Remote Sensing for Science, Education, and Natural and Cultural Heritage
Rainer Reuter (Editor)
EARSeL, 2010

Tree Species Classification Based on the Analysis of Hyperspectral Remote Sensing Data

Petra KRAHWINKLER, and Jürgen ROSSMANN
Ahornstraße 55, 52074 Aachen, Germany

Abstract. Current tree species classification algorithms operate on a stand-wise level and therefore high-resolution satellite data is sufficient. Local forest inventories at a small-scale level, as well as detailed biodiversity monitoring, need single tree based classification to meet their demands on granularity. The optimal choice of multispectral bands can highly improve classification results. In order to find suitable bands and to evaluate their impact on tree species classification a set of images, satellite data and hyperspectral data is analyzed and the result is tested in a simple decision tree approach to species classification

Keywords: Tree species classification, hyperspectral data, forest inventory, remote sensing, multispectral data.

Introduction

Sustainable forest management is an important strategy in both preserving biodiversity in forest areas and environmentally compatible forestry. Especially in habitat monitoring, detailed information on the forest and the existing tree species is needed. Field measurements are both expensive and time consuming.

Tree species classification algorithms are based on different data sets which make it hard to compare different approaches. Additional data, like hyperspectral sensors or multitemporal imaging, can be used to improve the classification. Multitemporal images were found in [1] to be less significant in improving classification results than additional spectral information.

The discrimination of 6 tree species and herbaceous vegetation was analyzed in [2]. However spectral variability within the same species was not taken into consideration. Instead only one sample per tree species was used for the analysis.

In [3] measurements were performed with a spectroradiometer from directly above the crown. One reading consisted of the average of 10 samples and several data points were collected per tree species. The variability within tree species was taken into account and a stepwise discriminant analysis was performed, but the used tree species are not common in central Europe.

A smaller number of spectral bands was found to generate more accurate identification than the use of all available hyperspectral bands of a given data set in [4]. These experiments indicate that the visible bands yield more information for discrimination in forest areas than near-infrared bands.

A pixel-based approach was compared in [5] to an object-based classification and it was stated that high spatial resolution can cause frequent misclassification when using a pixel-based approach and that the classification accuracy could be improved using a object-based approach.The spectral separability of five tree species - beech, douglas fir, spruce, larch and fir - by the mean values and the standard deviations of the grey-level values in the infrared, red and blue band was studied in [6] and very interesting findings on the separability of several tree species were provided. However they could not separate douglas fir from spruce and larch from beech. According to [7] shape patterns of the tree crowns like Gaussian curvature and crown roughness allow a better distinction between beech and oak than spectral or textural features. However a reliable tree delineation is needed to calculate those parameters.

A decision tree approach for tree species classification of 3 species was used in [8] and 87 image object metrics were analyzed. They found that the classification was possible using only four of these metrics, however, they admit that the small sample size for the analysis might have contributed to the good results.

We compared different data sources and spectral regions in order to gain a better understanding of the properties and differences of the spectra of several tree species. To identify relevant spectral bands is an important prerequisite for the decision making process to find a suitable sensor or a suitable combination of available remote sensing data at a preferable low cost for large area tree species classification.

This analysis is part of the project Virtual Forest, which aims at providing a database containing stand-based as well as single tree based information that can be used for forest planning, management and monitoring. The Virtual Forest is supported by the State of North Rhine Westphalia (NRW), Germany, the forest administration of North Rhine-Westphalia and the European Union (European Regional Development Fund – ERDF).

1. Study site, sensors and reference data

The study site is located north-east of the city of Arnsberg in Germany. For the RGB and CIR images an Ultracam X camera with a resolution of 10 cm per pixel was used. The infrared band was acquired at a wavelength of 840 nm. In addition satellite data from the SPOT satellite provided red, green, near-infrared and a short wavelength infrared band around 1640 nm. This data set has a resolution of 10 m per pixel for the red, green and near infrared bands and 20 m per pixel for the short wavelength infrared band.

To gain additional spectral information a hyperspectral data set was recorded with a AISA HAWK hyperspectral sensor, which is a small airborne near and short wavelength infrared sensor. The used configuration provided 320 pixels per scan line and 235 spectral bands in the range of 975 nm to 2449 nm, with a band width of 6.3 nm for each spectral band. The spatial resolution of the final product is 1.5 m per pixel. This data set covers an area of about 80 km^2.

As ground truth data for the analysis 7 forest stands were measured. Each tree in the forest stands was measured with its position, tree species, height and additional forest parameters.

The most important tree species in the federal state of North Rhine-Westphalia are oak (quercus robur), beech (fagus sylvatica), pine (pinus), european larch (larix decidua), spruce (picea), douglas fir (pseudotsuga menziesii) and poplar (populus), whereof the latter is not present in our data set. Other deciduous tree species which do occur in the designated area are less significant in the context of forestry in North Rhine-Westphalia and therefore are not part of this study.

2. Reference data and data sets

For each tree species half of the samples with the lowest reflectances were omitted to avoid the influence of shadow pixels in our analysis. Shadow pixels occur despite georeferencing due to measurement errors in the field, understorey trees that are measured in the field but are not visible from above the canopy and due to displacement of the tree crowns in respect to the position of the roots. The remaining samples sum up to 1637 single trees of 6 tree species whereof 605 belong to Douglas fir, 165 to larch, 249 to spruce, 6 to pine, 542 to beech and 70 to oak. Although the number of samples for pine is small and therefore might not be representative, we kept it in our study to get an estimate of the position of the spectrum of pine relative to the other spectra. The mean and the standard deviation were calculated.

2.1. Airborne image data and Spot satellite data

From the airborne images and color infrared images combined with the SPOT satellite data the spectral characteristics of each species were extracted. The within-species variability showed to be high which a low between-species variability and significant overlaps.

Aside of the spectral values the form of the spectrum is an interesting criterion. The spectral values may change due to lighting conditions or water content as described in [9] but parts of the form of the spectrum are specific for different species. Difference bands are calculated by subtracting one band from another band and can decrease the variability. The SW-R band is calculated from the short wave infrared band (SW) by subtracting the red band (R). In the same manner ratio bands like SW/R are calculated by dividing the short wave infrared band by the red band. Displaying the mean values and the standard deviation is a common representation. To ensure that it is a valid representation the Kolmogorov-Smirnov-Test as described in [10] was performed on several bands comparing the distribution of the samples to the Gaussian distribution. The estimated cumulative distribution function $S_N(x)$ of the data points is compared to the cumulative distribution function of the Gaussian distribution $P(x)$ and the maximum value of the absolute difference between the two functions is calculated as in equation 1.

$$D = \max_{-\infty < x < \infty} \left| S_N(x) - P(x) \right| \tag{1}$$

The probability of similarity is given by equation 2.

$$p = Q_{KS}\left(\left[\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}}\right] D\right) \tag{2}$$

with N denoting the number of samples and Q_{KS} the significance as given in equation 3.

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{2j^2\lambda^2} \tag{3}$$

Difference and ratio bands did not pass the test. Therefore, instead of the mean and standard deviation the 0.1587 quantile, the 0.5 quantile, also known as the median, and the 0.8413 quantile are used for the visual representation of difference and ratio bands. These quantiles show the same percentage of the data within the boundaries as does the standard deviation. The connecting lines between the individual difference and ratio bands for each tree species do not have any meaning. They do not approximate the spectrum but are used for a better visible representation only.

Three distinct characteristics can be observed in Figure 1 which shows the calculated difference bands. Beech is separated from the other tree species in the differences of the near infrared band (IR) with the red (R), green (G) and blue (B) bands. The difference in the IR-G band is slightly larger than in the other 2 bands with only small overlaps with the standard deviations of pine and oak. European larch is set apart from the rest of the species in the difference between the short wavelength infrared band (SW) and the near infrared band and Douglas fir and spruce are separated in the SW-R and SW-G bands with only a small overlap with the mean error of oak. The representation of pine in the difference and ratio bands shows an extremely large variation which is probably due to the small number of samples and therefore has to be revised.

In addition to the difference bands, ratio bands were calculated. The attributes of these bands slightly differ from the ones in the difference bands although the overall characteristics is similar in both band types, as for example spruce and Douglas fir differ more in the SW-R and SW-G bands than they do in the SW/R and the SW/G bands. The bands used so far are red, green, blue, one near infrared band and one short wavelength infrared band. All of the available bands show relevant information content, but in some cases the overlaps are substantial and more information is desired for a reliable discrimination.

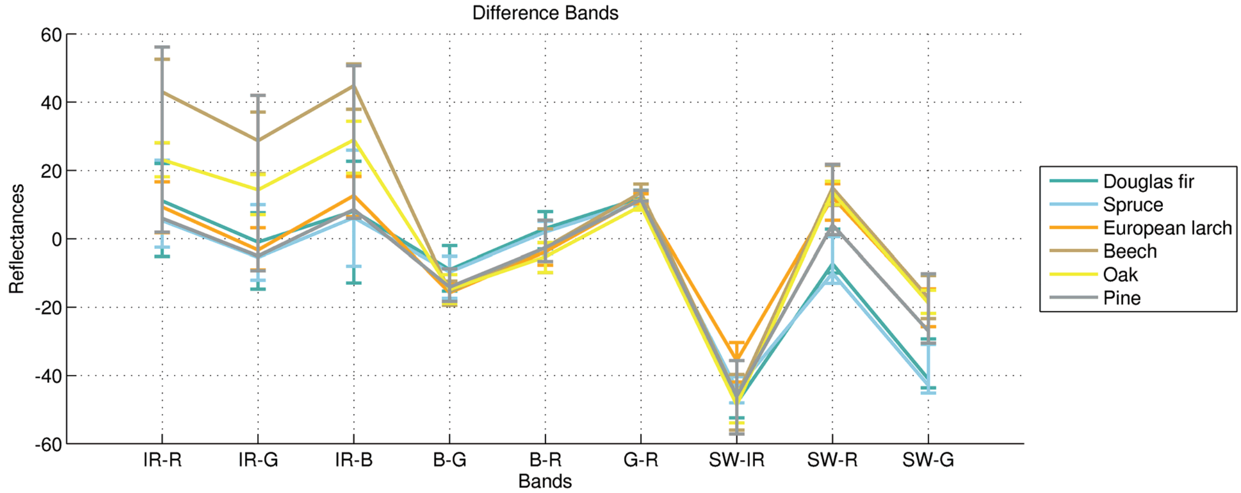

Figure 1. Difference bands.

2.2. Hyperspectral data

Although very high resolution hyperspectral data is not affordable for large scale forest inventories, the purpose is to find relevant bands and to use that information on the interesting bands for tree species classification to find a suitable sensor or data provider or to add single band sensors to an existing sensor setup for airborne data acquisition.

The spectra of the 6 tree species used in this study have a high overlap in our hyperspectral data set. Methods like penalized discriminant analysis (PDA) as described in [11] and used in [12] produce linear combinations that show how the components of the predictor vector contribute to the discrimination rule. The goal in this work is not to find discrimination rules, but to give an overview of the possible contributions of spectral bands. Hyperspectral data tends to be noisy and has small peaks and valleys in the spectrum. Therefore a 3 point mean was calculated to decrease the noise present in the collected samples without blurring the spectral information. From the whole spectrum 22 bands at local maxima and minima of the spectra were selected as local minima and maxima give a good representation of the form of the overall spectrum. The differences and ratios of these local extrema were calculated and analyzed.

The difference bands showed a high overlap for all the band combinations. The ratio bands showed several interesting characteristics for band combinations including the 146th band with a wavelength of 1894.5 nm, the 152nd with a wavelength of 1925 nm and the 155th band with a wavelength of 1944.9 nm. The most significant ratios were calculated with the 26th band with a wavelength of 1132.7 nm and are shown in Figure 2. At those bands two groups of tree species can be separated from each other. One contains Douglas fir, oak and beech and the other one consists of spruce, larch and pine.

Based on the visible data spruce and Douglas fir as well as European larch and oak seem to be difficult to separate. Therefore we separated those tree species spectra from the rest and analyzed them individually. The difference bands did not yield additional information for the discrimination of Douglas fir and spruce. The spectra of European larch and oak show significant differences for the same ratio bands, as did the Douglas fir and spruce spectra and are shown in the second chart of Figure 2.

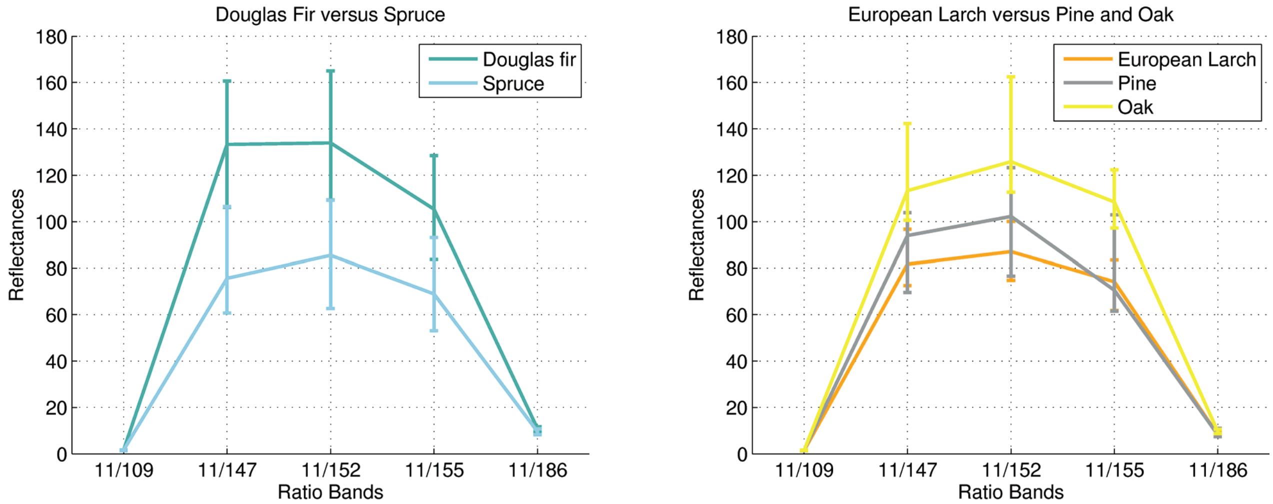

Figure 2. Douglas fir, <spruce, European larch, pine and oak in the AISA ratio bands.

3. Classification

To verify the found distinguishing features a decision tree was implemented as shown in Figure 3. The feature values like differences and ratios of values proved to be more stable and provided a better indicator for the tree species.

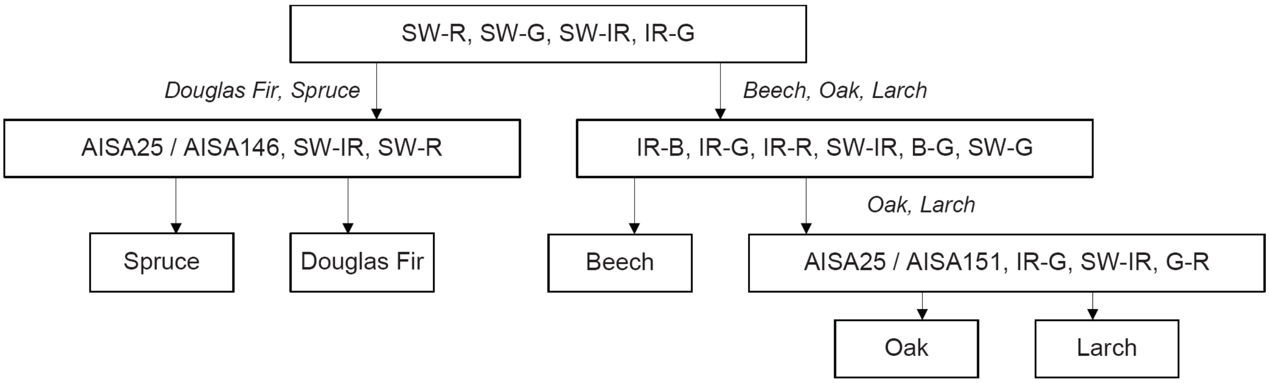

Figure 3. Decision Tree for the Tree Species Classification

As described in [5] we also found that pixel-based classification leads to frequent misclassification. Therefore we use objects which are calculated from the digital canopy model (DCM) and represent candidates for single trees. In young forest stands they may contain more than one tree but usually young forest stands are planted and therefore consist of only one tree species. In mature stands one region might contain only a part of a tree. If all parts are classified correctly the same tree species is assigned to them and it will not influence the final result.

We found that tree species can be distinguished based on few specific features. These features differ based on the species that are to be differentiated. Depending on the used algorithms, taking insignificant features into account can worsen the classification result. Especially valuable combinations of features are those where one feature allows to reliably classify variations to higher values and the other reliably classifies variations to lower values. An example is the combination of the SW-IR and the SW-R band for the discrimination of spruce and Douglas fir.

The first step of the decision tree separates non-forest and shadow areas from forest areas using the DCM as well as the overall brightness and the differences between the infrared band and the red, green and blue band respectively.

Therefore in each step of the decision tree a measure for the soundness of the classification is estimated based on the spectral overlaps in the analysis. These reliability images give a hint on the areas where the result needs to be confirmed by an expert and the algorithm needs to be refined.

Table 1. Confusion matrix.

	Douglas Fir	Spruce	Beech	Oak	Larch
Douglas Fir	79.3%	17.4%	0%	3.3%	0%
Spruce	4.8%	79.1%	7.6%	0.4%	8.0%
Beech	0%	0.9%	79.7%	14%	5.4%
Oak	0%	2.9%	14.3%	74.3%	8.6%
Larch	1.8%	9.8%	7.3%	10.4%	70.7%

3.1. Classification results

The results of the classification process are saved as tree species maps with an adjustable resolution.

The classification result and the reliability estimates can be blended into one image. Reduced reliability occurs mostly along the borders between deciduous and coniferous forest stands or next to non-forest areas as in those areas the coarse resolution of the SPOT satellite data makes it difficult to estimate the exact border between coniferous and deciduous trees or the spectral characteristics of the ground interferes with the spectral characteristics of the forest areas.

Unfortunately a reference data set was not available and the given data set could not be divided evenly due to the small amount of samples for oak. Therefore the only available validation at this moment is the comparison of the result to the training sample data set. This yielded a classification rate of 78%.

4. Conclusion and future work

The information given in this study can help with the search for a suitable data source for tree species classification. Although the data from the visible and the first part of the near infrared region showed to be more useful for tree species discrimination the hyperspectral data set also showed that short wavelength infrared regions contain helpful information. The area around 1640nm can for example help in the separation of Douglas fir and spruce from the other tree species, and the region around 1900nm in combination with the near infrared regions at around 1100nm can help in the distinction between Douglas fir versus spruce and larch versus oak.

An alternative source for a band in the short wavelength infrared region might be airborne laser scanner data, which is used in most forest inventories based on remote sensing data to estimate tree heights and other forest parameters. Several common lasers used for airborne data acquisition operate with a wavelength in the short wavelength infrared area. Evaluation of laser scanner data not only for the distance of the reflection but also for the intensity of the reflection might provide a useful additional band for classification in forest areas.

No data on the red edge region was available for this study. Therefore, we cannot give a statement as to whether it is a useful band for tree species classification as stated in [4]. A reliability estimate is a valuable asset for data verification and provides information on problematic areas which should be considered in the development of new classification techniques.

The next steps will be to use the information given in this paper for the acquisition of a suitable data set and for the development of a new classification algorithm. The results will be tested for its applicability in single tree based forest inventory and will be compared to other approaches common in tree species classification. Furthermore the addition of other european tree species like maple, ash, birch and alder will be tested.

References

[1] T. Key, T.A. Warner, J.B. McGraw, and M.A. Fajvan, A Comparison of Multispectral and Multitemporal Information in High Spatial Resolution Imagery for Classification of Individual Tree Species in a Temperate Hardwood Forest, *Remote Sensing of Environment*, vol. 75, p. 100-112, Jan. 2001.

[2] V. Pinard and A. Bannari, Spectroradiometric analysis in a hyperspectral use perspective to discriminate between forest species, *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol. 7, p. 4301-4303, 2003.

[3] J.A. van Aardt, Spectral Separability among Six Southern Tree Species, Mai. 2000.

[4] P. Gong, R. Pu, and B. Yu, Conifer Species Recognition: An Exploratory Analysis of In Situ Hyperspectral Data, \textit{Remote Sensing of Environment}, vol. 62, p. 189-200, Nov. 1997.

[5] N. Kamagata, Y. Akamatsu, M. Mori, Y. Qing Li, Y. Hoshino, and K. Hara, Comparison of pixel-based and object-based classifications of high resolution satellite data in urban fringe areas, *Proceedings of the 26th Asian Conference on Remote Sensing*, Hanoi, Vietnam. November 2005.

[6] B. Koch, J. Svoboda, P. Adler, and M. Dees, Automatische Baumartenerkennung auf der Grundlage digitalisierter CIR-Luftbilder, *Allg. Forst- u. Jagd-Zeitung*, vol 173, no 7-8, 2002, p.p. 131-140.

[7] H.-J. Fuchs, Methodische Ansätze zur Erfassung von Waldbäumen mittels digitaler Luftbildauswertung, Thesis, Georg-August-Universität Göttingen, Fakultät für Forstwissenschaften und Waldökologie, 2003.

[8] M.S. Chubey, S.E. Franklin, and M.A. Wulder, Object-based Analysis of Ikonos-2 Imagery for Extraction of Forest Inventory Parameters, *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 4, p. 383-394, April 2006.

[9] J. Stoffels, Einsatz einer lokal adaptiven Klassifikationsstrategie zur satellitengestützten Waldinventur in einem heterogenen Mittelgebirgsraum, *Thesis*, University Trier, 2009.

[10] W.H. Press and S.A. Teukolsky, Kolmogorov-Smirnov test for two-dimensional data, *Computers In Physics 1*, pp. 7477, 1988.

[11] T. Hastie, A. Buja, and R. Tibshirani, Penalized discriminant analysis, *Annals of Statistics*, vol. 23, p. 73-102, 1995.

[12] B. Yu, M. Ostland, P. Gong, and R. Pu, Penalized discriminant analysis of in situ hyperspectral data for conifer species recognition, *Geoscience and Remote Sensing, IEEE Transactions,* vol. 37, p. 2569-2577, 1999.

[13] E. Tomppo, H. Olsson, G. Stahl, M. Nilsson, O. Hagner, and M. Katila, Combining national forest inventory field plots and remote sensing data for forest databases, *Remote Sensing of Environment*, vol. 112, p. 1982-1999, May 2008.