). SEaTH - A new tool for feature analysis. In *Object- based Image Analysis and Treaty Verification* (51-62). Netherlands: Springer Netherlands

[11] Pal, M. & Mather, P. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, *86*(4), 554-565. doi: 10.1016/S0034-4257(03)00132-9

[12] Song, Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. http://doi.org/10.11919/j.issn.1002-0829.215044

[13] Spalding, M., McIvor, A., Tonneijck, FH., Tol, S. & van Eijk, P. (2014). Mangroves for coastal defence: Guidelines for coastal managers & policy makers. Netherlands: Wetlands International. Retrieved from http://www.nature.org/media/oceansandcoasts/mangroves-for-coastal-defence.pdf

[14] Stephens, D., & Diesing, M. (2014). A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PLoS ONE*, *9*(4). doi:10.1371/journal.pone.0093950

[15] Suárez, J. C., Smith, S., Bull, G., Malthus, T. J., Donoghue, D., & Knox, D. (2005). The use of remote sensing techniques in operational forestry. *Quarterly Journal of Forestry, 99*(1), 31-42. Retrieved from http://www.forestry.gov.uk/pdf/QJFarticle.pdf/$FILE/QJFarticle.pdf

[16] Taati, A., Sarmadian, F., Mousavi A., Pour, C., & Shahir A. (2014). Land use classification using support vector machine and maximum likelihood algorithms by Landsat 5 TM images. *Planet Sciences and Area Based Research*, *12*(8), 681-687. Retrieved from http://wjst.wu.ac.th/index.php/wjst/article/view/1225/515.

[17] Trimble. (2014). *eCognition Developer 9.0 User Guide*. Munchen, Germany.