International Journal of Applied Earth Observation and Geoinformation 26 (2014) 49–63

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales

Aniruddha Ghosh[a], Fabian Ewald Fassnacht[b], P.K. Joshi[a,*], Barbara Koch[b]

[a] Department of Natural Resources, TERI University, Vasant Kunj, New Delhi, India
[b] Remote Sensing and Landscape Information Systems, University of Freiburg, Tennenbacherstraße. 4, D-79106 Freiburg, Germany

ARTICLE INFO

Article history:
Received 19 February 2013
Accepted 31 May 2013

Keywords:
Hyperspectral
nDSM
Support Vector Machines (SVM)
Random Forest (RF)
Scale
Tree species classification

ABSTRACT

Knowledge of tree species distribution is important worldwide for sustainable forest management and resource evaluation. The accuracy and information content of species maps produced using remote sensing images vary with scale, sensor (optical, microwave, LiDAR), classification algorithm, verification design and natural conditions like tree age, forest structure and density. Imaging spectroscopy reduces the inaccuracies making use of the detailed spectral response. However, the scale effect still has a strong influence and cannot be neglected. This study aims to bridge the knowledge gap in understanding the scale effect in imaging spectroscopy when moving from 4 to 30 m pixel size for tree species mapping, keeping in mind that most current and future hyperspectral satellite based sensors work with spatial resolution around 30 m or more.

Two airborne (HyMAP) and one spaceborne (Hyperion) imaging spectroscopy dataset with pixel sizes of 4, 8 and 30 m, respectively were available to examine the effect of scale over a central European forest. The forest under examination is a typical managed forest with relatively homogenous stands featuring mostly two canopy layers. Normalized digital surface model (nDSM) derived from LiDAR data was used additionally to examine the effect of height information in tree species mapping. Six different sets of predictor variables (reflectance value of all bands, selected components of a Minimum Noise Fraction (MNF), Vegetation Indices (VI) and each of these sets combined with LiDAR derived height) were explored at each scale. Supervised kernel based (Support Vector Machines) and ensemble based (Random Forest) machine learning algorithms were applied on the dataset to investigate the effect of the classifier. Iterative bootstrap-validation with 100 iterations was performed for classification model building and testing for all the trials.

For scale, analysis of overall classification accuracy and kappa values indicated that 8 m spatial resolution (reaching kappa values of over 0.83) slightly outperformed the results obtained from 4 m for the study area and five tree species under examination. The 30 m resolution Hyperion image produced sound results (kappa values of over 0.70), which in some areas of the test site were comparable with the higher spatial resolution imagery when qualitatively assessing the map outputs. Considering input predictor sets, MNF bands performed best at 4 and 8 m resolution. Optical bands were found to be best for 30 m spatial resolution. Classification with MNF as input predictors produced better visual appearance of tree species patches when compared with reference maps. Based on the analysis, it was concluded that there is no significant effect of height information on tree species classification accuracies for the present framework and study area. Furthermore, in the examined cases there was no single best choice among the two classifiers across scales and predictors. It can be concluded that tree species mapping from imaging spectroscopy for forest sites comparable to the one under investigation is possible with reliable accuracies not only from airborne but also from spaceborne imaging spectroscopy datasets.

1. Introduction

Information on spatial distribution of forest cover types/classes and species composition is a primary and foremost requirement for sustainable forest management (Franklin, 2001). It is also important for answering scientific research questions related to

* Corresponding author. Tel.: +91 11 26122222; fax: +91 11 26122874.
E-mail addresses: aniru123@gmail.com (A. Ghosh), fabian.fassnacht@felis.uni-freiburg.de (F.E. Fassnacht), pkjoshi27@hotmail.com, pkjoshi@teri.res.in (P.K. Joshi), barbara.koch@felis.uni-freiburg.de (B. Koch).

0303-2434/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2013.05.017

Table 1
Overview of studies using Hyperspectral data for tree species mapping.

Biome	Sensor	Classification model	Tree species	Study
Temperate broadleaf and mixed	AVIRIS	MLC[a]	3 Conifer and mixed deciduous	Martin et al. (1998)
Temperate rainforests	Hyperion	MLC	5 Conifer	Goodenough et al. (2003)
Tropical and subtropical moist broadleaf	HYDICE	LDA[b], MLC,SAM[c]	2 Evergreen and 5 deciduous	Clark et al. (2005)
Temperate broadleaf and mixed	AVIRIS	DA[d]	3 Types of pine	vanAardt and Wynne (2007)
Temperate Natural/Alpine forest	AISA Eagle	SVM[f], GML-LOOC[g], LDA	8 Commonly occurring species	Dalponte et al. (2009)
Mangrove	AISA+	MD[e], MLC, SAM	Mangrove species	Yang et al. (2009)
Tropical and subtropical savannas	CAO system	SAM	10 Types of savanna species	Cho et al. (2010)

[a] Maximum likelihood.
[b] Linear discriminant analysis.
[c] Spectral angle mapper.
[d] Discriminant analysis.
[e] Minimum distance/Mahalnobis distance.
[f] Support Vector Machine.
[g] Gaussian maximum likelihood classifier with leave-one-out-covariance estimator.

functioning of forested ecosystems (Linke et al., 2006). During the last few decades, both ground-based inventories and remote sensing approaches have been used with varying accuracy for the quantitative assessment. While the ground-based measurements have been criticized for requiring more time, manpower and economic resources (Mairs, 1976), information derived from remote sensing images have been promoted as providing the best alternative (Holmgren and Thuresson, 1998). Among many remote sensing systems, hyperspectral sensors, which allow continuous sampling of the electromagnetic spectrum from the visual to the shortwave infrared region, have been found to be more effective than broadband multispectral sensors for this type of application (Thenkabail et al., 2004; Ustin et al., 2004). Numerous studies have validated the success of hyperspectral data for tree species classification in different biomes (Table 1).

However, similarity of spectral signatures for different tree species, as well as assemblage of tree crowns with little to no inter crown distance and occurrence of over storey canopy, increases the challenges for successful tree species mapping from optical remote sensing data (Leckie et al., 2005). In some cases the described problems can be diminished by integrating canopy structural information from other remote sensing sources. LiDAR is a promising and widely used remote sensing technology which can be used to measure canopy structural information (Castillo et al., 2012; Donoghue et al., 2007; Lim et al., 2003). Several studies have demonstrated that simultaneous use of hyperspectral and LiDAR derived information is beneficial as these are complementary to each other (Table 2).

The review of the literature reveals that in the works tabulated above, classification accuracy has been dependent on factors such as spectral and spatial resolution, classification model, fusion framework and study area. However, the majority of these studies have been restricted to a single spatial resolution. Dalponte et al. (2009) addressed the role of spectral resolution on the classification accuracy in conjunction with three different classifiers for forested areas. Like spectral resolution, the impact of spatial measurement scale (spatial resolution/pixel size) on the information content and classification accuracy for forested ecosystems is well established (e.g., Marceau et al., 1994a,b). It is important to understand the difference between spatial resolution or pixel size with operational scale. Resolution, an inherent property of the sensor, refers to the size of the smallest object that can be identifiable in an image; whereas the operational scale refers to the scale at which physical features appear or operate. A single tree in a forest operates at a smaller scale than the group of trees or the forest. Pixel sizes can be different from the operational scales and a critical resolution refers to the resolution where these two become equal. Investigating the change in model performance (for example classification accuracy) with changing resolution can help to identify the critical resolution and to understand the scale dependence of the model (Bian, 2005; Cao and Lam, 2005). However, the optimal critical resolution found within such experiments is also always closely related to the ecosystem under consideration. Within a highly diverse tropical forest, tree species discrimination most likely has to be realized on a level in which individual trees are covered by more than a single pixel to reduce adjacency effects from neighboring trees of other species. In contrast, for managed temperate forests with fewer species and rather homogenous stands a coarser resolution may be more suitable since coarser resolution reduces the within class variability of the spectral signature which leads to an increased inter-class separability. Furthermore, adjacency effects are less severe in managed temperate forests since neighboring trees are often from the same species. So far, to the best of our knowledge, no study using hyperspectral data for forested ecosystems in moderate climate zones have investigated the performance of different classifiers across scale to make assumptions about the optimum spatial resolution and classifier for this kind of classification task.

Furthermore, Cho et al. (2012), Naidoo et al. (2012) and Asner et al. (2008) concluded tree height derived from LiDAR is an important variable in the mapping tree species in two completely different forest ecosystems. These studies essentially confirm that LiDAR-derived height information is one of the most important parameters for a multisensor framework involving hyperspectral and LiDAR data. Therefore along with the classifiers, it is also necessary to re-investigate the LiDAR derived height information for mapping tree species in a temperate forest and to understand how additional LiDAR information, which should mainly be correlated with tree age and silvicultural measures, alters classification results across spatial scales.

The overarching aim of this study is to produce accurate tree species maps and to examine the influence of spatial resolution on the derived maps. More precisely it addresses the following research questions:

1. Is there an optimum spatial resolution among the examined ones to map tree species in temperate forest ecosystems with hyperspectral data independent of classifiers and input variables?
2. Is there a single classification approach among the examined ones that produces the highest accuracies across scales in the selected forest ecosystem?
3. Is there any single desirable predictor layer to identify tree species independent of classifiers and scale?
4. How significant is LiDAR-derived height information for tree species classification in temperate forests? How does this significance vary across classifiers and scales?