

RESEARCH ARTICLE

# Classification of Tree Species in Overstorey Canopy of Subtropical Forest Using QuickBird Images

Chinsu Lin<sup>1\*</sup>, Sorin C. Popescu<sup>2</sup>, Gavin Thomson<sup>3</sup>, Khongor Tsogt<sup>1</sup>, Chein-I Chang<sup>4</sup>

**1** Department of Forestry and Natural Resources, National Chiayi University, 300 University Road, Chiayi 60004, Taiwan, **2** Department of Ecosystem Science and Management, Texas A&M University, College Station, TX, 77843, United States of America, **3** Department of Applied Foreign Languages, National Formosa University, 64, Wenhua Road, Huwei Township, Yunlin (63201), Taiwan, **4** Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD, 21250, United States of America

\* [chinsu@mail.ncyu.edu.tw](mailto:chinsu@mail.ncyu.edu.tw)



**OPEN ACCESS**

**Citation:** Lin C, Popescu SC, Thomson G, Tsogt K, Chang C-I (2015) Classification of Tree Species in Overstorey Canopy of Subtropical Forest Using QuickBird Images. PLoS ONE 10(5): e0125554. doi:10.1371/journal.pone.0125554

**Academic Editor:** Krishna Prasad Vadrevu, University of Maryland at College Park, UNITED STATES

**Received:** November 20, 2014

**Accepted:** March 24, 2015

**Published:** May 15, 2015

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All data, except the QuickBird image, underlying the findings in this study are freely available in this manuscript. The QuickBird image used is a commercial product. The authors are not allowed to distribute this product under the contract limits. But the produced composite of the study site, for example [Fig 1](#), is available to readers.

**Funding:** The author(s) received no specific funding for this work.

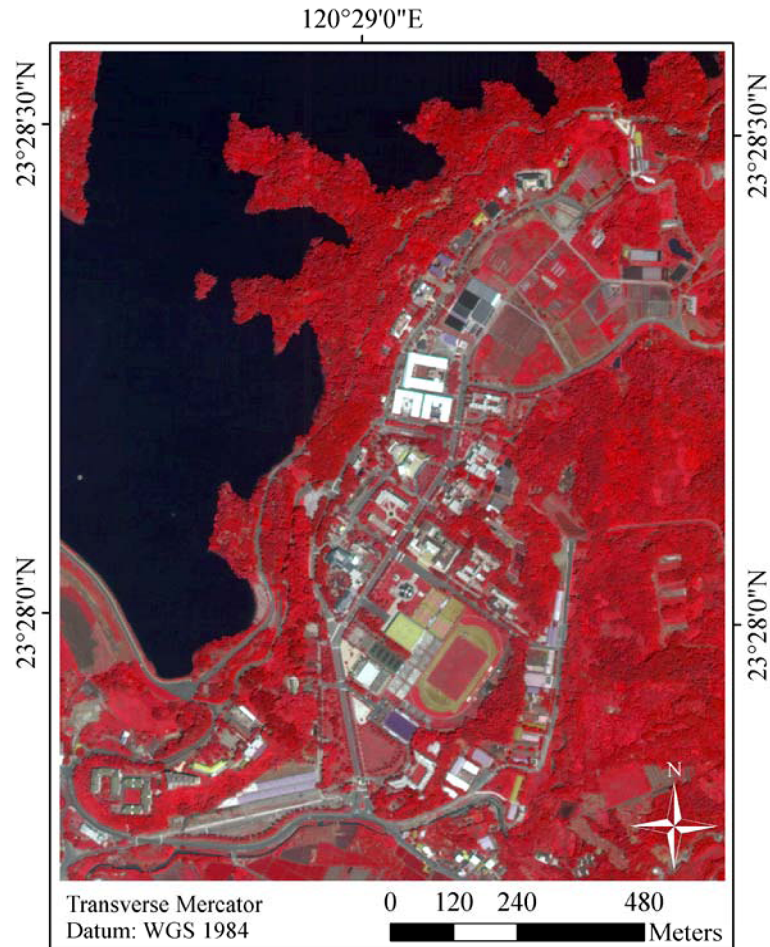
**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

This paper proposes a supervised classification scheme to identify 40 tree species (2 coniferous, 38 broadleaf) belonging to 22 families and 36 genera in high spatial resolution QuickBird multispectral images (HMS). Overall kappa coefficient (OKC) and species conditional kappa coefficients (SCKC) were used to evaluate classification performance in training samples and estimate accuracy and uncertainty in test samples. Baseline classification performance using HMS images and vegetation index (VI) images were evaluated with an OKC value of 0.58 and 0.48 respectively, but performance improved significantly (up to 0.99) when used in combination with an HMS spectral-spatial texture image (SpecTex). One of the 40 species had very high conditional kappa coefficient performance ( $SCKC \geq 0.95$ ) using 4-band HMS and 5-band VIs images, but, only five species had lower performance ( $0.68 \leq SCKC \leq 0.94$ ) using the SpecTex images. When SpecTex images were combined with a Visible Atmospherically Resistant Index (VARI), there was a significant improvement in performance in the training samples. The same level of improvement could not be replicated in the test samples indicating that a high degree of uncertainty exists in species classification accuracy which may be due to individual tree crown density, leaf greenness (inter-canopy gaps), and noise in the background environment (intra-canopy gaps). These factors increase uncertainty in the spectral texture features and therefore represent potential problems when using pixel-based classification techniques for multi-species classification.

## Introduction

The past few decades have witnessed an increase in research concerning classification of land cover/land use via high spatial resolution satellite multispectral imagery, such as SPOT5,



**Fig 1. A false color picture of the QuickBird multispectral image (a composite of band 4, 3, 2 for red, green, and blue) showing the land covers over the study site.**

doi:10.1371/journal.pone.0125554.g001

IKONOS, QuickBird, WorldView-2 and GeoEye. High resolution images capture fine spatial details required for vegetation cover mapping. For example, the QuickBird and GeoEye satellites provide images with high spatial resolution pixel size in the range of 0.61m/2.44m and 0.41m/1.65m for panchromatic and multispectral (MS) data respectively. In addition, the spatial resolution of MS data can be upgraded to the decimeter level by a method of data fusion or pansharpening. Such high spatial resolution (HSR) images can be used to support forest management, for example, by examining biological habitat and estimating tree growth and harvestable timber volume. Other uses include delineation of tree crowns [1–5], tree parameter estimation [6, 7], biodiversity exploration [8], the identification of herbaceous plant species [9], and thinning scenario simulation in recreational forest management [10]. Recently, many studies have focused on species-level forest classification using various satellite images, such as airborne/satellite MS images [11–14], airborne lidar with MS images [15–18], and ground-based lidar or airborne/in situ hyperspectral data [19–24]. In addition to the above studies, research has also been conducted to assess species density and/or species richness [25], vegetation phenology [26, 27], forest succession [28], and invasive woody species [29, 30]. Interestingly, very little work has been done in the past using satellite images to identify and differentiate a large number of tree species, particularly, in tropical/subtropical forest where many tree species

share the overstorey canopy. If overstorey species can be clearly identified, the forest carbon sink could be estimated more accurately according to tree species-based allometric formulas.

This paper explores the potential of using HSR QuickBird MS images (abbreviated as HMS) to identify multiple tree species. Two special features, i.e., spectral and texture information of the images were utilized based on two considerations. First, vegetation indices expand the limited spectral features of vegetation in the QuickBird HMS image and hold potential for tree classification. Second, leaf greenness, crown density (inter-canopy gaps), and background pixels (intra-canopy gaps) often cause noise and uncertainty in spectral analysis of satellite imagery. Therefore devising a method for pixel noise reduction may also increase the potential for tree classification. Four alternative classification schemes that combine input datasets and classifiers were examined in this study. The first used the original set of pansharpened 4-band image (HMS). The second applied a set of 5-band vegetation index images (HMS5VI) derived from the original HMS images. The third classification scheme used 4-band spectral-spatial textural images (SpecTex) obtained from HMS. The final scheme used 13-band image (HMS13B) that integrated HMS, HMS5VI, and SpecTex images. The classifiers applied to the classification schemes were maximum likelihood classifier, Mahalanobis distance, spectral angle mapper, support vector machine, spectral information divergence, and neural network. This paper aims to classify a large number of tree species (40) and has the following main objectives:

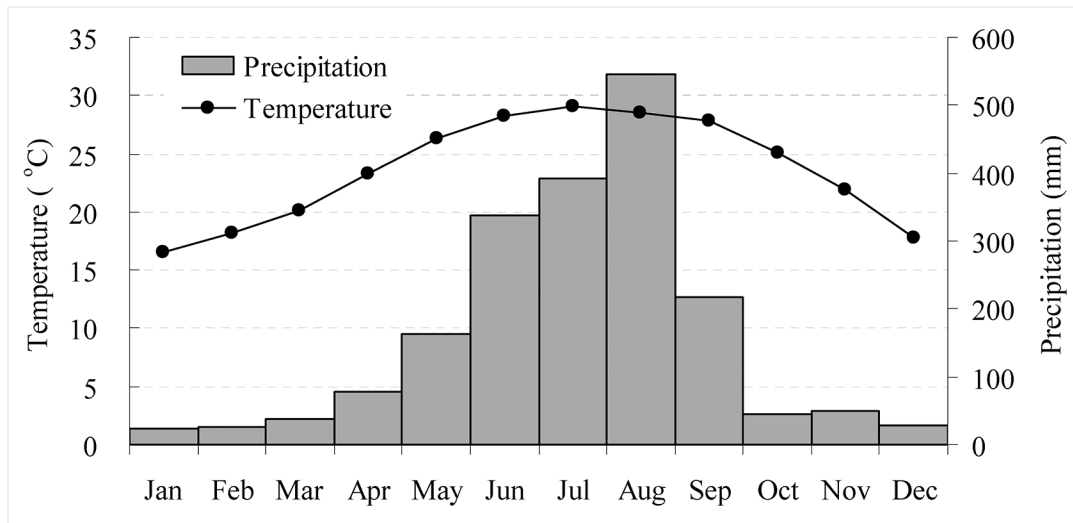
- Determine the baseline accuracy of the proposed classification scheme using training samples of only the original 4-band HMS QuickBird images and their derived vegetation indices.
- Examine if the baseline performance in the training samples can be improved by incorporating spatial texture features derived from the HMS images derived.
- Consider whether a marginal signature can be introduced to improve identification of hard to classify accurately (HCA) species in the training samples (i.e. the conditional kappa coefficient for any specific HCA species is less than 0.95).
- Evaluate the accuracy of classifications using the four datasets and classifiers by retraining with test samples.

## Materials and Methods

### Study site and species-based stratified sampling

Taiwan is located in a tropical/subtropical zone where 12 zonal and 9 azonal vegetation types are widely distributed. Diversity of forest vegetation in Taiwan is strongly influenced by temperature and moisture gradient [31]. The study site shown in Fig 1 is located in the watershed area of Lantan Dam in southern Taiwan centered at 120° 29'03"E and 23° 28'09"N. The average elevation is around 60m above sea level. Historic records from the weather station near the study site showed that the average 10-year annual temperature and precipitation of the area between 2006 and 2014 was 23.6°C and 1940mm respectively. As shown in Fig 2, average monthly temperature and precipitation varied from 17°C to 29°C and 23mm to 545mm and displayed a bell-shaped curve that peaked in July and August respectively.

A ground survey revealed that 40 tree species (Table 1) inhabited the forest overstorey canopy in the study site. A species-based stratified sampling method was applied to extract the identified tree species pixels as reference samples ( $N_1 = 5727$ ) for signature training. After training, an additional dataset also comprised of 40 species was randomly selected for use as test samples ( $N_2 = 4102$ ). Training and test samples were collected using a laser range finder and Leica



**Fig 2. Generalized trends of the average monthly temperature (curve) and precipitation (bars) of the study site for the years 2006 to 2014.**

doi:10.1371/journal.pone.0125554.g002

GS5+ GPS which offers positioning accuracy up to one meter. The image resolution was sufficiently fine to identify the sample trees via image interpretation. In addition to explicitly identifying the neighborhood of the trees in the study site, each of the training and test samples was identified by double-checking the image interpretation with the on-site inventory. Table 1 tabulates the number of pixels of each species found in the training dataset and test dataset.

### Procedures for multispectral image processing and analysis

QuickBird uses a pushbroom sensor to collect image data in a panchromatic band and 4 multispectral bands with a pixel size of 0.61m and 2.44m at nadir, respectively. The coverage of spectral bandwidths is blue: 450–520 nm, green: 520–600 nm, red: 630–690 nm, and near infrared: 760–900 nm, and panchromatic: 445–900 nm. The images used for this study were collected on 24 Aug, 2002, at 10:33am. The images were re-sampled to a pixel size of 0.6m and 2.4m for both the panchromatic and multispectral band images respectively after geometrical correction by the data distribution center. A series of image processing techniques such as atmospheric correction, data fusion (pansharpening), vegetation-index calculation, spatial-texture analysis, and layer stacking were carried out to produce variant datasets for species classification and the evaluation of training performance, test accuracy, and classification uncertainty (Fig 3). Please refer to the following sections for details of the image processing and analysis.

### Multispectral image preprocessing

**Atmospheric correction.** As shown in Eq (1), the digital number of the original 4-band multispectral image was first converted to at-sensor radiance  $L_{\lambda,i}$  ( $Wm^{-2} \mu m^{-1} sr^{-1}$ ) using calibration coefficients ( $a_i$  and  $k_i$ ) and the nominal bandpass width of each band ( $\Delta\lambda_i$ ) accompanying the metadata of the image (Table 2). Then  $L_{\lambda,i}$  was used to derive the reflectance (dimensionless) of surface materials by the atmospheric correction module (ATMOSC) in the commercial software IDRISI. The reflectance at the blue, green, red, and near infrared bands

**Table 1. Tree species used for spectral-spatial texture classification.**

Family	Scientific name	Denoted by	Training samples	Test samples	Family	Scientific name	Denoted by	Training samples	Test samples
Mimosaceae	<i>Acacia auriculiformis</i>	A.a	60	67	Sapindaceae	<i>Koelreuteria formosana</i>	K.f	47	405
Palmae	<i>Areca catechu</i>	A.ca	77	330	Hamamelidaceae	<i>Liquidambar formosana</i>	L.f	89	40
Araucariaceae	<i>Araucaria cunninghamii</i>	A.cu	59	142	Mimosaceae	<i>Leucaena leucocephala</i>	L.l	197	60
Araucariaceae	<i>Araucaria excelsa</i>	A.e	40	159	Lythraceae	<i>Lagerstroemia subcostata</i>	L.s	112	58
Moraceae	<i>Artocarpus heterophyllus</i>	A.h	73	31	Myrtaceae	<i>Melaleuca leucadendrea</i>	M.l	39	15
Apocynaceae	<i>Alstonia scholaris</i>	A.s	54	31	Euphorbiaceae	<i>Mallotus paniculatus</i>	M.p	75	25
Poaceae	<i>Bambuseae</i>	B.	633	81	Sapotaceae	<i>Palaquium formosanum</i>	P.f	46	55
Euphorbiaceae	<i>Bischofia javanica</i>	B.i	40	20	Apocynaceae	<i>Plumeria rubra</i>	P.r	46	180
Caesalpiniaceae	<i>Cassia fistula</i>	C.f	70	69	Strelitziaceae	<i>Ravenala madagascariensis</i>	R.m	63	142
Fagales	<i>Cyclobalanopsis glauca</i>	C.g	55	349	Palmae	<i>Roystonea regia</i>	R.r	38	34
Apocynaceae	<i>Cerbera manghas</i>	C.m	32	26	Bignoniaceae	<i>Spathodea campanulata</i>	S.c	189	758
Annonaceae	<i>Cananga odorata</i>	C.o	315	41	Myrtaceae	<i>Syzygium formosanum</i>	S.f	108	27
Caesalpiniaceae	<i>Cassia siamea</i>	C.si	89	52	Sapindaceae	<i>Sapindus longana</i>	S.l	183	52
Caesalpiniaceae	<i>Cassia suratten</i>	C.su	196	37	Meliaceae	<i>Swietenia macrophylla</i>	S.m	150	64
Caesalpiniaceae	<i>Delonix regia</i>	D.r	274	69	Combretaceae	<i>Terminalia catappa</i>	T.ca	371	134
Fabaceae	<i>Dalbergia sissoo</i>	D.s	118	94	Bignoniaceae	<i>Tabebuia chrysantha</i>	T.ch	58	56
Moraceae	<i>Ficus eladtica</i>	F.e	296	45	Verbenaceae	<i>Tectona grandis</i>	T.g	45	25
Moraceae	<i>Ficus religiosa</i>	F.r	164	20	Ulmaceae	<i>Trema orientalis</i>	T.o	340	48
Caesalpiniaceae	<i>Haematoxylon campechianum</i>	H.c	262	45	Ulmaceae	<i>Ulmus parvifolia</i>	U.p	414	59
Palmae	<i>Heritiera littoralis</i>	H.l	36	16	Ulmaceae	<i>Zelkova serrata</i>	Z.s	122	141

doi:10.1371/journal.pone.0125554.t001

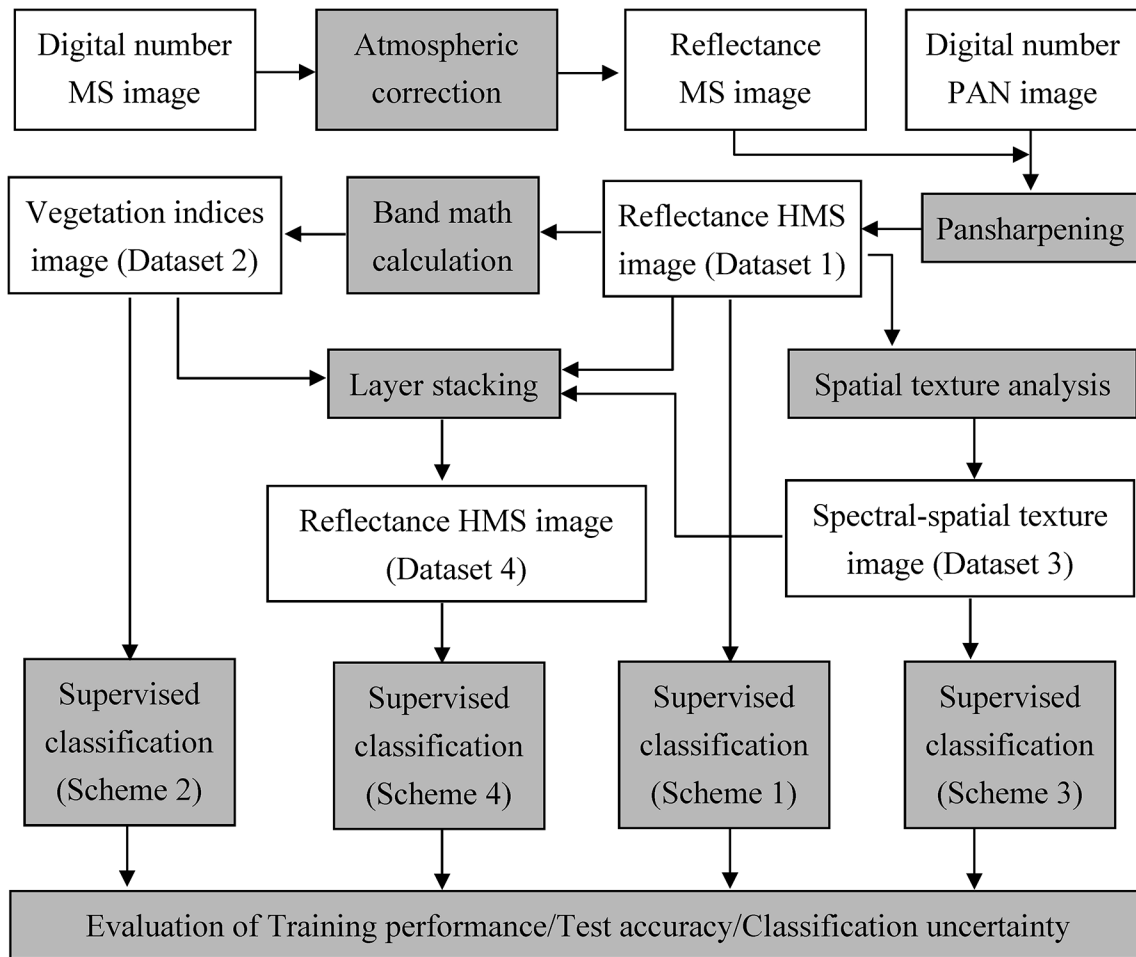
define the spectral signature of the material in a pixel.

$$L_{\lambda_i} = \frac{L_i}{\Delta\lambda_i} = \frac{a_i \cdot k_i \cdot DN_i}{\Delta\lambda_i} \tag{1}$$

**Data fusion.** Geometrically registered QuickBird multispectral images (pixel size: 2.4m) were fused using principal components transformation (PCT)-based fusion techniques according to the following Eq (2) to create a QuickBird HMS image (pixel size: 0.6m).

$$H = \min DN_{PC_1} + \frac{(DN_{Pan} - \min DN_{Pan})(\max DN_{PC_1} - \min DN_{PC_1})}{(\max DN_{Pan} - \min DN_{Pan})} \tag{2}$$

The above PCT-based image fusion is a spectral domain technique with keeping the original spectral bands intact [33]. So, the fused HMS image obtained both the spectral and spatial



**Fig 3. Flowchart of image processing and analyses used in this study.**

doi:10.1371/journal.pone.0125554.g003

information from the original QuickBird MS image and the panchromatic image for individual tree identification [34].

**Expanding image spectral information by including vegetation indices.** Recent efforts have employed several different vegetation indices (VIs) in modeling vegetation biophysical variables such as leaf area index (LAI), canopy closure or percentage green cover, green

**Table 2. Calibration coefficients for absolute radiance conversion of 16-bit QuickBird images<sup>1)</sup>.**

Spectral band	Revised factor “k” ( $Wm^{-2}sr^{-1}count^{-1}$ )	Calibration gains “a” ( $Wm^{-2}sr^{-1}count^{-1}$ )	Nominal bandpass width (nm)
Pan	0.0838188	0.06269	455
Blue	0.0160412	0.01431	70
Green	0.0143847	0.01045	80
Red	0.0126735	0.00968	60
NIR	0.0154242	0.01568	140

<sup>1)</sup> Revised factor for the panchromatic band varies with its exposure levels using time-delayed-integration (TDI); factor value is positively proportional to the TDI level, which can be found in. IMD files. Revised factor was set to 1 for each band for the product generated after June 6, 2003. Coefficients listed in the table address the case for 10 TDI levels. Absolute calibration coefficients for each band are also available from the. IMD file [32].

doi:10.1371/journal.pone.0125554.t002

biomass, and chlorophyll content. Many VIs combine information contained in red and NIR bands and normalize external effects, e.g. solar and viewing angles, and internal effects such as soil variation or topographic condition. Some VIs, investigated in this study and described below also included transformed spectral information resulting from expansion of data dimensionality of the original spectral bands.

The normalized difference vegetation index (NDVI) was developed by Tucker [35] and defined by

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (3)$$

where  $\rho_{nir}$  and  $\rho_{red}$  represents the reflectance of near infrared and red bands respectively. NDVI is a nonlinear index influenced by atmospheric path radiance and affected by an additive noise effect. Since NDVI was derived after image atmospheric correction, it can be used to monitor the seasonal and inter-annual changes in vegetation growth and activity. In particular, it can also help to reduce many forms of multiplicative noise. Huete [36] demonstrated that a potential problem with NDVI is that it is easily affected in the field by background soil. Nevertheless, since this index is widely used for mapping global vegetation cover, the NDVI was included in this study.

In order to cope with the signature biases introduced by soil background, Huete [36] introduced a soil adjusted vegetation index (SAVI) defined as

$$SAVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red} + L} (1 + L) \quad (4)$$

where  $\rho_{nir}$  and  $\rho_{red}$  denotes the reflectance of the near infrared and red bands respectively, and “ $L$ ” is an adjustment factor which minimizes the influence of soil brightness in the NDVI. Huete et al. [37] found that the  $L$  in the SAVI was negatively related to the vegetative percentage in a pixel and suggested that a moderate value of 0.5 for  $L$  could be applied to general cases for various conditions of green cover. Therefore, the SAVI was included in this study.

Kaufman and Tanré [38] indicated that scattering and absorption caused by aerosols spatially varied in the atmosphere on images where continental, maritime, desert, or heavily vegetation cover varied significantly. To address this issue they developed an atmospherically resistant vegetation index (ARVI) defined by

$$ARVI = \frac{\rho_{nir}^* - \rho_{rb}^*}{\rho_{nir}^* + \rho_{rb}^*} \quad (5)$$

where  $\rho^*$  is a reflectance value with prior correction for molecular scattering and ozone absorption,  $\rho_{rb}^* = \rho_{red}^* - \gamma(\rho_{blue}^* - \rho_{red}^*)$  and a priori gamma ( $\gamma$ ) function is introduced into the difference in reflectance between the blue and the red bands ( $\rho_{blue}^* - \rho_{red}^*$ ) to correct reflectance in the red band ( $\rho_{red}^*$ ). The ARVI specified by Eq (5) helps to minimize atmospheric-induced variations in pixel-based vegetation indices as well as stabilize temporal and spatial variations of the indices in atmospheric aerosol content. If the aerosol model is not known a priori, the  $\gamma$  function is normally set to 1.0 to minimize atmospheric effects.

Gitelson et al. [39] developed the visible atmospherically resistant index (VARI) for remote estimation of vegetation fraction. They found that VARI had minimal sensitivity to atmospheric effects because the estimation error of vegetation fraction was less than 10% in a wide range of atmospheric optical thickness. VARI has a strong linear relationship with vegetation fraction

as demonstrated by Viña et al. [40]

$$VARI_{green} = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}} \tag{6}$$

where  $\rho_{blue}$ ,  $\rho_{green}$ , and  $\rho_{red}$  represent the reflectance of the blue, green, and red bands respectively.

The enhanced vegetation index (EVI) was also introduced by Huete et al. [41] to optimize the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction of atmospheric influences. It is defined by

$$EVI = G \times \frac{\rho_{nir}^* - \rho_{red}^*}{\rho_{nir}^* + C_1 \cdot \rho_{red}^* - C_2 \cdot \rho_{blue}^* + L} \tag{7}$$

where  $\rho^*$  is the atmospherically corrected or partially atmosphere-corrected reflectance,  $G$  is a gain factor,  $L$  is the canopy background adjustment factor that addresses nonlinear, differential NIR and red radiant transfer through a canopy, and  $C_1$ ,  $C_2$  are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band. According to Huete et al. [37], the coefficients used in the EVI index are generally set by  $L = 1$ ,  $C_1 = 6$ ,  $C_2 = 7.5$ , and  $G = 2.5$ .

For vegetation species classification, leaf greenness, percentage variations of mesophyll, and physiological photosynthesis effects need to be considered because these factors might directly or indirectly influence the reflectance of vegetation. According to Larcher [42], the maximum value for net photosynthesis of plants varies dramatically from 17 to 68 ( $\mu \text{ mol m}^{-2} \text{ s}^{-1}$ ). This indicates that trees use the light energy of blue and red spectra differently. Moreover, chlorophyll (the component makes the leaf greenish) is a major factor in photosynthesis. The concentration of leaf pigments generally changes with the seasons and this is the basis of phenological studies. So, it is assumed that signals of blue and green spectra which are usually neglected in forest classification will become essential variables in species recognition. This is the main reason that the three VIs, ARVI, VARI, and EVI were introduced in this study. These spectra indexes have potential in species classification.

### Deriving the spatial texture images from the spectral band images

A spectral image takes into account context, edges, texture and color (tonal variation). The ability to recognize detail is strengthened by considering texture as well as the spatial features of objects of interest. Therefore, each spectral band image produced a spatial texture image which was used in conjunction with the spectral images for data analysis. As a result, an  $L$ -band remotely sensed image produced an  $L$ -band spectral-spatial texture (SpecTex) image. These images possess not only spectral information provided by the original spectral band images, but also additional spatial information characterized by texture features in spatial texture images derived from each of  $L$  spectral band images.

QuickBird image quality is frequently degraded by noise such as random spikes and impulse imperfections [43] which can be limited by using a Lee-sigma filter, a kind of focal filtering analysis [44, 45]. The Lee-sigma filter is a standard deviation-based (sigma) filter that processes data based on statistics calculated within individual filter windows. Unlike a typical low-pass smoothing filter, the Lee-sigma filter produces a texture image which preserves image sharpness and details while suppressing noise. It has been shown to be useful for paddy field extraction from various similar agricultural crops [46] and was therefore appropriate to this study. The value of the pixel being processed, denoted by  $DN_c$  with the subscript “c” indicating the



center pixel in a surrounding window, was replaced by an averaged value calculated from the surrounding pixels in that specific window whose DNs are within a specified threshold. In our experiments, this specific threshold was chosen according to the local noise DN standard deviation,  $\sigma_{local}$  and the resulting acceptable range of the center pixel DN<sub>c</sub> was  $DN_c \pm 2\sigma_{local}$  [46, 47].

## Assessment of pixel-based classification and accuracy

The supervised maximum likelihood classifier (MLC) [48] was applied to each of the images, i.e., HMS, HMS5VI, SpecTex, and HMSB13, for species classification where the overall kappa coefficient (OKC) and users' accuracy based species conditional kappa coefficient (SCKC) were calculated based on the error matrix for accuracy assessment. Allouche et al. [49] indicated that the kappa value depended on the prevalence of a particular species. But Vach [50] also demonstrated that of the prevalence of a particular species would have negligible influence. Based on a simulation study of kappa variance, Stehman [51] concluded that the kappa estimator offers little bias even at relatively small sample size. In this study, the true prevalence of a particular species was unknown *a priori*. We applied the stratified random sampling technique to make sure that training and test samples could be selected from each of 40 species. The OKC and SCKC of different datasets were calculated using the same data samples. As a result, if any prevalence dependence existed, its marginal influence would have been the same among those classifications so as to make such classification comparison acceptable.

The introductory section of this paper raised four important issues that can now be addressed. Firstly, the training performance of each of the four classification schemes can be assessed based on the OKC values obtained from using the training samples. The highest OKC value can be considered as the potential highest accuracy (PHA) achievable via high spatial resolution imagery. Secondly, any specific HCA species may be identified if the species conditional kappa coefficient (SCKC) is less than 0.95. Finally, the test accuracy of the classification schemes using MLC classifier is determined from the test samples which are totally independent of the training samples and then classification uncertainty is derived as the difference of training performance and test accuracy. After that, the training performance/test accuracy/classification uncertainty of the four classification schemes were calculated using the same training and test samples and other classifiers such as Mahalanobis distance (MD) [48], spectral angle mapper (SAM) [52], support vector machine (SVM) [53], signature information divergence (SID) [54], and neural net (NN) [48]. A comparison of the classifiers' efficiency in terms of raising training performance, test accuracy, and uncertainty was then made.

## Statistical Analysis of the classification scheme accuracy

Training performance and test accuracy were based on the OKC value of a classification scheme which combined the input datasets and classifiers. Variations in training performance and test accuracy involved two independent variables (the classifiers and the datasets). In the other words, a particular combination of classifiers and datasets produced performance-OKC and accuracy-OKC was determined via the training and test samples. For a particular type of OKCs, the minimum OKC (OKC<sub>min</sub>) in the classifications using the original HMS image (classification scheme 1) by variant classifiers can be selected as a standard level for statistical comparisons. As a result, the accuracy improvement efficiency (AIE) of any particular combination of classifiers and datasets can be determined as the ratio of OKC increment using Eq (8).

$$AIE = (OKC - OKC_{min}) / OKC_{min} \quad (8)$$

A fixed model-based, 4x6 two-way factorial experiment design defined by Eq (9) was

applied to account for the main effects and the interaction of the two variables (classifiers and datasets) on the dependent variable  $Y$ , that is the AIE in Eq (8).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon \quad (9)$$

In Eq (9), the coefficients  $\beta_i$  and  $\beta_{ij}$  represent the effects of the two main variables ( $X_1$  and  $X_2$ ) and the two-variable interactions ( $X_1 X_2$ ) on the dependent variable  $Y$ , i.e., the AIE, and  $\varepsilon$  is the error term. In the factorial ANOVA analysis, an effect is said to be significant if its coefficient has a probability less than or equal to the significant probability 0.05. Then, a moderate conservative post hoc test, Duncan's new multiple range test was also applied to interpret multiple comparisons.

## Results

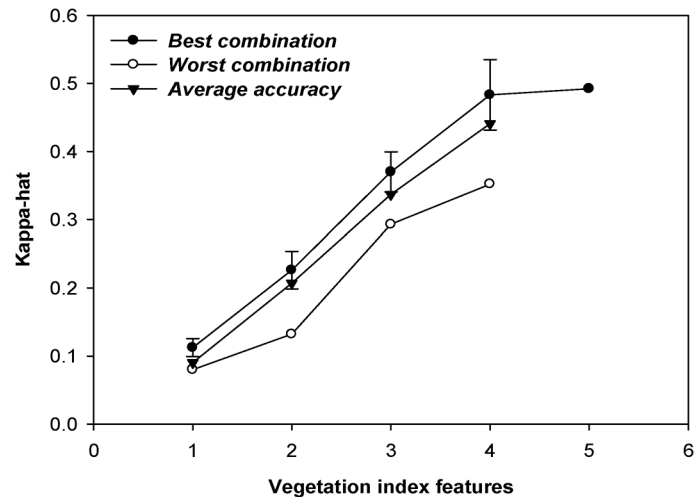
### Performance baseline of multispectral signatures and their derived VI signatures

The overall training performance of the 40-species classification system using dataset 1, the HMS images, was used as a baseline for classification performance. It was evaluated at  $OKC = 0.58$  with a standard deviation 0.20 and the minimal and maximal SCKC values for one of the 40-species were 0.21 and 0.99. If the classification was performed using only dataset 2, i.e., the HMS5VI images, the performance was  $OKC = 0.48$  with a standard deviation 0.21, and SCKC values of every species varied between 0.18 and 0.99. These results demonstrate that baseline classification performance using solely the vegetation indices was around 18% lower than the HMS images. Some of the species could be successfully detected, but the overall results were generally not acceptable if the classification was done using only the original 4-band HMS images or its derived 5-band HMS5VI images.

It was also interesting to consider the possibility of a dependency among the VIs in the species classification. Based on our experimental results, the  $OKC$  of the tested 40-species classification was below 0.10 when a single VI was used for classification. This is due to similarities between the spectral signatures of these 40 tree species. This situation was somewhat alleviated by including more VIs in the classification, but the improvement in training performance was dependent on the number of VIs used. In order to enhance the findings, various combinations of the five VIs, i.e., NDVI, SAVI, EVI, ARVI, and VARI were used. In particular, classification was performed using combinations of 2-, 3-, 4-, and 5-VIs and then compared and analyzed. Fig 4 shows the  $OKC$  of the 40-species classification for dataset 2 using various numbers of used HMS5VI images where the best, average and worst performances were plotted. The improvement in performance appeared to be linear with the number of HMS5VI images used for dataset 2.

### Use of spectral-spatial texture signatures to improve tree species classification

**Difference between tree-species' spectral signatures in HMS and SpecTex images.** Table 3 lists the transformed divergence (TD) of any two features of the 40 species being compared derived from the 4-band HMS data (the entries in the upper triangle with a gray background) and the 4-band SpecTex data (the entries in the lower triangle with a green background). A blank entry indicates a value of  $TD = 2000$  corresponding to cases where two classes of spectral features are separable. The entry with  $1900 \leq TD < 2000$  is highlighted in red indicating that good separation was observed between the corresponding two classes. Other cases with  $TD < 1900$  meant the two classes of spectral features were not clearly separable.



**Fig 4. Relationship of classification accuracy using dataset 2 with various numbers of used VI images.**

doi:10.1371/journal.pone.0125554.g004

Obviously, the 40 tested tree species whose spectral features are highly similar in the HMS image (most of the entries in the upper triangle of the TD matrix) have values less than 1900. In other words, more than 50% of the species showed a lower level of spectral dissimilarity compared with other species. Fortunately, the poor separation can be significantly improved by using spectral-spatial texture transformation. This can be seen in the lower triangle entries of the TD matrix in Table 3. There are only 12 pairs of between-class TDs among all of the tested 780 pairs that are less than 2000, which was the maximum upper margin of the TD value. Actual TD values between-species are: 1999 for *Z.s-S.l*, *U.p-L.f*, *A.e-H.l*; 1998 for *H.c-T.o*; 1997 for *S.m-B.i*; 1996 for *A.cu-H.l*; 1992 for *S.m-T.o*; 1985 for *T.ch-D.r*; 1975 for *F.r-C.o*; 1961 for *T.ch-C.o*; 1902 for *R.m-M.l*; and 1734 for *H.c-S.m*. According to Jensen [55], Lin [56], and Dwivedi et al. [57], we can conclude that there was only one pair of two-species among all the 40-species whose signatures were inseparable in the SpecTex dataset. The two species are *Hae-matoxylon campechianum* vs. *Swietenia macrophylla* (TD = 1734).

**Improvement of training performance after integrating HMS images, VIs, and spectral-spatial texture images.** The QuickBird HMS data pixel size is smaller than one meter, so an overstorey tree crown will encompass more than one pixel of contiguous space. The height variations and the vertical structure changes in crown pixels might introduce significant noise which may make spectral signatures of tree species complicated. Applying the Lee-sigma filter to the original image can remove the noise and produce a more homogeneous spectral-spatial texture image for improved recognition of species spectra. The spectral-spatial texture not only improved the statistical distance between species, but also maximized their signature divergences. These benefits are demonstrated in Table 3. In addition, Table 4 shows the results of a comprehensive comparative study among the multiple species using four data sets; (1) dataset 1: four original bands, HMS (2) dataset 2: five VI images, HMS5VI (3) dataset 3: four spectral-spatial textural images, SpecTex and (4) dataset 4: 13 layer of images (HMS13B) which are combined by the HMS, HMS5VI, and SpecTex images.

SI Table, shows the SCKC of the 40 tree species. When the HMS image was used for classification, only the species *T.ca* had a nearly perfect conditional accuracy with SCKC = 0.99 compared to the species *P.f* with SCKC = 0.94, three species: *B.*, *M.l*, and *P.r* with  $0.80 < \text{SCKC} < 0.90$ , 7-species: *A.s*, *B.i*, *D.r*, *L.l*, *R.m*, *S.c*, and *T.o* with  $0.70 < \text{SCKC} < 0.80$ , 5 species: *D.s*, *F.e*, *H.*

**Table 3. Full-bands transformed divergences of between-species in training samples for the original spectral signatures and the spectral-spatial texture signatures. \***

	HMS image																			
	<i>A.cu</i>	<i>T.ch</i>	<i>P.f</i>	<i>Z.s</i>	<i>U.p</i>	<i>A.ca</i>	<i>A.s</i>	<i>R.m</i>	<i>L.S</i>	<i>C.si</i>	<i>C.m</i>	<i>D.s</i>	<i>A.h</i>	<i>M.p</i>	<i>T.ca</i>	<i>C.g</i>	<i>A.a</i>	<i>H.c</i>	<i>C.f</i>	<i>L.I</i>
<i>A.cu</i>	–	1983	1997	1937	1885		1955	1998		1769		1798						1998	1912	1623
<i>T.ch</i>		–		1952	1986		1998	1999	1998	1898		1976			1995			1993	1984	1783
<i>P.f</i>			–	1992								1999								
<i>Z.s</i>				–	1798		1968		1999	1827		1745						1999	1804	1307
<i>U.p</i>					–		1887	1997	1724	1154	1997	1104	1966		1998	1966		1517	702	1059
<i>A.ca</i>						–				1955			1888		1911	1911		1980		
<i>A.s</i>							–			1411	1978	1911						1998	1772	1803
<i>R.m</i>								–	1944				1999					1998		1950
<i>L.S</i>									–	1986		1970	1444	1964	1981		1804	1365	1956	1843
<i>C.si</i>										–	1998	1601	1999		1999			1946	724	827
<i>C.m</i>											–	1998						1999	1999	
<i>D.s</i>												–	1995		1999			1753	1079	1404
<i>A.h</i>													–	1998	1775	1998	1942	1883	1992	1962
<i>M.p</i>														–						
<i>T.ca</i>															–	1998		1785	1998	1999
<i>C.g</i>																–		1991		1998
<i>A.a</i>																	–	1999		
<i>H.c</i>																		–	1870	1891
<i>C.f</i>																			–	1124
<i>L.I</i>																				–
<i>F.r</i>																				
<i>S.I</i>				1999																
<i>F.e</i>																				
<i>S.f</i>																				
<i>L.f</i>					1999															
<i>A.e</i>																				
<i>T.g</i>																				
<i>S.m</i>																			1734	
<i>S.c</i>																				
<i>R.r</i>																				
<i>D.r</i>		1985																		
<i>M.I</i>								1902												
<i>C.su</i>																				
<i>H.I</i>	1996																			
<i>P.r</i>																				
<i>K.f</i>																				
<i>B.</i>																				
<i>T.o</i>																			1998	
<i>C.o</i>		1961																		
<i>B.i</i>																				
	SpecTex image																			

(Continued)

*c*, *M.p*, and *U.p* with  $0.70 < \text{SCKC} < 0.60$ ; and the others with SCKC less than 0.60. It was also observed that the misclassification rate was increased when the classification was made on the 5-band VI image. Two of the tested species, i.e, *B.* and *T.ca* had good accuracy ( $\text{SCKC} > 0.90$ )

Table 3. (Continued)

	HMS image																					
	<i>F.r</i>	<i>S.J</i>	<i>F.e</i>	<i>S.f</i>	<i>L.f</i>	<i>A.e</i>	<i>T.g</i>	<i>S.m</i>	<i>S.c</i>	<i>R.r</i>	<i>D.r</i>	<i>M.I</i>	<i>C.su</i>	<i>H.I</i>	<i>P.r</i>	<i>K.f</i>	<i>B.</i>	<i>T.o</i>	<i>C.o</i>	<i>B.i</i>		
<i>A.cu</i>	<b>1986</b>	1887		<b>1970</b>	<b>1998</b>	1833		<b>1992</b>		1868	<b>1991</b>	<b>1998</b>		<b>1993</b>							<b>1988</b>	
<i>T.ch</i>	<b>1998</b>	<b>1953</b>	1824	<b>1993</b>	<b>1986</b>	<b>1997</b>		<b>1975</b>	<b>1999</b>	<b>1982</b>	1816			<b>1962</b>				<b>1993</b>	<b>1932</b>			
<i>P.f</i>				<b>1993</b>																		
<i>Z.s</i>	<b>1999</b>	<b>1931</b>	<b>1999</b>	1343	<b>1992</b>	1451		<b>1982</b>		<b>1999</b>	<b>1975</b>	<b>1940</b>		1859	<b>1999</b>							
<i>U.p</i>	1447	991	<b>1997</b>	1738	1642	1661		1593	<b>1996</b>	<b>1965</b>	1208	<b>1990</b>	<b>1999</b>	<b>1968</b>		<b>1938</b>	<b>1993</b>	<b>1990</b>			1434	
<i>A.ca</i>								<b>1993</b>			<b>1997</b>					<b>1999</b>	<b>1979</b>				<b>1999</b>	
<i>A.s</i>	<b>1997</b>	<b>1994</b>		1799	<b>1997</b>	<b>1958</b>		<b>1999</b>		1812	1896			1836							<b>1987</b>	
<i>R.m</i>	<b>1972</b>	<b>1999</b>	<b>1992</b>			<b>1997</b>		<b>1956</b>			<b>1999</b>		<b>1999</b>					<b>1999</b>			<b>1998</b>	
<i>L.S</i>	1238	1854	1880		1863	<b>1976</b>	<b>1996</b>	1341	<b>1905</b>		1754		1812			1790	1263	<b>1914</b>			1322	
<i>C.si</i>	<b>1944</b>	1684	<b>1989</b>	1118	1662	1719		<b>1904</b>	<b>1997</b>	1552	1219			1888	<b>1999</b>			<b>1993</b>			<b>1904</b>	
<i>C.m</i>						<b>1984</b>															<b>1998</b>	
<i>D.s</i>	<b>1947</b>	1533	<b>1998</b>	<b>1909</b>	<b>1979</b>	1780		1661		<b>1986</b>	1775	<b>1932</b>		<b>1964</b>	<b>1997</b>	<b>1967</b>		<b>1998</b>			<b>1927</b>	
<i>A.h</i>	<b>1959</b>	<b>1966</b>	1752		<b>1978</b>	<b>1994</b>		1418	<b>1905</b>		<b>1902</b>		1871				1429	1885			<b>1969</b>	
<i>M.p</i>																		<b>1981</b>				
<i>T.ca</i>	<b>1995</b>	<b>1965</b>	1617		<b>1996</b>			<b>1964</b>	<b>1992</b>		<b>1975</b>		<b>1999</b>					<b>1942</b>	<b>1949</b>	<b>1984</b>	<b>1999</b>	
<i>C.g</i>	<b>1997</b>					<b>1999</b>		<b>1919</b>			<b>1967</b>					<b>1998</b>					<b>1984</b>	
<i>A.a</i>	<b>1999</b>		<b>1976</b>						<b>1950</b>				1869				1408	1987			1998	
<i>H.c</i>	1429	1001	<b>1901</b>	<b>1997</b>	<b>1944</b>	<b>1925</b>		1419	<b>1790</b>		1704		<b>1945</b>			<b>1993</b>	1849	1847			1135	
<i>C.f</i>	<b>1958</b>	1603	<b>1992</b>	1539	1653	1872		1727	<b>1999</b>	<b>1938</b>	1069	<b>1994</b>		<b>1922</b>	<b>1999</b>	1818	<b>1999</b>	<b>1993</b>			<b>1917</b>	
<i>L.I</i>	1655	1269	<b>1957</b>	1610	1572	1446		1623	<b>1982</b>	1885	1416	<b>1998</b>	<b>1999</b>	1859			<b>1995</b>	<b>1971</b>			1808	
<i>F.r</i>	–	1632	<b>1982</b>	<b>1994</b>	1895	1869		1724	<b>1953</b>	<b>1996</b>	1765			<b>1980</b>			<b>1998</b>	<b>1964</b>	<b>1979</b>		1001	
<i>S.J</i>		–	<b>1988</b>	1872	1705	1543		1722	<b>1980</b>	<b>1980</b>	1515	<b>1998</b>		1997		<b>1999</b>	<b>1994</b>	<b>1955</b>			1814	
<i>F.e</i>			–		<b>1990</b>	<b>1999</b>		<b>1941</b>	1684		<b>1931</b>		1167				1721	955	<b>1998</b>	<b>1988</b>		
<i>S.f</i>				–	<b>1920</b>	1566				<b>1981</b>	<b>1916</b>			1823							<b>1997</b>	
<i>L.f</i>					–	<b>1938</b>		<b>1946</b>	<b>1900</b>	<b>1992</b>	1124		<b>1998</b>	<b>1998</b>		<b>1998</b>	<b>1984</b>	<b>1987</b>			<b>1955</b>	
<i>A.e</i>						–		<b>1958</b>		<b>1990</b>	1948			<b>1938</b>				<b>1999</b>			<b>1929</b>	
<i>T.g</i>							–				<b>1999</b>							<b>1999</b>				
<i>S.m</i>								–	<b>1971</b>		1567		<b>1969</b>			<b>1912</b>	1799	1883			1628	
<i>S.c</i>									–		<b>1936</b>		1729				1756	1477			<b>1924</b>	
<i>R.r</i>										–	<b>1967</b>				<b>1998</b>						<b>1996</b>	
<i>D.r</i>											–		<b>1997</b>	<b>1981</b>		1897	<b>1953</b>	<b>1967</b>			1702	
<i>M.I</i>												–										
<i>C.su</i>													–				1471	1250			<b>1981</b>	
<i>H.I</i>						<b>1999</b>								–								
<i>P.r</i>																–						
<i>K.f</i>																		–	<b>1992</b>		<b>1987</b>	
<i>B.</i>																			–	1440	<b>1921</b>	
<i>T.o</i>									<b>1992</b>										–		<b>1969</b>	
<i>C.o</i>	<b>1975</b>																				–	
<i>B.i</i>									<b>1997</b>													–

SpecTex image

\*: the “blank” entries have TD = 2000, the bold number highlighted 1900<TD≤2000 represents spectral features are good separable.

**Table 4. Comparisons of the training-samples-based SCKC for the five HCA-species and the OKC for all 40 species in the classification using various data sets.**

Data sets	SpecTex4 bands	SpecTex+ARVI	SpecTex+EVI	SpecTex+NDVI	SpecTex+SAVI	SpecTex+VARI	HMS13B13bands
SCKC							
<i>C.o</i>	0.68	0.77	0.80	0.76	0.79	0.94	1.00
<i>H.l</i>	0.84	0.84	0.93	0.86	0.93	1.00	1.00
<i>P.f</i>	0.94	0.94	0.96	0.94	0.96	0.94	1.00
<i>R.m</i>	0.90	0.92	0.94	0.90	0.91	1.00	1.00
<i>S.m</i>	0.94	0.95	0.98	0.95	0.98	0.97	1.00
OKC							
40 species	0.9861	0.9898	0.9931	0.9900	0.9925	0.9945	0.9971

doi:10.1371/journal.pone.0125554.t004

and 9 species: *D.r*, *F.r*, *M.l*, *P.r*, *R.m*, *R.r*, *S.c*, *S.l*, and *T.o* were classified with a relatively better accuracy (SCKC>0.6), most of the 40 species were classified at very low SCKC values compared to the results produced by using the 4-band original image.

The SpecTex image signatures of the 40 tree species were nearly correctly recognized (SCKC≥0.95) with the exception of 5 species (Table 4) whose SCKC values ranged from 0.68–0.94. Those five species are *C.o*: *Cananga odorata*, *H.l*: *Heritiera littoralis*, *R.m*: *Ravenala madagascariensis*, *S.m*: *Swietenia macrophylla*, and *P.f*: *Palaquium formosanum*; they can be considered as the hard to classify accurately (HCA) species. Fortunately, those HCA species can be successfully recognized using the HMS13B: the integrated image of HMS, HMS5VI, and SpecTex.

Combining the 5-VI (NDVI, SAVI, ARVI, VARI, and EVI) into one image (HMS5VI) only achieved a 0.48 OKC value which was worse than the 0.58 OKC obtained using the original 4-band HMS image. The reduced OKC of 0.10 was probably due to the fact that each of the VIs use only a partial set of spectral signatures from the HMS image. On the other hand, using only 4-band SpecTex image achieved as high as OKC = 0.98 with improvements of 104% and 69% compared to the 5-band HMS5VI image and the original 4-band HMS image respectively. By contrast, all the tested 40 tree species were successfully classified with an overall performance OKC = 0.98 when the 4-band SpecTex image was used for classification. Moreover, an overall performance OKC = 0.9971≈1.00 was achieved when the integrated image HMS13B was used for classification (Table 4).

## Discussion

### VI as the critical signature for recognizing HCA species

The two issues that were of particular interest in this study were mentioned previously, (1) What is the potential highest accuracy (PHA) that can be achieved using the high spatial resolution satellite images? (2) What is the critical marginal signature that could be used to improve species classification up to the PHA level? These signatures can be used as a guide for figuring out key features that contain crucial information for recognition of those species that are difficult to discriminate, as well as features that contain redundant information which can be removed from the classification. To satisfy such needs, we have to understand the potential of the various combinations of high spatial resolution images.

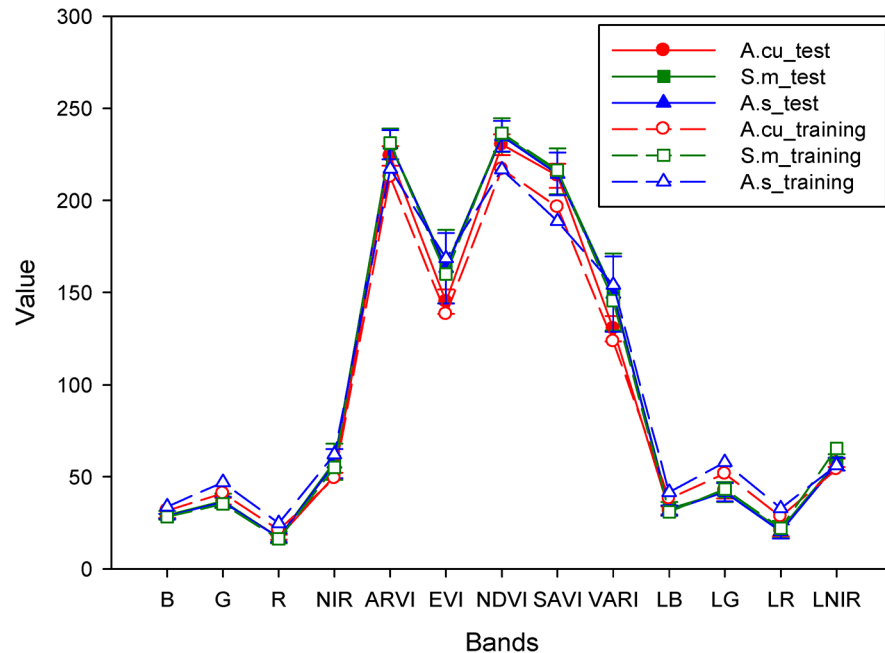
Overall performance was OKC = 0.99 for the 40 species. Very few pixels were misclassified in terms of between-species discrimination (TD can be seen in the lower triangle matrix in Table 3). PHA for each of the tested species is shown in column “HMS13B” next to “SpecTex” in S1 Table. Five species proved hard to classify accurately (HCA). They were, *Cananga*

*odorata*, *Heritiera littoralis*, *Palaquium formosanum*, *Ravenala madagascariensis*; and *Swietenia macrophylla* and are highlighted in [S1 Table](#). Two data sets, those derived from the SpecTex images and those derived from the HMS13B images proved to be effective according to the results in [S1 Table](#). In addition to these two data sets, data sets that combined SpecTex images with one VI image were included to see whether or not including a VI image into the 4 spectral-spatial texture images could improve classification performance. VARI was shown to be a very good predictor for vegetation fraction estimation in [\[39\]](#) and [\[40\]](#). The VARI vegetation index can organize blue, green, and red spectra more efficiently to present leaf chlorophyll, color information and crown coverage. Accordingly, it is suggested that VARI is able to retain heterogeneous spectra of those visible spectral bands and thus contribute better marginal performance for species classification as shown in [Table 4](#). Adding the VARI to the SpecTex images significantly improved the training performance, especially in three species, *Cananga odorata*, *Heritiera littoralis*, and *Ravenala madagascariensis*; while another four species showed little additional performance. Thus, classification using HMS13B images, which integrated heterogeneous signatures from HMS, HMS5VI, and SpecTex, offered the highest SCKC and OKC performance for the tree species tested in this study ([Table 4](#)).

### Species classification uncertainty

Assessment of the training samples in the previous section showed good potential for species classification using HMS13B images. However, a large uncertainty of species mapping was still visible in the integrated spectral-spatial texture QuickBird images because of unsatisfactory overall kappa accuracy ( $OKC < 0.20$ ) obtained in the test samples of those species. Based on ground checks after classification, examples of individual trees of the same species were observed with different foliage color and foliage density in their test and training samples. The amount of leaf greenness reflects the chlorophyll concentration and the ability of leaf photosynthesis [\[58\]](#). Foliage density meanwhile is the most widely used and most rapid method of determining forest health [\[59\]](#). High variations of reflectance caused by inter- and intra-canopy gaps and leaf color changes were the major factors that reduced species mapping accuracy. [Fig 5](#) demonstrates the signature variations of tree species in the study site. In this figure, signatures of the training samples and test samples of *Araucaria cunninghamii*, *Alstonia scholaris*, and *Swietenia macrophylla* significantly overlap each other.

The uncertainty of tree species classification can also be examined visually by a close look at [Fig 6](#). It can be seen, that the training samples of the 40 species highlighted by polygons (denoted r-Species code) were mostly classified accurately; however, the test samples (abbreviated e-Species code) were mostly confused. Most of the test samples of the 40 species showed a diverse colors indicating the test accuracy was not optimistic and the classification of a large number of tree species suffered from high uncertainty that was probably caused by tree physiology and physical properties of tree crowns. In contrast to [Fig 6B](#), [S2 Table](#) shows the test-samples-based SCKC of the MLC classification. It reveals a serious problem of misclassification in the species classification. Most of the 40 species have a SCKC less than 0.20 in every classification scheme. In other words, the number of species with  $SCKC \geq 0.20$  was 6, 5, 5, and 9 for classification schemes 1, 2, 3, and 4 respectively. The largest SCKC was in classification scheme 4 and equaled 1.00 which was greater than 0.73, 0.39, and 0.53, the largest SCKC in classification scheme 1, 2, and 3. A difference between SCKC values in training- and test-samples has also been demonstrated.



**Fig 5. An example of the species classification uncertainty caused by spectral signatures variations.**

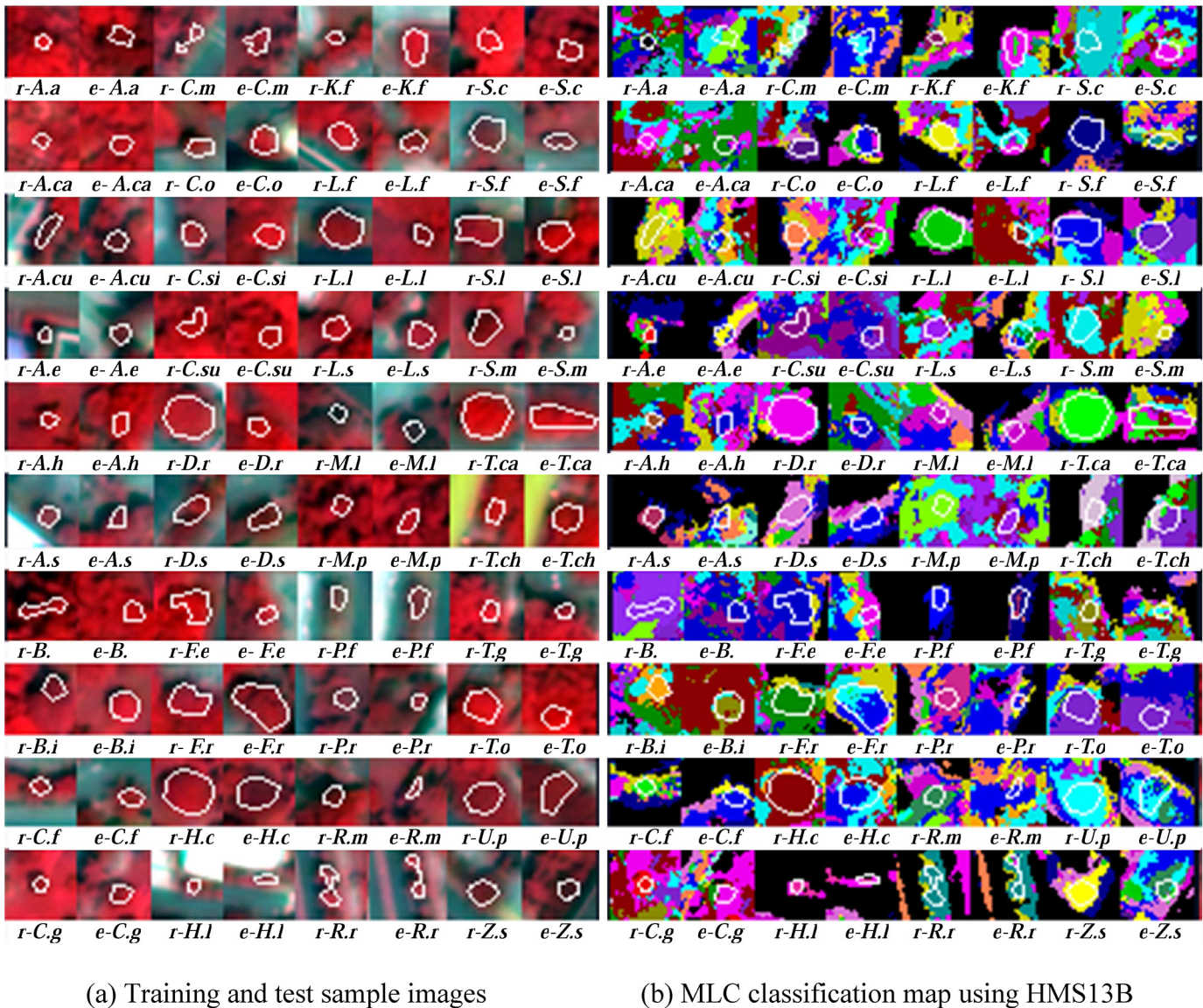
doi:10.1371/journal.pone.0125554.g005

### Comparison of classifiers and input datasets on classification efficiency

Fig 7A–7C shows the diversity of training performance, test accuracy, and classification uncertainty in the classifiers. This kind of variation also appeared in the four datasets by variant classifiers. The highest and the lowest training performance (0.98 and 0.04) were achieved by MLC and NN that was carried out using the HMS13B and HMS5VI images respectively. The average and standard deviation of the classifiers’ performance derived from the four datasets are MLC  $0.79 \pm 0.16$ , SVM  $0.66 \pm 0.16$ , MD  $0.51 \pm 0.21$ , NN  $0.46 \pm 0.33$ , SAM  $0.33 \pm 0.16$ , and SID  $0.23 \pm 0.17$ . Differences between performance and accuracy were repeated in each of the other classifiers. The classifiers’ test accuracy was NN  $0.10 \pm 0.03$ , MLC  $0.09 \pm 0.01$ , MD  $0.08 \pm 0.06$ , SVM  $0.08 \pm 0.05$ , SAM  $0.08 \pm 0.04$ , and SID  $0.06 \pm 0.06$ . Classification uncertainty was defined as the loss of OKC which was calculated as the difference between training performance and test accuracy. The vanished OKC of the classifiers was MLC  $0.70 \pm 0.13$ , SVM  $0.58 \pm 0.13$ , MD  $0.43 \pm 0.15$ , NN  $0.36 \pm 0.31$ , SAM  $0.25 \pm 0.13$ , and SID  $0.17 \pm 0.12$ .

Based on the minimum OKC of classification using the HMS image (classification scheme 1), that is performance-OKC = 0.12 and accuracy-OKC = 0.03 for the classifier SID, a further statistical test of the OKC increment achieved by all combination of classifiers and datasets was carried out by a two-way ANOVA test. As shown in S3 Table, the effect of classifiers ( $F = 3.234$ , sig. probability = 0.023) and datasets ( $F = 10.320$ , sig. probability < 0.001) on the change of AIE was significant, and the interaction was insignificant ( $F = 0.215$ , sig. probability = 0.998). In addition, S4 and S5 Tables showed the Duncan’s grouping of AIEs in the aspect of classifiers and datasets respectively. As shown in S4 Table, the classifications using the HMS13B image (classification scheme 4) achieved the highest value of AIE = 4.35 which is significantly greater than the others. AIE of the SpecTex (classification scheme 3) was 2.78 which is statistically equal to 1.55 the AIE of the HMS (classification scheme 1) at the significant probability = 0.067. Of all the datasets, the HMS5VI has the smallest AIE of 1.09 (S4 Table). S5 Table shows the AIE of the classifiers ranging from 0.97 to 3.83. Although the values have been





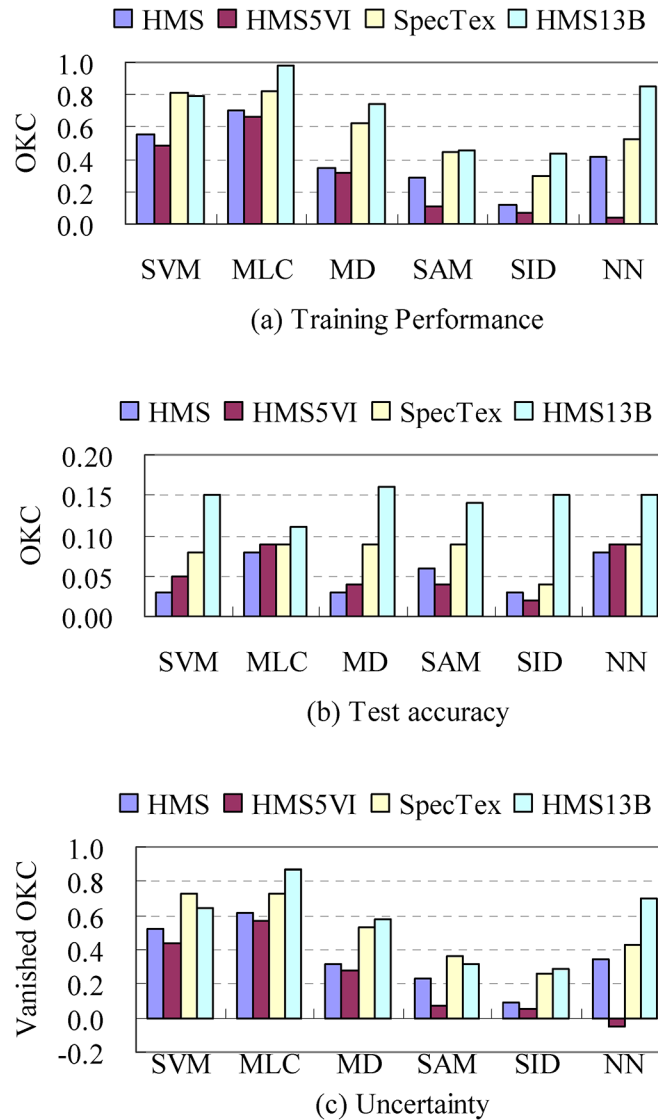
**Fig 6. A magnified view of the images (a) and classifications result of tree species (b).** Below each image, the prefixes “r-” and “e-” before the code of tree species denote training and test samples respectively. The white polygon in an image depicts the training samples or test samples of that specific species.

doi:10.1371/journal.pone.0125554.g006

graded into 3 groups by Duncan’s test, the AIE values were mostly insignificant in relation to each other. But the MLC and SVM, with AIE 3.83 and 3.04, respectively, were significantly higher than the SID. In brief, the AIE statistical test suggested the HMS13B image when combined with the MLC or SVM classifier was able to achieve improved multiple species classification results.

### Recommendations for further study

The classifiers MLC, SVM, SAM, MD, SID, and NN have been intensively applied in LULC and forest classifications. In this study, excellent training performance was obtained by



**Fig 7. A comparison of the training performance, test accuracy, and uncertainty among classifiers in variant classification schemes.**

doi:10.1371/journal.pone.0125554.g007

integrating the high spatial resolution multispectral image, and their derived vegetation indices and texture images. Unfortunately, no improvement in accuracy could be achieved using the test samples. This indicates that high variation of species spectra and wide-spread forest are major factors in performance. In addition, a possible variation of phenological events and/or tree health might cause changes in individual trees of a particular species leading the species classification into a more complicated situation. In these cases, additional spectral signatures and nonparametric methods might be helpful. This may be worthy of further study in the future.

In tropical and sub-tropical forest, many species compete for dominance in the overstorey canopy. Changes in the composition of the overstorey canopy might influence the extent of biodiversity, and the structure of the forest community, and succession. As Chytrý et al. [60]

indicated, vegetation surveys that define vegetation types and help understand differences among them are essential for basic ecological research and applications in biodiversity conservation and environmental monitoring. To improve our understanding of forest data, particularly, tree species composition, efforts in using remote sensing imagery to effectively expand more heterogeneous spectra and textural information could be explored further. Imagery with high spatial resolution and hyperspectral signatures in the visible-short wave infrared region would also be highly desirable.

WorldView-2 is capable of collecting bands of Coastal, Blue, Green, Yellow, Red, Red Edge, Near-IR1, and Near-IR2. WorldView-3 is able to collect an additional 8 SWIR (Short-Wave Infrared) bands and 12 CAVIS (Clouds, Aerosols, Vapors, Ice, and Snow) bands. These two HSR multispectral satellites offer more diverse spectral information than Quickbird-2 multispectral image. Therefore, a variety of additional vegetation indices could be derived to explore extra spectral information to diagnose or describe properties of tree species. For example, a new spectral feature (effective chlorophyll index, ECI) developed by Lin et al. [58] is an index that derived by blue, red, and red edge spectra was approved to be able to accurately predict total chlorophyll content of foliage and therefore can help to identify the variations of greenness in tree crown or canopy. In other words, by taking of the availability of diverse spectral bands at variable wavelengths highly accurate mapping of tree species using satellite/airborne remote sensing images is achievable.

Although, object-based classifiers have been successfully applied in LULC and forest mapping recently [54, 61, 62], there still a big challenge to overcome the high complexity of tree-species spectra to reduce classification uncertainty or accomplish a classification results with great agreement of both training performance and test accuracy. Tree spectral reflectance varies in respect to factors such as crown structure, phenology [61], and physiological and disease stress [63] which vary with time as well. If an object-based methodology can integrates spectral information with crown-shape parameters, it will greatly assist the process of tree mapping and more accurately fit the needs of forest ecosystem management. This may provide a fruitful direction for future studies.

## Conclusions

There has been a lot of interest recently in species classification using remote sensing techniques, but few studies have attempted to classify a large number of tree species using satellite images. This paper introduced a novel remote-sensing-based method for the classification of 40 tree species in subtropical forest using high-spatial-resolution QuickBird multispectral images. Classification performance and accuracy was measured using OKC and SCKC. Several interesting findings emerged from the study and some potential issues of concern were raised. Suggestions are offered below:

- The atmospherically corrected high-spatial-resolution QuickBird multispectral image is not able to identify 40 tree species simultaneously. Although the vegetation indices NDVI, SAVI, EVI, ARVI, and VARI help to supplement the regular multiple spectra with additional information, without the original spectral signatures, no improvement could be gained in classification.
- Texture transformation is also capable of retrieving additional spectral information in respect to the spectra of a multispectral image. Using a dimensionality expansion technique to integrate multiple images is helpful to gather diverse spectral information and for deriving tree species signatures. However, this classification scheme still has problems with tree classification outside the training samples and can not achieve a general level of accuracy, for example

an overall kappa coefficient of 0.80 or better in land use and land cover (LULC) classification.

- Ground survey observations revealed some substantial differences among individual samples of the same tree species such as the leaf density of tree crown, leaf color, and the surrounding trees. Leaf color and the density of the tree crown are major factors regulating tree physiology and may influence the spectral absorption and reflection features. In addition, low foliage density creates gaps in the tree crown or canopy and also increases the amount of light reaching the ground. The effect increases the influence of the background material. As a result, there is a much greater variation in spectral reflectance that may contribute to misclassification. This issue presents a significant challenge to forest species mapping using remote sensing which we intend to investigate further in future studies.
- Integrating high-spatial-resolution canopy height model data and satellite images with diverse spectral information could further enrich the spectral signatures and spatial texture data specific to individual tree species and therefore obtain a higher potential accuracy rate in overstorey-canopy species mapping.

## Supporting Information

**S1 Table. The training-sample-based species conditional kappa coefficient (SCKC) of each species for the MLC classification using variant data sets<sup>#</sup>.**

(DOC)

**S2 Table. The test-sample-based species conditional kappa coefficient (SCKC) of each species for the MLC classification using variant data sets.**

(DOC)

**S3 Table. ANOVA test of the accuracy improvement efficiency (AIE) of tree species classification.**

(DOC)

**S4 Table. Duncan's new multiple range method determined grouping for the average AIE of the dataset used in tree species classification.**

(DOC)

**S5 Table. Duncan's new multiple range method determined grouping for the average AIE of the classifiers in tree species classification.**

(DOC)

## Author Contributions

Conceived and designed the experiments: CL SP. Performed the experiments: CL KT. Analyzed the data: CL SP CIC GT KT. Contributed reagents/materials/analysis tools: CL SP CIC GT KT. Wrote the paper: CL GT SP CIC KT.

## References

1. Gougeon FA (1995) A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing* 21: 274–284.
2. Brandtberg T, Walter F (1998) Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple scale analysis, *Machine Vision and Application* 11: 64–73.

3. Popescu SC, Wynne RH (2004) Seeing the trees in the forest: using LIDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing* 70: 589–604.
4. Lin C, Thomson G, Lo CS, Yang MS (2011) A multi-level morphological active contour algorithm for delineating tree crowns in mountainous forest. *Photogrammetric Engineering and Remote Sensing* 77: 241–249.
5. Lin C, Lo CS, Thomson G (2011) A textural modification of the MMAC algorithm for individual tree delineation in forest stand using aerial bitmap images. *Proceeding of the 4th International Congress on Image Signal Processing (CISP 2011)* 3: 1633–1637.
6. Lin C, Lo CS, Thomson G (2011) Estimating individual tree characteristics using the MMAC algorithm and a lidar-derived canopy height model. *Journal of Earth Science and Engineering* 1: 35–41.
7. Lo CS, Lin C (2013) Growth-competition-based stem diameter and volume modeling for tree-level forest inventory using airborne LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 51: 2216–2226.
8. Gairola S, Proches S, Rocchini D (2013) High resolution satellite remote sensing: a new frontier for biodiversity exploration in Indian Himalayan forests. *International Journal of Remote Sensing* 34: 2006–2022.
9. Ishihama F, Watabe Y, Oguma H (2012) Validation of a high-resolution, remotely operated aerial remote-sensing system for the identification of herbaceous plant species. *Applied Vegetation Science* 15: 383–389. PMID: [23009627](#)
10. Lin C, Thomson G, Hung SH, Lin YD (2012) A GIS-based protocol for the simulation and evaluation of realistic 3-D thinning scenarios in recreational forest management. *Journal of Environmental Management* 113: 440–446. doi: [10.1016/j.jenvman.2012.09.001](#) PMID: [23062273](#)
11. Franklin SE, Hall RJ, Moskal LM, Maudie AJ, Lavigne MB (2000) Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing* 21: 61–79.
12. Leckie DG, Tinis S, Nelson T, Burnett C, Gougeon FA et al. (2005) Issues in species classification of trees in old growth conifer stands. *Canadian Journal of Remote Sensing* 31: 175–190.
13. Hill RA, Wilson AK, George M, Hinsley SA (2010) Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science* 13: 86–99.
14. Feilhauer H, He KS, Rocchini D (2012) Modeling species distribution using niche-based proxies derived from composite bioclimatic variables and MODIS NDVI. *Remote Sensing* 4: 2057–2075.
15. Hill RA, Thomson AG (2005) Mapping woodland species composition and structure using airborne spectral and lidar data. *International Journal of Remote Sensing* 26: 3763–3779.
16. Holmgren J, Persson , Söderman U (2008) Species identification of individual trees by combining high resolution lidar data with multi-spectral images. *International Journal of Remote Sensing* 29: 1537–1552.
17. Reitberger J, Krzystek P, Stilla U (2008) Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing* 29: 1407–1431.
18. Ke Y, Quackenbush LJ, Im J (200) Remote sensing of environment synergistic use of QuickBird multi-spectral imagery and lidar data for object-based forest species classification. *Remote Sensing of Environment* 114: 1141–1154.
19. Boschetti M, Boschetti L, Oliveri S, Casati L, Canova I (2007) Tree species mapping with airborne hyperspectral MIVIS data: the Ticino Park study case. *International Journal of Remote Sensing* 28: 1251–1261.
20. Pu R (2009) Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing* 30: 2759–2779.
21. Wang L, Sousa WP (2009) Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance. *International Journal of Remote Sensing* 30: 1267–1281.
22. Ghiyamat A, Shafri HZM (2010) A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing* 31: 1837–1856.
23. Cochrane MA (2010) Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing* 21: 2075–2087.
24. Jones TG, Coops NC, Sharma T (2010) Employing ground-based spectroscopy for tree-species differentiation in the Gulf Islands National Park Reserve. *International Journal of Remote Sensing* 31: 1121–1127.
25. Rocchini D, Ricotta C, Chiarucci A (2007) Using satellite imagery to assess plant species richness: The role of multispectral systems. *Applied Vegetation Science* 10: 325–331.

26. Hoare D, Frost P (2004) Phenological description of natural vegetation in southern Africa using remotely-sensed vegetation data. *Applied Vegetation Science* 7: 19–28.
27. Hernández-Stefanoni JL, Dupuy JM, Castillo-Santiago MA (2009) Assessing species density and abundance of tropical trees from remotely sensed data and geostatistics. *Applied Vegetation Science* 12: 398–414.
28. McDanold RI, Halpin PN (2007) Urban D.L. Monitoring succession from space: A case study from the North Carolina Piedmont. *Applied Vegetation Science* 10: 193–203.
29. Hantson W, Kooistra L, Slim PA (2012) Mapping invasive woody species in coastal dunes in the Netherlands: a remote sensing approach using LIDAR and high-resolution aerial photographs. *Applied Vegetation Science* 15: 536–547.
30. Ouyang ZT, Gao Y, Xie X, Guo HQ, Zhang TT, Zhao B (2013) Spectral Discrimination of the Invasive Plant *Spartina alterniflora* at Multiple Phenological Stages in a Saltmarsh Wetland. *PLoS ONE* 8(6): e67315. doi: [10.1371/journal.pone.0067315](https://doi.org/10.1371/journal.pone.0067315) PMID: [23826265](https://pubmed.ncbi.nlm.nih.gov/23826265/)
31. Li CF, Chytry M, Zeleny D, Chen TY, Hsieh CF (2012) Classification of Taiwan forest vegetation. Reference # 6448087. Symposium on The East Asian Flora and Its Role in the Formation of the World's Vegetation. Valadivostok, Russia. September 23–27.
32. Krause K (2005) Radiometric Use of QuickBird Imagery. Technical Note. Available: [http://www.digitalglobe.com/downloads/products/QuickBird\\_technote\\_raduse\\_v1.pdf](http://www.digitalglobe.com/downloads/products/QuickBird_technote_raduse_v1.pdf).
33. Chavez PS Jr, Sides SC, Anderson JA (1991) Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and Remote Sensing* 57: 295–303.
34. Lin C, Huang CP, Ker CC, Chang CI (2005) Open Land Individual Tree Crown Identification and Delineation using High Resolution Satellite Imagery. *Quarterly Journal of Chinese Forestry* 38: 163–178.
35. Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8:127–150.
36. Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25: 295–309.
37. Huete AR, Liu HQ, Batchily K, Van Leeuwen WJ (1997) A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment* 59: 440–451.
38. Kaufman YJ, Tanré D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing* 30: 261–270.
39. Gitelson AA, Kaufman YJ, Stark R, Rundquist D (2002) Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* 80: 76–87.
40. Viña A, Gitelson AA, Rundquist DC, Keydan G, Leavitt B, Schepers J (2004) Monitoring Maize (*Zea mays* L.) Phenology with Remote Sensing. *Agronomy Journal* 96: 1139–1147.
41. Huete AR, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83:195–213.
42. Larcher W (2003) *Physiological plant ecology—Ecophysiology and stress physiology of functional group*. 4th ed. Springer, New York, USA.
43. Kang J, Zhang W (2008) Quickbird Remote Sensing Image Denoising Using Wavelet Packet Transform. *Proceeding of the Second International Symposium on Intelligent Information Technology Application* 3: 315–318.
44. Lee JS (1980) Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2*, 165–168.
45. Gaikwad RN, Sharma JR, Jayaraman V (2006) Qualitative and quantitative assessment of high resolution multispectral image using fusion technique by measuring entropy—a case study using QuickBird dataset. *Proceedings of Symposium of ISPRS Commission IV. ISPRS Archives—Volume XXXVI Part 4*, 3 p.
46. Lin C (2003) Study on the application of digital image of aerially color photographs in the paddy field patches and area survey. *Journal of National Chiayi University* 75: 181–198.
47. Schowengerdt RA (1997) *Remote sensing-models and methods for image processing*. 2nd ed. Academic Press, NY, USA.
48. Richards J.A., 1999, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, p.240.
49. Allouche O, Tsoar A, Kadmon R (2006) Test the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223–1232.
50. Vach W (2005) The dependence of Cohen's kappa on the prevalence does not matter. *Journal of Clinical Epidemiology* 58: 655–661. PMID: [15939215](https://pubmed.ncbi.nlm.nih.gov/15939215/)

51. Stehman SV (1996) Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing* 62: 401–407.
52. Kruse FA, Lefkoff AB, Boardman JB, Heidebrecht KB, Shapiro AT, Barloon PJ, et al. (1993) The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* 44: 145–163.
53. Wu TF, Lin CJ, Weng RC (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research* 5: 975–1005.
54. Du Y, Chang CI, Ren H, Chang CC, Jensen JO, D'Amico FM, et al. (2004) New hyperspectral discrimination measure for spectral characterization. *Optical Engineering* 43(8): 1777–1786.
55. Jensen JR (2004) *Introductory digital image processing: a remote sensing perspective*. 3rd ed. Upper Saddle River, Pearson Prentice Hall, New Jersey, USA.
56. Lin C (1999) Study of seasonal changes of spectral reflectance of forest vegetation. *Taiwan Journal of Forest Science* 14: 289–305.
57. Dwivedi RS, Kothapalli RV, Singh AN (2009) Generation of farm-level information on salt-affected soils using IKONOS-II multispectral data, in: Metternicht G, Zinck JA (Eds.), *Remote Sensing of Soil Salinization—Impact on Land Management*. CRC Press, Florida, pp 73–88. doi: [10.2134/jeq2009.0127](https://doi.org/10.2134/jeq2009.0127) PMID: [25269314](https://pubmed.ncbi.nlm.nih.gov/25269314/)
58. Lin C, Popescu SC, Huang SC, Chang PT, Wen HL (2015) A novel reflectance-based model for evaluating chlorophyll concentration of fresh and water-stressed leaves. *Biogeosciences* 12: 49–66, doi: [10.5194/bg-12-49-2015](https://doi.org/10.5194/bg-12-49-2015), [2015](https://doi.org/10.5194/bg-12-49-2015)
59. Frampton CM, Pekelharing CJ, Payton IJ (2001) A fast method for monitoring foliage density in single lower-canopy trees. *Environmental Monitoring and Assessment* 72: 227–234. PMID: [11720225](https://pubmed.ncbi.nlm.nih.gov/11720225/)
60. Chytrý M, Schaminée JHJ, Schwabe A (2011) Vegetation survey: a new focus for Applied Vegetation Science. *Applied Vegetation Science* 14: 435–439.
61. Lin C, Wu CC, Tsogt K, Ouyang YC, Chang CI (2015) Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery. *Information Processing in Agriculture*, doi: [10.1016/j.inpa.2015.01.003](https://doi.org/10.1016/j.inpa.2015.01.003), In Press.
62. Duro DC, Franklin SE, Dubé MG (2012) A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118: 259–272.
63. Anderson DA, Armstrong RA, Weil E (2013) Hyperspectral Sensing of Disease Stress in the Caribbean Reef-Building Coral, *Orbicella faveolata*—Perspectives for the Field of Coral Disease Monitoring. *PLoS ONE* 8(12): e81478. doi: [10.1371/journal.pone.0081478](https://doi.org/10.1371/journal.pone.0081478) PMID: [24324697](https://pubmed.ncbi.nlm.nih.gov/24324697/)