Jurnal Teknologi

Full paper

Selection of Vegetation Indices in Hierarchical Structural Based Vegetation Species Classification

W. C. Chew,[a] A. M. S. Lau,[a*] T. Okuda[b]

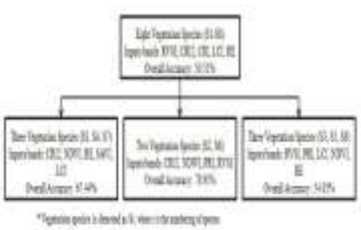
[a]Department of Geoinformation, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor Malaysia
[b]Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Higashi-Hiroshima, 739-8521, JAPAN

*Corresponding author: alvinlau@utm.my

Article history

Received : 31 July 2014
Received in revised form :
23 November 2014
Accepted : 1 December 2014

Graphical abstract



Abstract

A pilot study has been carried out to test the effectiveness and feasibility of a multi-level classification procedure in handling a large number of vegetation species and aims to improve the classification accuracy. In this study, the research question is whether a selected vegetation index has its significance towards a specific vegetation species group. Ten vegetation indices extracted from *in-situ* hyperspectral remote sensing data has been applied and one-way ANOVA analysis (with significance level, α = 0.01, 0.05, and 0.1) has been carried out to evaluate significant difference of vegetation indices in pair-wise vegetation species. In the multi-level classification procedure, vegetation species were classified continuously from one level to the next level until a good classification result has been achieved. Results indicated that multi-level classification procedure has effectiveness in handling vegetation species where accuracy has been improved from the first classification level to the second level. The study also has highlighted the significance of selected vegetation indices subsets toward different vegetation species in the multi-level classification procedure where the highest accuracy has been given by the selection in dedicated vegetation species groups.

Keywords: Vegetation indices; vegetation species classification; hyperspectral remote sensing

Abstrak

Satu kajian awal telah dijalankan untuk menguji keberkesanan dan kebolehan prosedur pengkelasan pelbagai peringkat dalam mengendalikan spesies tumbuh-tumbuhan secara jumlah besar dan bertujuan untuk meningkatkan ketepatan pengkelasan. Dalam kajian ini, persoalan kajian ialah sama ada suatu indeks tumbuhan yang dipilih mempunyai kepentingannya terhadap kumpulan spesies tumbuh-tumbuhan tertentu. Sepuluh indeks tumbuhan dijana daripada data lapangan penderiaan jauh hiperspektral telah digunakan dan analisis ANOVA sehala (dengan aras keertian, α = 0.01, 0.05, dan 0.1) telah dijalankan untuk menilai perbezaan ketara indeks tumbuhan di antara spesies tumbuh-tumbuhan dalam bentuk pasangan. Dalam prosedur pengkelasan pelbagai peringkat, spesies tumbuh-tumbuhan telah dikelaskan secara berterusan dari satu peringkat ke peringkat seterusnya sehingga hasil pengkelasan yang baik telah dicapai. Hasil kajian menunjukkan pelbagai peringkat prosedur pengkelasan mempunyai keberkesanan dalam mengendalikan spesis tumbuh-tumbuhan di mana ketepatan telah meningkat daripada peringkat pengkelasan pertama ke pengkelasan kedua. Kajian ini juga menekankan kepentingan subset indeks tumbuhan yang dipilih terhadap spesis tumbuh-tumbuhan yang berbeza dalam prosedur pengkelasan pelbagai peringkat di mana ketepatan tertinggi telah diberikan oleh pilihan atas kumpulan spesies tumbuh-tumbuhan yang berdedikasi.

Kata kunci: Vegetation indices; vegetation species classification; hyperspectral remote sensing

■1.0 INTRODUCTION

Vegetation species distribution mapping is important in ecosystem studies or natural resources management propose and detailed species maps in good accuracy are highly on demand for different vegetated areas. Remote sensing is a technique that uses remote sensors to capture electromagnetic radiation in certain range of wavelengths which are reflected from specific target area. The process of vegetation species mapping using remote sensing data is done based on analyzing reflected electromagnetic radiation (so called reflectance) in different wavelengths over a forest. Hyperspectral remote sensing captures reflectance in hundreds or thousands wavelengths. Massive spectral information could be extracted from hyperspectral remote sensing data by experts to conduct successful vegetation species mapping works. Thus, hyperspectral remote sensing data has been widely used in vegetation species classification and mapping.

In previous hyperspectral remote sensing studies for vegetation species mapping, experts applied statistical analysis to select a number of best wavelengths among hundreds hyperspectral wavelengths to classify vegetation species within an area of interest. However, some studies have highlighted that species classification accuracies have been compromised with similar set of selected best wavelengths in classifying a high number of target species.[1-2] It is found that classification accuracy was reducing gradually from a lower species number to a higher number as the possibility of misclassification among species has been increased. However, the reality is high number of vegetation species are growing in tropical rainforests while previous studies have met some limitations in handling species mapping for these regions.

Some recent studies have shown that selection of bands (combination of different wavelengths in hyperspectral data or different types of vegetation indices) are useful in identifying vegetation species meanwhile the input bands are species sensitive most of the time during the classification process.[3] Also, an adaptive selection in hyperspectral bands subset has been introduced to improve spectral separability in Hawaiian rainforests tree species classification.[4] In their study, an approach with different bands subset for specific species groups from a time series of hyperspectral data has given a better classification result than using an identical band subset for all tree species at one time.

An idea is introduced in this study to improve tree species classification accuracy where different subsets of hyperspectral data are used to classify different vegetation species in a multi-level classification procedure as shown in Fig. 1. With this classification procedure, a number of vegetation species is classified by using a set of selected best bands from hyperspectral remote sensing data. The vegetation species classification procedure will be continued for the same species at the next level but it is done on different parallel smaller groups with corresponding selected best bands. The process terminates at the n-level which optimal species classification is achieved. In this idea, all selected bands are species sensitive where each vegetation species group undergoes classification process with a dedicated set of selected bands. Starting at the second level, vegetation species are grouped into smaller groups as confusing species during classification due to high spectral similarity among each other could be classified to further improve classification accuracy. An adaptive bands subset tends to optimize the spectral separability among vegetation species which may lead to good classification accuracy eventually and effectively uses hyperspectral remote sensing data as more bands are selected from hundreds or thousands wavelengths compared to previous studies. The main research question in this study is whether a selected hyperspectral band subset has its significance towards a particular vegetation species group and aims to improve species classification accuracy with a multi-level procedure.

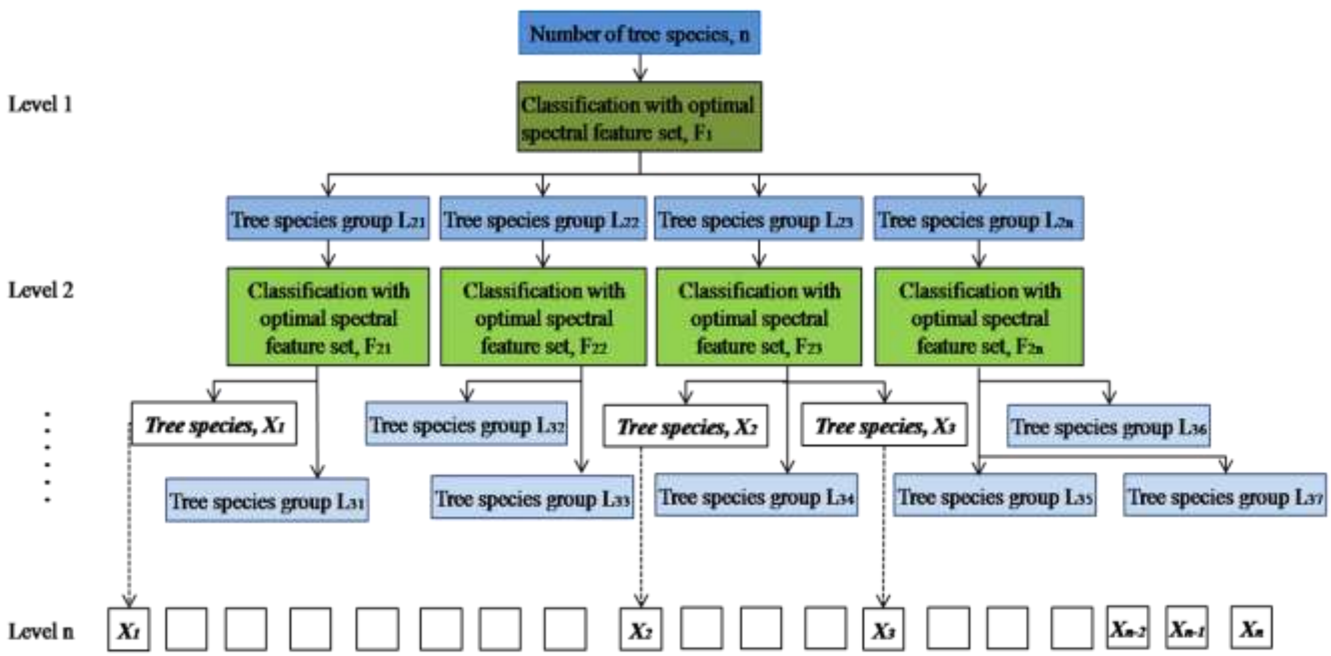


Figure 1 The structure of a multi-level vegetation species classification procedure

■2.0 DATA COLLECTION AND PROCESSING

This was a pilot study for testing the feasibility of a multi-level vegetation species classification procedure in handling a high number of target species. In this study, *in-situ* hyperspectral data of eight tropical rainforest vegetation species within the Pasoh Forest Reserve, Negeri Sembilan, Malaysia were collected with a portable spectroradiometer where 15 to 20 samples have been collected for each vegetation species as stated in Table 1. Vegetation indices are good indicators to delineate characteristics of different vegetation in remote sensing vegetation analysis purposes. Vegetation indices are mathematical formulae which use reflectance values of at least two wavelengths in remote sensing data. Ten vegetation indices (refer to Table 2) which were widely used in previous vegetation species classification studies [5-6] were applied in this study. Values of the ten selected vegetation indices have been extracted from different wavelengths of the collected hyperspectral remote sensing data for each sample in order to conduct classification purpose. For running classification, about a half of sample size was selected randomly from total samples of each vegetation species meanwhile the remaining samples were used in accuracy assessment. Also, the selected training samples were used in one-way ANOVA statistical analysis for vegetation indices selection.

Table 1 The list of vegetation species in this study

Label	Species Name	Number of Sample
S1	Dipterocarpus sublamellatus	20
S2	Pterygota alata	20
S3	Shorea roxburgii	20
S4	Aquilaria malaccensis	15
S5	Shorea acuminata	17
S6	Dyera costulata	20
S7	Artocarpus rigidus	15
S8	Shorea macroptera	19

Selection of good vegetation indices to classify different vegetation species was important in this study where those selected indices were used as input parameters into Maximum Likelihood classifier for identifying among tree species in a specific group. One-way ANOVA analysis (with significance level, 0.01, 0.05, and 0.1 respectively) has been carried out on each vegetation index to evaluate significant difference in the training samples for pair-wise vegetation species. Weighting was used to emphasize the important of vegetation index which has gained a lot of significant difference at higher significance level for pair-wise vegetation species. The assigned weights were three marks for significance level, α = 0.01, two marks for α = 0.05 and one mark for α = 0.1 respectively.

Table 2 The list of vegetation indices which was used in this study

Vegetation Index	Formula
Normalized Difference Vegetation Index (NDVI)	(R*798* - R*679*) / (R*798* + R*679*)
Red Edge NDVI (RE)	(R*750* - R*705*) / (R*750* + R*705*)
Soil-Adjusted Vegetation Index (SAVI)	1.5 * (R*798* - R*679*) / (R*798* + R*679* + 0.5)
Anthocyanin Reflectance Index 2 (ARI 2)	R*798* * ((1 / R*550*) - (1 / R*699*))
Carotenoid Reflectance Index (CRI)	R*800* * ((1 / R*520*) - (1 / R*550*))
Carotenoid Reflectance Index 2 (CRI 2)	(1 / R*511*) - (1 / R*699*))
Leaf Chlorophyll Index (LCI)	(R*850* - R*710*) / (R*850* + R*680*)
Normalized Difference Water Index (NDWI)	(R*862* - R*1239*) / (R*862* + R*1239*)
Photochemical Reflectance Index (PRI)	(R*532* - R*568*) / (R*532* + R*568*)
Red-edge Vegetation Stress Index (RVSI)	(R*719* - R*752*) / (2 - R*733*)

Remark: R*i* is the reflectance value of wavelength *i* nanometer in hyperspectral data

Starting from the first level of classification procedure, the top five highest score vegetation indices (refer to Table 3) were selected and put into Maximum Likelihood to run classification for the eight selected species. Accuracy assessment was done on the classification result to give misclassification errors among vegetation species. Based on user defined conditions, the eight species were then re-assigned into smaller species groups for further classification at the next level in order to improve the accuracy. Total score of each vegetation index in one-way ANOVA was calculated for each group respectively for the next input parameters selection. In this case, the total score was lower than the previous level due to the number of pair-wise species in smaller groups was lesser. Another top five vegetation indices were selected specifically for each group at the second classification level and the procedure was continued with running classification and accuracy assessment. In the multi-level classification procedure, the steps were repeated for a few levels until a good accuracy for vegetation species classification was achieved

.

Table 3 The result of one-way ANOVA among all eight vegetation species at the first level of species classification

Vegetation Index	α = 0.01	α = 0.05	α = 0.1	Total Score
RVSI	12	4	4	**20**
CRI2	3	12	4	**19**
CRI	7	5	5	**17**
LCI	9	3	3	**15**
RE	8	5	1	**14**
NDVI	2	4	4	10
SAVI	2	4	4	10
PRI	6	2	1	9
ARI2	4	3	0	7
NDWI	0	1	2	3

■3.0 RESULTS AND DISCUSSION

In this study, two classification levels were required to classify the eight vegetation species since the number of species is small. More classification levels are required for any larger number of target vegetation species and the structure of the multi-level classification procedure will be more complicated in order to complete a large number vegetation species mapping. From this pilot study, multi-level classification procedure has shown its effectiveness and feasibility in handling vegetation species mapping by using different sets of species sensitive vegetation indices in running Maximum Likelihood classification. As shown in Fig. 2, the classification overall accuracy has been improved from 50.51% to at least 54.05% and the highest accuracy was 78.95% at the second classification level. At the second level, the lowest accuracy was 54.05% where the species group contained Shorea roxburgii (S3), Shorea acuminata (S5) and Shorea macroptera (S8). This result has clearly indicated that vegetation indices with Maximum Likelihood classifier made Shorea species to be separated from other species effectively but did not perform so well to classify among the Shorea species. More hyperspectral remote sensing information should be included during the classification to improve the classification accuracy among Shorea species.

Cross-validation has been performed where vegetation indices subsets have been used to run classification in other species groups to answer whether a selected hyperspectral band subset has its significance towards a particular vegetation species group. Result in Table 4 has proven that the significance of dedicated vegetation indices in specific species group classification. The selection of vegetation indices sets have given the highest accuracy in the first (S1, S4, S7) and the second (S2, S6) species groups where the accuracies were 67.44% and 78.95% respectively. Although these two vegetation indices subsets have performed better than the dedicated selection of vegetation indices in the third group (S3, S5, S8), the third selection subset still gave the highest accuracy in the third group which was 54.05% when compared to another two groups.

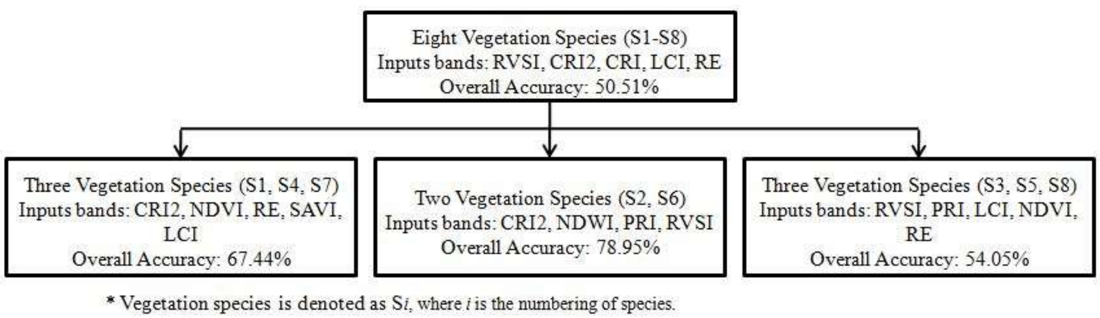


Figure 2 The overall accuracy of the eight vegetation species in a multi-level classification procedure.

Table 4 Cross-validation classification accuracy (in percentage) among three different vegetation species groups at the second level in classification procedure

Input Vegetation Indices (bands)	S1, S4, S7	S2, S6	S3, S5, S8
CRI2, NDVI, RE, SAVI, LCI	**67.44**	73.68	64.86
CRI2, NDWI, PRI, RVSI	62.79	**78.95**	78.38
RVSI, PRI, LCI, NDVI, RE	39.53	36.84	**54.05**

■4.0 CONCLUSION

Previous studies have highlighted the limitation of a single set of hyperspectral remote sensing wavelengths in handling a larger number of target vegetation species where classification accuracy is reduced significantly. Multi-level classification procedure which was introduced in this study has shown its potentials and effectiveness in classifying a number of vegetation species where overall classification accuracy was improved from the first level to the second level in the classification procedure. Thus, this idea should be enhanced and applies in future works.

Acknowledgement

A heartfelt appreciation is given by authors to Universiti Teknologi Malaysia and all kinds of assistances in conducting the field campaign. The current study was granted by the Ministry of Higher Education (MOHE) under the research scheme "UTM RG Flagship" with the VOT number Q.J130000.2427.02G19.

References

[1] K.L. Castro-Esau, G.A. Sánchez-Azofeifa, B. Rivard, S. J. Wright, M. Quesada. 2006. *Am. J. Bot.* 93: 517.
[2] J.B. Féret, G.P. Asner. 2013. *IEEE T. Geosci. Remote.* 51: 73.
[3] L. Wang, W.P. Sousa. 2009. *Int. J. Remote Sens.* 30: 1267.
[4] A. Somers, G.P. Asner. 2014. *Int. J. Appl. Earth Obs. Geoinf.* 31:57.
[5] M.L. Clark, D. A. Roberts. 2012. *Remote Sens.* 4: 1820.
[6] R. Pu. 2009. *Int. J. Remote Sens.* 30:2759.