International Journal of Applied Earth Observation and Geoinformation 13 (2011) 228–235

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Evaluation of classifiers for processing Hyperion (EO-1) data of tropical vegetation

Dhaval Vyas[a], N.S.R. Krishnayya[a,*], K.R. Manjunath[b], S.S. Ray[b], Sushma Panigrahy[b]

[a] Ecology Lab, Department of Botany, Faculty of Science, The M.S. University of Baroda, Baroda 390 002, India
[b] Space Applications Centre, ISRO, Ahmedabad 380 015, India

ARTICLE INFO

Article history:
Received 7 May 2010
Accepted 25 November 2010

Keywords:
Hyperion (EO-1) data
Tropical forest
Band selection
Artificial Neural Network
Spectral Angle Mapper
Support Vector Machine

ABSTRACT

There is an urgent necessity to monitor changes in the natural surface features of earth. Compared to broadband multispectral data, hyperspectral data provides a better option with high spectral resolution. Classification of vegetation with the use of hyperspectral remote sensing generates a classical problem of high dimensional inputs. Complexity gets compounded as we move from airborne hyperspectral to Spaceborne technology. It is unclear how different classification algorithms will perform on a complex scene of tropical forests collected by spaceborne hyperspectral sensor. The present study was carried out to evaluate the performance of three different classifiers (Artificial Neural Network, Spectral Angle Mapper, Support Vector Machine) over highly diverse tropical forest vegetation utilizing hyperspectral (EO-1) data. Appropriate band selection was done by Stepwise Discriminant Analysis. The Stepwise Discriminant Analysis resulted in identifying 22 best bands to discriminate the eight identified tropical vegetation classes. Maximum numbers of bands came from SWIR region. ANN classifier gave highest OAA values of 81% with the help of 22 selected bands from SDA. The image classified with the help SVM showed OAA of 71%, whereas the SAM showed the lowest OAA of 66%. All the three classifiers were also tested to check their efficiency in classifying spectra coming from 165 processed bands. SVM showed highest OAA of 80%. Classified subset images coming from ANN (from 22 bands) and SVM (from 165 bands) are quite similar in showing the distribution of eight vegetation classes. Both the images appeared close to the actual distribution of vegetation seen in the study area. OAA levels obtained in this study by ANN and SVM classifiers identify the suitability of these classifiers for tropical vegetation discrimination.

1. Introduction

As we strive to understand the impact of human activities on our planet, concerns over global land use and land cover change are rising (Okin et al., 2001). For decades remote sensing has been used as a tool to generate land use/land cover maps (Chan and Palinckx, 2008). Many studies have been carried out to generate land cover maps of various forest ecosystems with the help of broadband multispectral data (Carpenter et al., 1999; Knorn et al., 2009; Rogan et al., 2002; Sedano et al., 2005; Wang et al., 2004). Landsat TM imagery is predominantly used in the classification studies of forest growth stages (Helmer et al., 2000; Nelson et al., 2000). Arroyo-Mora (2002) has studied different successional stages of dry deciduous forest with the help of combined Landsat TM and IKONOS data set. Asner et al. (2009) made an attempt to monitor forest degradation and deforestation over different types of tropical forests with the help of multispectral sensors. However, conventional broadband multispectral data proved to be less accurate because of coarse spectral resolution (Ellis et al., 2006). A better option would be hyperspectral remote sensing with high spectral resolution (hundreds of contiguous narrow bands). High spectral resolution of hyperspectral sensors is useful in sorting out finer differences among traditional land cover classes (Turner et al., 2003). Hyperspectral remote sensing techniques have great potential for characterizing the diversity and richness of tropical environment (Sánchez-Azofeifa et al., 2003). Ustin et al. (2004) have affirmed that hyperspectral remote sensing has the potential to separate vegetation into taxonomic levels with higher accuracies. Air borne hyperspectral data have been applied to tree species identification in various ecosystems such as temperate forest stand types (Martin et al., 1998; Melgani and Bruzzone, 2004; Gong et al., 1997; Plaza et al., 2009), boreal forest trees discrimination (Van Aardt and Wynne, 2001).

Most of the hyperspectral studies were performed on temperate forests (Buddenbaum et al., 2005; Camps-Valls and Bruzzone, 2005; Martin et al., 1998; Melgani and Bruzzone, 2004). Few

* Corresponding author.
E-mail address: krish14@gmail.com (N.S.R. Krishnayya).

0303-2434/$ – see front matter
doi:10.1016/j.jag.2010.11.007

researchers (Clark et al., 2005; Sanchez-Azofeifa and Castro, 2006; Zhang et al., 2006) made an attempt to map tropical trees by using high spatial and spectral imagery. Thenkabail et al. (2004) concluded that Hyperion data is better in classifying Land use pattern of African rain forests. Tropical forests are more heterogeneous than temperate ones. Tropical forests house most of the earth's terrestrial biodiversity and account for a large proportion of terrestrial carbon stores and photosynthesis (Clark and Clark, 2001). Locating individuals of a particular tree species will be a challenging task due to the high number of tree species per unit area in many tropical ecosystems (Sánchez-Azofeifa et al., 2003). Complexity gets compounded as we move from airborne hyperspectral to Spaceborne technology. It is unclear how different classification algorithms will perform on a complex scene of tropical forests collected by spaceborne hyperspectral data (EO-1, Hyperion). Main problems with hyperspectral image processing is the huge amount of data involved (high dimensionality). Powerful statistical methods are required to reduce dimensionality (Chan and Palinckx, 2008). Dimensionality reduction is mostly done by band selection. Feature extraction by band selection can minimize computational time for high dimensional hyperspectral data. Many methods such as principal component analysis (PCA) minimum noise fraction transform (MNF), discriminant analysis, decision boundary feature extraction (DBFE), spectral mixture analysis, Sequential Forward Floating (SFF), Steepest Ascent (SA) (Asner and Heidebrecht, 2002; Landgrebe, 2003; Melgani and Bruzzone, 2004; Myint, 2001; Neville et al., 2003; Okin et al., 2001; Platt and Goetz, 2004; Rashed et al., 2003; Thenkabail et al., 2004) have been used for dimensionality reduction. Each method works differently. PCA and MNF work on the principle of identification of uncorrelated bands, to segregate noise components, to reduce the dimensionality of data sets. Another approach proposed to reduce the dimensionality of feature space is by means of feature selection in techniques such as SFF, SA. The feature extraction approach (DBFE) addresses the problem of feature reduction by transforming the original feature space into a space of lower dimensionality, which contains most of original information (Melgani and Bruzzone, 2004). Thenkabail et al. (2004) employed Stepwise Discriminant Analysis (SDA) to find out the most optimal wavebands to discriminate different vegetation classes. Lucas et al. (2008) classified tree species with high accuracy (87%) by means of identified wavelengths from Stepwise Discriminant Analysis (SDA). Discriminant analysis categorizes samples using multivariate separability measures such as Wilks' lambda (Thenkabail, 2002). Though there are many other advanced methods for band selection (e.g. genetic algorithm, clustering based technique, clonal selection), SDA is comparatively a simpler method and hence is being used by many workers for optimum band identification (Jain et al., 2007; Lucas et al., 2008; Ray et al., 2010; Van Aardt and Wynne, 2001).

Identifying suitable classifiers for hyperspectral data is very important. Dalponte et al. (2009) have stated that classifiers with higher complexity are potentially more effective than the ones with smaller complexity, especially for difficult classification problems. In recent years, many advanced classification approaches were used for classification of vegetation, such as Artificial Neural Networks (Erbek et al., 2004; Foody, 2004; Kavzoglu and Mather, 2004), Decision tree classifier (Lawrence et al., 2004; Pal and Mather, 2003), Support Vector Machine classifier (Dalponte et al., 2009; Melgani and Bruzzone, 2004; Plaza et al., 2009), Random forest and Adaboost (Chan and Palinckx, 2008), linear discriminant analysis (Du and Ren, 2003; Clark et al., 2005), Spectral Angle Mapper (Christian and Krishnayya, 2009; Clark et al., 2005). Some of these classification algorithms fully exploit the huge amount of data provided by hyperspectral sensors. Among these, SAM, SVM and ANN are more widely used. The Spectral Angle Mapper (SAM) is a physically based spectral classification that uses an n-dimensional angle to match reflectance spectra coming from each pixel to reference spectra (Kruse et al., 1993). The algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra, treating them as vectors in a space with dimensionality equal to the number of bands (Shafri et al., 2007). The core advantage of SAM is, when used on calibrated reflectance data, it is relatively insensitive to illumination and albedo effects. Plaza et al. (2009) have successfully classified different tree species from urban area with the help of SVM classifier. The Support Vector Machine (SVM) is an effective distribution free classifier that has been widely used in the recent years for solving hyperspectral classification problems (Camps-Valls and Bruzzone, 2005). Melgani and Bruzzone (2004) pointed out the effectiveness of SVM to analyze hyperspectral data directly in the hyper dimensional feature space, without the need of any feature reduction procedure. They also mentioned about the advantage of SVM in classifying heterogeneous data (like the one of tropical system) for which only few training samples are available for each identified class. The ANN technique uses standard back propagation for supervised learning. This is the most widely used model and its design consists of one input layer, at least one hidden layer and one output layer (Shafri et al., 2007). This algorithm is a promising technique for a number of situations such as non-normality, complex feature spaces and multivariate data types (Atkinson and Tatnall, 1997). Complexity in structure, distribution and diversity of vegetation is common in tropics. It is extremely essential to monitor natural resources like forests in developing countries such as India as the rapid strides in the economy are putting tremendous pressure on existing tropical cover. Spaceborne technology is more appropriate for monitoring forest cover in India for its cost effectiveness. Classification of this data for a clear understanding is challenging. The present study addresses the following objective,

To compare the performance of three different classifiers (ANN, SAM, SVM) over highly diverse tropical forest vegetation.

2. Materials and methods

2.1. Study area

The study was conducted in tropical dry deciduous forests of Shoolpaneshwar Wildlife Sanctuary (SWS) located at 21°29′–21°52′N lat. and 73°29′–73°54′E long., Narmada District, Gujarat, India. The SWS is one of the important protected areas supporting sizeable biota. It occupies an area of 675 km^2. Topography of the area is undulating with both continuous and discontinuous hilly tracts up to an elevation of ~800 m intermingled with valleys, streams and sporadic clearings for agriculture. The area is intermixed with tribal settlements. Parts of the sanctuary area are cleared by tribals for agriculture. Annual rainfall of the area is in the range of 900–1200 mm. Rainfall is restricted to the months of June–October. Minimum and maximum annual mean temperatures are 8 °C and 42 °C respectively. Teak (*Tectona grandis* L.) and Bamboo (*Dendrocalamus strictus* Nees.) are the most dominant species of the study area. Other major tree species growing in the sanctuary are *Madhuca indica* Gmel. *Mangifera indica* L., *Pongamia pinnata* L. and *Ficus glomerata* L. Leaf shedding is complete in deciduous trees like *Tectona*, *Dendrocalamus* and *Madhuca*. In other trees like *Mangifera*, *Ficus* and *Pongamia* complete shedding is not seen. Because of this, these trees appear greener throughout. Other tropical tree species occurring in mixed patches in SWS are, *Butea monosperma* Lamk., *Holarrhena antidysenterica* R.Br, *Mitragyna parviflora* Korth., *Dalbergia latifolia* Roxb., *Anogeissus latifolia* Wall., *Bridelia retusa* L., *Albanian lebbeck* L., *Garuga pinnata* Roxb. Mixed patches also have *Tectona* and *Dendrocalamus*.