ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 612–620

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Spectral discrimination of papyrus vegetation (*Cyperus papyrus L.*) in swamp wetlands using field spectrometry

Elhadi Adam *, Onisimo Mutanga

University of KwaZulu-Natal, Discipline of Geography, P. Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

ARTICLE INFO

Article history:
Received 28 August 2008
Received in revised form
7 March 2009
Accepted 23 April 2009
Available online 5 June 2009

Keywords:
Papyrus
Greater St. Lucia Wetlands Park
Field spectrometer measurements
CART
Jeffries-Matusita

ABSTRACT

Techniques for mapping and monitoring wetland species are critical for their sustainable management. Papyrus (*Cyperus papyrus L.*) is one of the most important species-rich habitats that characterize the Greater St. Lucia Wetlands Park (GSWP) in South Africa. This paper investigates whether papyrus could be discriminated from its co-existing species using ASD field spectrometer data ranging from 300 nm to 2500 nm, yielding a total of 2151 bands. Canopy spectral measurements from papyrus and three other species were collected in situ in the Greater St. Lucia Wetlands Park, South Africa. A new hierarchical method based on three integrated analysis levels was proposed and implemented to spectrally discriminate papyrus from other species as well as to reduce and subsequently select optimal bands for the potential discrimination of papyrus. In the first level of the analysis using ANOVA, we found that there were statistically significant differences in spectral reflectance between papyrus and other species on 412 wavelengths located in different portions of the electromagnetic spectrum. Using the selected 412 bands, we further investigated the use of classification and regression trees (CART) in the second level of analysis to identify the most sensitive bands for spectral discrimination. This analysis yielded eight bands which are considered to be practical for upscaling to airborne or space borne sensors for mapping papyrus vegetation. The final sensitivity analysis level involved the application of Jeffries-Matusita (JM) distance to assess the relative importance of the selected eight bands in discriminating papyrus from other species. The results indicate that the best discrimination of papyrus from its co-existing species is possible with six bands located in the red-edge and near-infrared regions of the electromagnetic spectrum. Overall, the study concluded that spectral reflectance of papyrus and its co-existing species is statistically different, a promising result for the use of airborne and satellite sensors for mapping papyrus. The three-step hierarchical approach employed in this study could systematically reduce the dimensionality of bands to manageable levels, a move towards operational implementation with band specific sensors.

1. Introduction

Papyrus swamps (*Cyperus papyrus L.*) characterize most wetland areas of eastern and central tropical Africa (Bemigisha, 2004). Specifically, the swamp covers great areas in Uganda and Sudan around the Lake Victoria and Nile basins (Beadle, 1974). Other extensive areas are in the Upemba basin, Zaire, and the Okavango Delta, Botswana (Thompson et al., 1979). Papyrus swamps usually create a buffer zone between terrestrial and aquatic ecosystems and play hydrological, ecological, and economic roles in the aquatic systems (Gaudet, 1980; Mafabi, 2000).

Previous studies found that tropical papyrus swamps are characterized by a tremendous amount of combined nitrogen (Mwaura and Widdowson, 1992; Muthuri and Kinyamario, 1989) and a high rate of biomass production (Muthuri and Kinyamario, 1989). In this regard, papyrus plays a vital role in hosting habitats for wildlife species such as the sitatunga antelope (*Tragelaphus spekei*) and African python (*Python sebae*) (Owino and Ryan, 2007). Papyrus also has some grazing potential and could be used as fodder with high nutritive value especially in the dry season when other forage is limited (Muthuri and Kinyamario, 1989). Further, studies found that the highest species richness of birds in marshland was associated with the areas where papyrus and natural vegetation were plentiful (Harper, 1992; Owino and Ryan, 2007). In addition to providing habitat for wildlife, the high biomass production characterising papyrus swamps has seen it being widely used for paper making. The Egyptians for example, were the first people who used papyrus to make paper more than five thousand years ago (Bucci, 2001). Recently, promising results have been obtained in using papyrus as an alternative source of fuel

* Corresponding author. Tel.: +27 332605779; fax: +27 332605344.
E-mail address: 205527619@ukzn.ac.za (E. Adam).

0924-2716/$ – see front matter © 2009 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2009.04.004

in many countries in central Africa such as Rwanda (Jones, 1983; Muthuri and Kinyamario, 1989).

Despite its relative importance, human encroachment and intensified agricultural activities in many parts of Africa have threatened the existence of papyrus (Mafabi, 2000; Owino and Ryan, 2007; Maclean et al., 2006). The continued degradation in papyrus habitat represents a significant threat to biodiversity conservation particularly for papyrus-specialist birds and other papyrus-reliant species in many African countries (Owino and Ryan, 2007; Maclean et al., 2006).

To establish sustainable management of such important species, up-to-date spatial information about the magnitude and distribution of papyrus at several scales is essentially required (Schmidt and Skidmore, 2003; Nagendra, 2001). This can be achieved through remote sensing techniques that can monitor the change in papyrus areas and assessing the species' percentage cover as compared to the other species.

Traditionally, species discrimination for floristic mapping needs intensive field work, including taxonomical information and the visual estimation of percentage cover for each species which are costly and time consuming and sometimes inapplicable due to the poor accessibility (Kent and Coker, 1992; Lee and Lunetta, 1995). Remote sensing, on the other hand, is a technique that gathers data regularly about the earth's features without actually being in direct contact with those features. The main advantages that make remote sensing preferable to field-based methods in land-cover classification, are that it has repeat coverage which allows continuous monitoring, and its digital data can be easily integrated into a geographic information system (GIS) for more analysis which is less costly and less time-consuming (Shaikh et al., 2001; Ozesmi and Bauer, 2002; Mironga, 2004; Schmidt and Skidmore, 2003).

Both multispectral and hyperspectral remote sensing techniques have been used to discriminate and map wetland species. Multispectral data such as Landsat TM and SPOT imagery have been used to identify general vegetation classes or to attempt to discriminate just broad vegetation communities (May et al., 1997; Harvey and Hill, 2001; Li et al., 2005), while hyperspectral data have been successful in mapping wetland vegetation at the species level (Schmidt and Skidmore, 2003; Belluco et al., 2006; Brown, 2004; Rosso et al., 2005; Pengra et al., 2007; Vaiphasa et al., 2005; Enrica et al., 2006). Hyperspectral data has also been used to study vegetation health, water content in vegetation, biomass, and other physico-chemical properties (e.g. Mutanga and Skidmore, 2004; Mutanga et al., 2003; Green et al., 1998; Zarco-Tejada et al., 2005).

In general, the use of multi-spectral data in discriminating and mapping wetlands species is challenging due to spectral overlap between the wetlands species and due to the lack of spectral and spatial resolution of the multi-spectral data (Rosso et al., 2005). On the other hand, hyperspectral data often consist of over 100 contiguous bands of 10 nm or less bandwidth. These contiguous bands and narrow ranges lead to the possibility of discriminating and mapping vegetation species more accurately and precisely than the standard multispectral bands (Ustin et al., 2004; Schmidt and Skidmore, 2003; Borges et al., 2007).

A few previous attempts at using multispectral remote sensing in studies of papyrus swamps have been concentrated mainly on economic benefit and management scenarios of papyrus swamps and promising results have been obtained (Bemigisha, 2004; Owino and Ryan, 2007). However, the spectral discrimination of papyrus (*Cyperus papyrus L.*) has been overlooked in scientific research. No attempt, to our knowledge, has been made to discriminate papyrus swamps using field spectrometry, let alone in South Africa where only a handful of studies have used hyperspectral data to characterize vegetation in general due to high cost and poor accessibility (Ismail et al., 2007; Mutanga et al., 2004).

Although hyperspectral data are critical in discriminating species, its high spectral resolution contains redundant information at band level (Bajwa et al., 2004; Kokaly et al., 2003). This high-dimensional complexity of hyperspectral data can be problematic in terms of image processing algorithms, an excessive demand for sufficient field samples, high cost, and overfitting when using multivariate statistical techniques (Borges et al., 2007; Vaiphasa et al., 2007; Bajcsy and Groves, 2004; Goetz, 1991) Mutanga and Kumar (2007). It is therefore imperative to identify the optimal bands required for discriminating and mapping wetland species without losing any important information. Different univariate and multivariate techniques for dimensionality reduction and band selection with different performance levels have been developed, such as canonical analysis, classification and regression tree (CART), discriminant analysis, principal component analysis, artificial neural network and Jeffries-Matusita (JM) (Vaiphasa et al., 2005; Milton et al., in press; Cochrane, 2000; Schmidt and Skidmore, 2003; Satterwhite and Ponder Henley, 1987). However, inconsistent results have been obtained for different species and environments and the use of a single technique in reducing data dimensionality to acceptable operational levels has not been very successful.

This study aimed to investigate whether field spectrometry data could be used effectively to discriminate papyrus swamps from other species occurring in the swampy wetlands of South Africa. In other words, spectral separability analysis was used to examine whether papyrus swamps could spectrally be discriminated from the other species using field spectrometer measurements at canopy level as well as reducing spectral data dimensionality. More specifically, the objectives of this study were: (1) to determine whether there is a significant difference between the mean reflectance at each measured wavelength (from 350 to 2500 nm) for *Cyperus papyrus L.* and the other co-existing three species (*Phragmites australis*, *Echinochloa pyramidalis*, and *Thelypteris interrupta*). (2) To identify key wavelengths that are most sensitive in discriminating *Cyperus papyrus* from the other three species. In order to achieve this, we used a field spectrometer to measure the spectral reflectance from papyrus swamps and the associated species in Greater St. Lucia Wetland Park in South Africa. To achieve an efficient optimal selection of bands, we propose a new hierarchical method that integrates analysis of variance (first level), classification regression trees (second level) and finally Jeffries-Matusita distance analysis (third level) to assess the relative importance of the selected bands.

2. Material and methods

2.1. Study area

The study area was the Greater St Lucia Wetlands Park (GSWP) which covers about 3,000,000 ha along the eastern coast of South Africa in the Province of KwaZulu-Natal, between longitudes 32° 21′ E and 32° 34′ E and latitudes 27° 34′ S and 28° 24′ S. The GSWP which includes the Futululu natural forest is considered to be the largest estuarine system in Africa (Taylor, 1995). The climate is sub-tropical. The mean annual rainfall varies from 1500 mm in the eastern shore to 700 mm in the western shore of the lake. It is characterized by a high diversity of ecosystems including, marine, inland lake, estuarine, forested dunes, mangrove, and coastal and swamp forest ecosystems. Therefore, it is recognized as both a UNESCO World Heritage Site and a Ramsar wetland of global significance. The papyrus (*Cyperus papyrus*) occurs with *P. australis*, *E. pyramidalis*, and *T. interrupta* in the large area between forested dunes and plantation forest. The area is either wet or flooded permanently with freshwater throughout the year.

Table 1
The papyrus swamp and its associated species, the number of sample plots and the total number of measurements collected.

Species name	Type code	No of plots	No of measurements
Cyperus papyrus	CP	15	134
Phragmites australis	PA	9	111
Echinochloa pyramidalis	EP	7	101
Thelypteris interrupta	TI	10	113

2.2. Field data collection

2.2.1. The identification of papyrus and its associated species

The most common plant species associated with papyrus swamps in the area were identified in the field in the summer of 2007 under the supervision of an experienced ecologist using field observation techniques. These species were then recorded based on their density and estimation of percentage cover (covering at least 40% of the area). In total three species were identified as being the most co-existent species with papyrus. These were *P. australis*, *E. pyramidalis*, and *T. interrupta* (Table 1).

2.2.2. Spectral data acquisition

The Analytical Spectral Devices (ASD) FieldSpec®3 spectrometer was used to measure the spectral reflectance from papyrus and the other species. This spectrometer has a wavelength ranging from 350 to 2500 nm with a sampling interval of 1.4 nm for the spectral region 350–1000 nm and, 2.0 nm for the spectral region 1000–2500 nm, and a spectral resolution of 3–10 nm (ASD, Analytical Spectral Devices, Inc., 2005).

A combination of random sampling and purposive sampling was used to select field sites. Hawth's Analysis Tool extension for ArcMap designed to perform spatial analysis, was used to generate random points in a land cover map developed using an ASTER image. These points were then input in GPS to navigate to the field sites. Purposive sampling was done when the random point was not accessible, or to increase the variation of reflectance measurements of the species. Once the sampling location was identified, a vegetation plot was defined to cover 3 by 3 m in area of each species; then a total of 10–15 field spectrometer measurements were taken randomly from the nadir at about 1.5 m and with a 5° field of view above the vegetation species on each plot. This resulted in a ground field of view of about 13 cm in diameter, which was large enough to cover a cluster of species and to reduce the effects of background such as soil and water in the in situ spectral measurement (Table 1). All the measurements were collected in December 2007 between 10:00 am and 02:00 pm under sunny and cloudless conditions. A white reference Spectralon calibration panel was used every 10–15 measurements to offset any change in the atmospheric condition and irradiance of the sun. Metadata such as the site description (coordinates and altitude, land cover class), and general weather conditions were also recorded accompanying field spectral measurements on each measured point (Milton et al., in press). Due to the atmospheric water absorption noise in the reflectance spectra, a number of bands around 1420 nm, 1940 nm, and 2400 nm were excluded from the analysis.

2.3. Data processing

It was difficult to use one technique to identify a reasonable number of wavelengths that are most sensitive from 350 to 2500 nm ($n = 2151$). This was because the dimensionality still remained high when one technique was used (412 wavelengths from analysis of variance). Moreover, there is no single technique that has universally proven to be superior for the optimal feature selection (Yang et al., 2005), and it is quite possible that more than one subset of wavelengths can discriminate the data equally well (Yeung et al., 2005). We therefore innovated a new hierarchical method for spectral analysis based on three integrated levels. In the first level, we used one-way ANOVA to test if the differences in the mean reflectance between papyrus swamps and the other three species were statistically significant. We tested the research hypothesis that the means of the reflectance between the pairs of papyrus swamp and each one of the co-existing species (PA, EP, and TI) were significantly different at each measured wavelength, from 350 to 2500 nm, viz. the null hypothesis Ho: $\mu 1 = \mu 2$, $\mu 1 = \mu 3$, $\mu 1 = \mu 4$ versus the alternate hypothesis Ha: $\mu 1 \neq \mu 2$, $\mu 1 \neq \mu 3$, $\mu 1 \neq \mu 4$ where: $\mu 1$, is the mean reflectance values from papyrus and $\mu 2$, $\mu 3$, $\mu 4$ the mean reflectance values from *P. australis*, *E. pyramidalis*, and *T. interrupta* respectively.

One-way ANOVA was used with a post-hoc Scheffé test at each measured wavelength for the individual class pair (CP vs. PA, CP vs. EP, and CP vs. TI). We tested ANOVA with two confidence levels: a 99% confidence level ($p < 0.01$), and a 95% confidence level ($p < 0.05$).

2.3.1. Classification and regression trees (CART)

We used CART in this second level of the hierarchical method to further reduce the number of significant wavelengths obtained from ANOVA analysis, with the purpose of reducing data dimensionality. CART, which was developed by Breiman et al. (1984), is a nonparametric statistical model that can select from a large dataset of explanatory variables (**x**) those that are best for the response variables (**y**) (Yang et al., 2003; Questier et al., 2005). CART was preferred in our study because the values of the predictor variables (spectral reflectance) are a continuous, as opposed to categorical target (plant species).

The CART model is built in accordance with the splitting rule. This rule performs the function of splitting the data into smaller parts according to the reduction of the deviance from the mean of the target variable (Y_{bar}) (or corrected total sum of the squares). (Y_i) is the target variable of each dataset. The decision tree begins a search from a root node (parent node) derived from all the predictors, and possible split points such that the reduction in deviance, D (total), is maximized (terminal node) as follows (Breiman et al., 1984).

$$D(\text{total}) = \sum (Y_i - Y_{bar})^2. \quad (1)$$

The cut point, or value, always splits the data into two child nodes, the left node and the right node with maximum homogeneity. The reduction in deviance is as shown in the following equation:

$$\Delta_{j,\text{total}} = D(\text{total}) - (D(L) + D(R)) \quad (2)$$

where $D(L)$ and $D(R)$ are the deviances of the left and right nodes.

Hence, the algorithm begins searching for the maximized $\Delta_{j,\text{total}}$ over all the predictor variables and the cut points subject to the constraint that the number of the members in the left and right nodes are larger than some criterion set by the user. The algorithm repeats the procedure of binary splitting for each node (left and right nodes) by treating each child node as a parent node splitting until the tree has a maximum size (Yang et al., 2003).

In this study, we used CART as the second level of the hierarchical method to select the most sensitive wavelengths from the number of significant wavelengths selected in the first level (ANOVA). Therefore, CART generated the optimal bands by selecting only the spectral bands that result in small misclassification rates to discriminate each class pair (CP vs. PA, CP vs. EP, and CP vs. TI) individually. The bands which were common in each class pair were then selected to get the optimal bands for all class pairs.

2.3.2. Distance analysis

After we had the optimal bands selected from the CART analysis, additional analysis was needed to identify the best band or band combinations that could be used for the best spectral separability between papyrus and each one of the three species. Hence, we tested the hypothesis that some bands are relatively more important than others in discriminating papyrus. The separability index used in this level of hierarchical method was JM distance analysis (Schmidt and Skidmore, 2003; Ismail et al., 2007; Vaiphasa et al., 2007). It was impossible to run JM distance analysis on all the significant bands ($n = 412$) from ANOVA analysis because of the singularity problem of matrix inversion (Vaiphasa et al., 2005; Ismail et al., 2007). Moreover, this high-dimensional complexity is very costly, time-consuming, and beyond the capacity of the common image processing algorithms (Schmidt and Skidmore, 2003; Vaiphasa et al., 2007; Borges et al., 2007). We therefore, used the bands derived from CART. The JM distance between a pair of probability functions is seen as a quantification of the mean distance between the two class density functions (Richards, 1993). When classes are normally distributed, this distance turns out to be the Bhattacharyya (BH) distance (Richards, 1993; Schmidt and Skidmore, 2003). The JM distance has upper and lower bounds that vary between 0 and $\sqrt{2}$ ($\approx$1.414), with the higher values indicating the total separability of the class pairs in the bands being used (Richards, 1993; ERDAS Field Guide, 2005). In this study we decided to use higher separability values $\geq$97% as a JM distance threshold to identify the most important band or band combinations for best discrimination of papyrus swamp. The formula for computing JM distance is as follows (ERDAS Field Guide, 2005):

$$\alpha = \frac{1}{8}\left(\mu_i - \mu_j\right)^{\mathrm{T}}\left(\frac{C_i + C_j}{2}\right)^{-1}\left(\mu_i - \mu_j\right) + \frac{1}{2}\ln\left(\frac{\left(\left|C_i + C_j/2\right|\right)}{\sqrt{|C_i| * \left|C_j\right|}}\right) \quad (3)$$

$$JM_{ij} = \sqrt{2\left(1 - \mathrm{e}^{\alpha}\right)} \quad (4)$$

where: i and j = the two classes being compared, C_i = the covariance matrix of signature i, μ_i = the mean vector of signature i, ln = the natural logarithm function, $|C_i|$ = the determinant of C_i (matrix algebra).

3. Results

3.1. First level: ANOVA test

ANOVA results indicate that there is no significant difference between the two class pairs (CP vs. EP, and CP vs. TI) when a 99% confidence level ($p < 0.01$) is used. However, the 95% confidence level ($p < 0.05$) indicates that there is a statistically significant difference in the spectral reflectance between all the class pairs (CP vs. PA, CP vs EP, and CP vs. TI) at $n = 412$ wavelengths. These significant wavelengths were highlighted using a histogram for every individual class pair. The results of the ANOVA test for each class pair (CP vs. PA, CP vs EP, and CP vs. TI) are shown in Fig. 1(a, b, and c). The shaded areas show the wavelengths where the spectral reflectance from the papyrus swamp is statistically different from the other three species, with a 95% confidence level (p-value < 0.05).

The conclusions from the ANOVA test are that the mean reflectance between papyrus and the other three species is significantly different in many measured wavelengths. These significant wavelengths are located in three different regions of the electromagnetic spectrum (red-edge, near-infrared, and mid-infrared).

Fig. 1. ANOVA results for each class pair (a) CP vs. PA, (b) CP vs. EP, and (c) CP vs. TI The grey areas show the wavebands where there are significant differences between the class pairs within the electromagnetic spectrum.

Table 2 shows the frequency of the significant bands adapted into the four spectral domains which is widely used in the hyperspectral remote sensing of vegetation (Kumar et al., 2001). The table shows that there are no statistically significant wavelengths located in the visible region for the class pairs CP vs. EP, and CP vs. TI. However, the class pair CP vs. PA has more significant wavelengths located all over the spectral regions than any other class pair (CP vs. EP, and CP vs. TI). All the wavelengths from 350 to 1300 ($n = 950$) are significant for CP vs. PA as well as 49.95% ($n = 600$) of wavelengths located in the mid-infrared region, whereas the statistically significant wavelengths for the pair CP vs. TI are located only in the red-edge and near-infrared portions of the electromagnetic spectrum ($n = 449$). It can also be seen from Table 2 that the red-edge and near-infrared are the most important regions where each class pair has the most statistically significant wavelengths.