Optimizing Tree-Species Classification in Hyperspectal Images

Etienne Barnard*, Moses Azong Cho†, Pravesh Debba‡, Renaud Mathieu†, Konrad Wessels*,
Charl van Heerden*, Christiaan van der Walt*, Gregory P. Asner§
*Meraka Institute, CSIR, Pretoria, South Africa
†Ecosystem Earth Observation, Natural Resources and Environment, CSIR
‡Logistics and Quantitative Methods, Built Environment, CSIR, Pretoria, South Africa
§Department of Global Ecology, Carnegie Institution for Science
260 Panama Street, Stanford, CA 94305, USA

Abstract—**We investigate the classification of eight prominent savanna tree species, based on hyperspectral reflectance data. Although two principal components account for 95% of the variance of the data, up to 20 components are found to be useful for classification. Scaling of these components so that all features have equal variance is found to be useful, and our best performance (88.9% accurate classification) is achieved with 15 scaled features and a support vector machine as classifier. A graphical analysis suggests that several exemplars ("endmembers") are required for each class, and this observation is confirmed by the large number of support vectors employed by the best classifier.**

I. Motivation

The availability of high spatial and spectral resolution imaging spectrometers, i.e., sensors that provide contiguous spectral data in narrow bands, has produced an embarrassment of riches within the field of remote sensing. Images obtained from such spectrometers provide a wealth of information on matters such as land use, biodiversity assessment and resource analysis; however, the optimal utilization of these images introduces novel computational and algorithmic challenges that have not been fully solved to date. In particular, we require classifiers that can achieve high classification accuracy at a limited computational cost (since classification results may be required for every pixel in a large set of two-dimensional images). This accuracy must be achieved from a small number of representative samples, since the cost of gathering "ground truth" samples is significant, and it is important to understand the minimal spectral requirements placed on the spectrometers, to avoid unnecessary costs associated with sensor procurement and maintenance.

In the current contribution, we investigate the ecologically important issue of the classification of tree species in hyperspectral images. As pointed out, for instance, by Cho *et al.* [1], the ability to map vegetation at the species level is of broad interest in ecology, and the particular task of tree classification can be used to compile resource inventories and to develop models of practically important variables such as fire-hazard status.

Within this task domain, we focus on two aspects, namely the number of training prototypes (also known as "endmembers") required for classification, and the overall accuracy that can be achieved with an optimized algorithm. Section II describes this task in more detail, and also summarizes the methods of data gathering and analysis that were employed. In Sections III and IV we provide, respectively, the qualitative and quantitative results obtained in this investigation. Finally, Section V places these results in perspective, and discusses a number of additional matters that are worthy of exploration.

II. Methods

A. Data set

Airborne hyperspectral data for the study site, the Kruger National Park ($31°20'32.41''E, 24°50'47.75''S$), South Africa were acquired in May 2008 with the Carnegie Airborne Observatory (CAO) system [2]. The CAO combines three major instrument sub-systems into a single airborne package: (i) High-fidelity Imaging Spectrometer (HiFIS); (ii) Waveform Light Detection and Ranging (LiDAR) scanner; and (iii) Global Positioning System-Inertial Measurement Unit (GPS-IMU). The CAO-Alpha configuration, which was used in this study, employs a pushbroom imaging array with 1500 cross-track pixels, and sampled the scenes in the visible-near infrared (VNIR) spectral region between 384.8-1054.3 nm (72 bands) at 9.23 nm spectral resolution (full-width-half-maximum) and a spatial resolution of 1.12 m. Apparent surface reflectance was derived from the radiance data using an automated atmospheric correction model, ACORN 5LiBatch (Imspec LLC, Palmdale, CA).

The Universal Transverse Mercator (UTM) coordinates of tree crowns were extracted from the image and downloaded into a handheld Leica differential global positioning system (GPS; < 1 m horizontal accuracy). With the help of the GPS and printed true colour composite maps of the study site, the various trees were located and identified in the field. Only the dominant trees have been used in this study; *Combretum apiculatum, Sclerocarya birrea, Terminalia sericea, Spirostachys africana, Pterocarpus rotundifolius, Dichrostachys cinerea, Acacia gerrardii* and *Acacia nigrescens*. The region of interest tool in Environment for Visualising Images (ENVI 4.7) software (ITT Visual Information Solution, 2009) was used to manually collect the image spectra of the various trees. The spectra of each tree crown were then averaged to a single

spectrum. We therefore have a single 72-dimensional vector representing the spectrum of each individual tree; there are a total of 334 trees in our set from the eight classes listed above. The class representations are not exactly uniform, with the number of samples in a class ranging between 24 and 79.

B. Analytic approach

Figure 1 shows typical spectra for each of the tree species studied in the current investigation. Clearly, these spectra are relatively smooth, which implies that the reflectances in the 72 spectral bands are not independent of one another. Our qualitative investigation therefore begins with Principal Component Analysis (PCA) of the spectra, which allows us to determine the number of linear components that contribute significantly to the variance across samples. Since PCA is not a scale-independent process (differential scaling of the input dimensions leads to changes in the directions and weightings (eigenvalues) of the principal components), we normalize the input features so that each individual dimension (i.e. the data in each wave-band) has zero mean and unit variance.

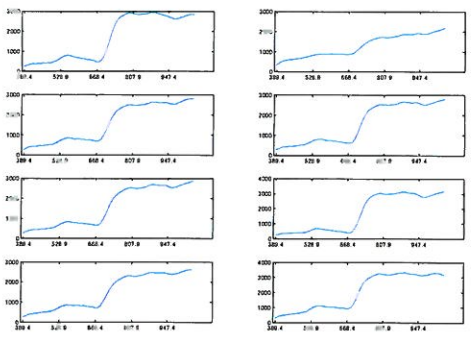

Fig. 1. Sample tree spectra for 8 classes of trees. Wavelengths in nm are shown on the horizontal axis, and the vertical axis shows reflectances (in units of 100 times percentage reflectance).

To investigate the relationship between the number of end-members per class and the achievable classification accuracy, we graph the projections of all data points onto pairs of principal components (eigenvectors). Since these projections consider only two dimensions at a time, they may be somewhat misleading regarding the separability of the various classes – they do, however, allow us to gain a basic intuitive understanding of the various class distributions in feature space.

This understanding is enhanced by a number of classification experiments. We compare the accuracies achievable with a number of state-of-the-art classifiers, such as multilayer perceptrons (MLPs), nearest-neighbour (NN) classifiers and support vector machines (SVMs). In this process, we investigate two matters related to the features used for classification:

- The preferred scaling of the different principal components is not obvious: on the one hand, normalizing each component to have the same variance seems like a sensible way to prevent the largest components from dominating the overall process. On the other hand, the smallest components are likely to be noise-dominated, implying that a boost to their variance may be harmful to classification accuracy.
- The optimal number of principal components to retain is also subject to experimental determination.

The overall sequence of steps involved in feature extraction are summarized in Fig. 2.

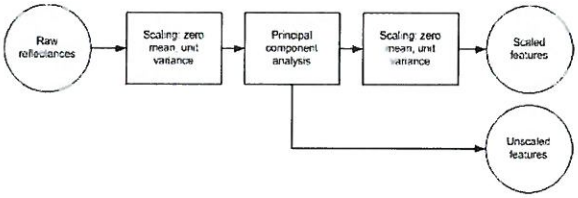

Fig. 2. Processing steps for feature extraction.

Our classifiers are trained with standard open-source tools such as Weka [3] and libsvm [4]; for the particular case of SVMs, we employ a radial kernel and a grid search to obtain the optimal hyperparameters. For our initial investigations into the feature-selection issues listed above, a nearest-neighbour classifier is used in a leave-one-out configuration (that is, each sample is classified based on its distances to all the remaining samples). The final classification results are obtained with 10-fold cross validation.

III. Qualitative Results

The magnitudes of the eigenvalues obtained with PCA are shown in Fig. 3; we see that there are two very strong components, which together account for 95.6% of the variance of the data set. The remaining components rapidly diminish in importance; by the eighth component, the weighting is less than 0.1% of that of the dominant component. The spectra corresponding to the two dominant components are shown in Fig. 4. These clearly correspond to short- and long-wavelength variations, the long-wavelength band (bottom panel in Fig. 4) being indicative of the red-edge transition between red and near infrared wavelength and the short-wavelength band (central panel in Fig. 4) corresponding mostly to pigment absorption in the blue and red spectral ranges.

Fig. 5 contains the projections of all data points onto several pairs of principal components. A number of observations can be made:

- We see that none of the features is able to achieve effective separation of all eight classes.
- As is often the case, the components with largest variance (s_1 and s_2) are also the most useful for discriminating between the different classes.
- The components with smaller eigenvalues are nevertheless not negligible from the perspective of classification – even the least significant component shown here (s_7) seems to add some discriminatory power to the feature set.

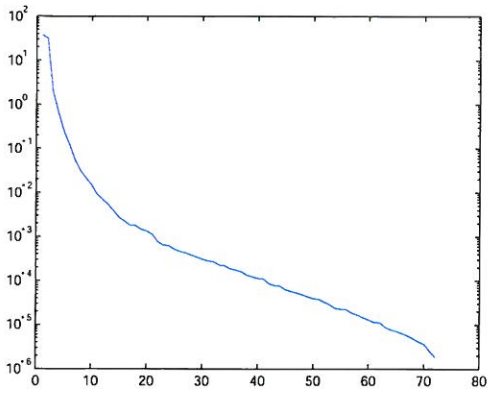

Fig. 3. Eigenvalues of the normalized covariance matrix – the largest two principal components account for more than 95% of the variance in the data.

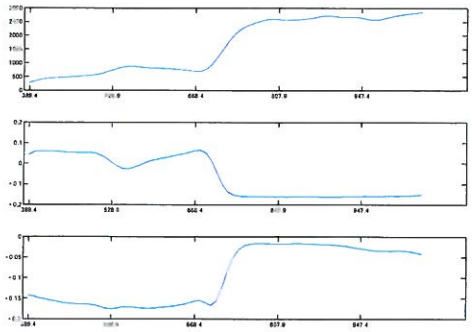

Fig. 4. Mean spectrum across all samples (top panel) as well as spectra corresponding to the two dominant principal components. The vertical axes are in arbitrary units, whereas the horizontal axes are wavelengths in nm.

- The classes appear to be rather diffuse and non-Gaussian in feature space, suggesting that a small number of endmembers per class will not be able to represent the full variability observed.

These observations are investigated quantitatively below.

IV. Quantitative Results

We compared the accuracies achievable with different numbers of principal components, as well as the two approaches to feature scaling mentioned in Sec. II-B, using leave-one-out nearest-neighbour classifiers. (We always keep the largest n principal components, rather than searching for the n components that give best classification accuracy.) The resulting classification accuracies are shown in Fig. 6. The importance of lower-ranking components is confirmed: despite the large magnitude of the two dominant components, as many as 15 or 20 components contribute to classification accuracy. Interestingly, it seems quite useful to scale the feature dimensions so that they have equal variance: for the largest 40 or so components, this improves accuracy over that achieved with unscaled features. Thereafter, the noise in smaller components erases the gains achieved; however, those components should in any case not be used for classification.

Fig. 6 also shows the accuracy achieved if each class is represented by a single endmember, located at its centroid. Classification accuracy is seen to be degraded, both with and without scaling, as would be expected from Fig. 5.

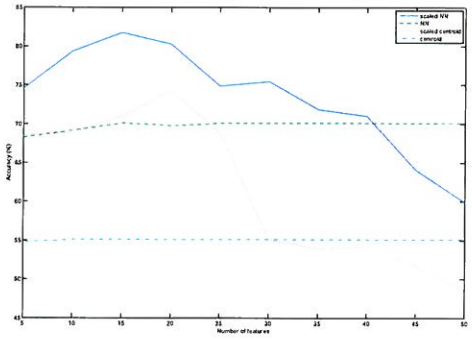

Fig. 6. Accuracy of nearest-neighbour classifiers and centroid-based classifier as a function of the number of principal components used for classification.

Finally, Table I contains the classification accuracies achieved with 10-fold cross validation and various combinations of classifiers and features. We see that the best performance (approximately 89% accuracy) is achieved with an SVM, using the top 15 scaled principal components. (There are unfortunately not other published results on this same data set; however, in [1] 57% accuracy was achieved for ten-class classification on a related data set.) Normalizing the various feature vectors so that classification is based on the angles between the vectors (as is done in the Spectral Angle Mapper (SAM) [5]) is somewhat harmful in this case. As with the NN classifier, a reduced feature set is seen to be useful, though the flexibility of the SVM enables it to narrow the gap between the reduced and complete feature sets.

Classifier	Feature Scaling	SAM	Number of features	Accuracy
Naive Bayes	No	No	72	55.7%
MLP	No	No	72	83.8%
NN	No	No	72	69.4%
SVM	No	Yes	72	84.4%
SVM	No	No	72	87.1%
SVM	Yes	Yes	15	84.7%
SVM	Yes	No	15	88.9%

TABLE I

Ten-fold cross-validation accuracies for various classifiers and feature sets.

The confusion matrix corresponding to the most successful classifier is shown in Table II, with the rows corresponding to the true classes and the columns to classes selected by the classifier.

The number of support vectors retained by the SVM is an approximate measure of the number of endmembers that

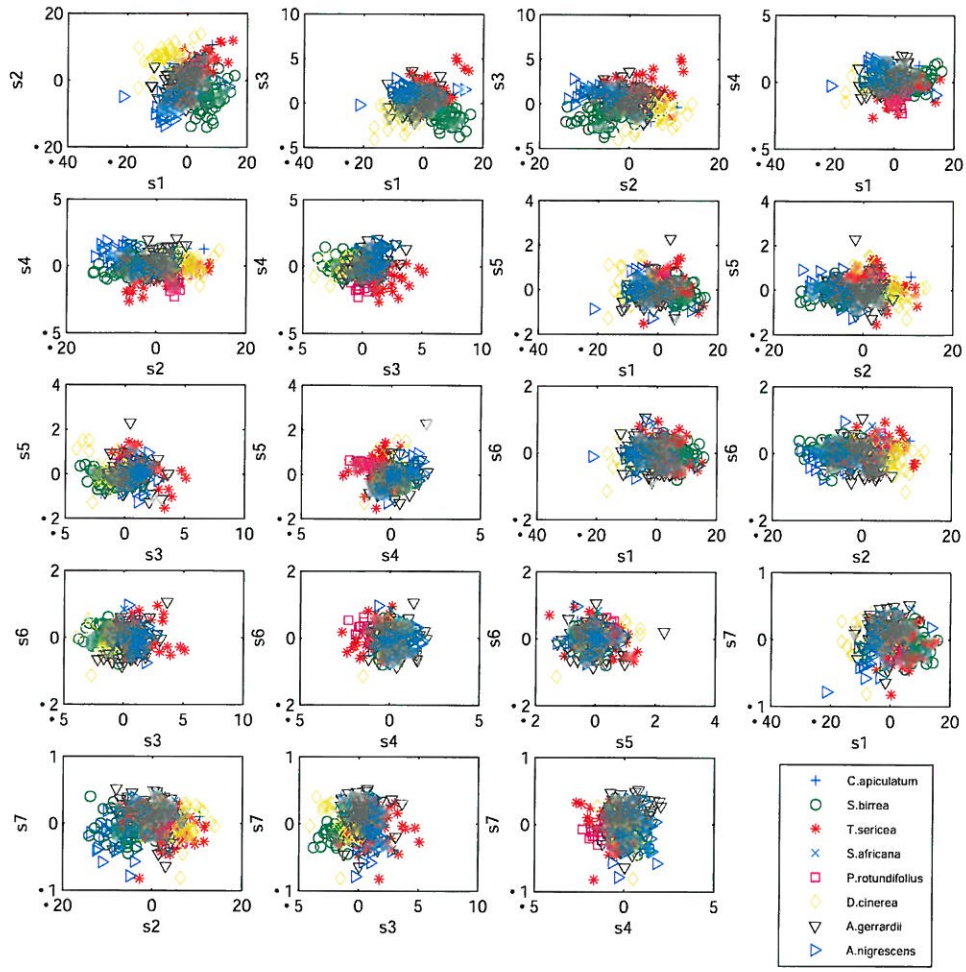

Fig. 5. Projections of all data on selected pairs of principal components.

	0	1	2	3	4	5	6	7
0	20	0	0	2	0	1	1	0
1	0	56	0	0	0	0	2	0
2	1	0	60	0	1	0	0	1
3	3	0	0	16	1	0	4	0
4	0	0	0	3	16	0	0	0
5	1	0	0	0	0	37	0	0
6	1	1	2	1	0	1	70	3
7	0	0	3	0	0	0	4	22

TABLE II
Confusion matrix of the most accurate classifier. Class labels as follows: 0: C.apiculatum, 1: S.birrea, 2: T.sericea, 3: S.africana, 4: P.rotundifolius, 5: D.cinerea, 6: A.gerrardii, 7: A.nigrescens

are useful for classification. Our best classifier employed an average of 202.0 support vectors for classification (thus, approximately 25 per class; average taken over the 10 folds), and the SVM trained on all 72 features required an average of 131.6 support vectors. These values can certainly be reduced without a large sacrifice in accuracy, but do suggest that a sizeable number of endmembers are required to give high accuracy (in agreement with the result of Cho *et al.* [1]). It is interesting to note that the larger feature vector required somewhat fewer support vectors – whether this difference will also be found in more general circumstances is a matter for

further exploration.

V. Conclusion and outlook

We have investigated the properties of spectra of several savanna species extracted from hyperspectral images, and shown that these targets have complex signatures in feature space. It is therefore unlikely that a small number of endmembers can represent these species adequately – a conclusion that was confirmed by our classification experiments. From a technical perspective, we have found the scaling of the largest principal components to be beneficial for classification accuracy, whereas Spectral Angle Mapping was not beneficial. A SVM was the most successful classifier on this task, producing the best results yet achieved on this data set.

Several interesting avenues for exploration are suggested by this work. Our results demonstrate that a relatively small number of features can produce accurate classification results; however, these features are currently derived by linear transformation from the entire spectrum, and it would be useful to investigate whether a smaller set of wavelength bands can produce comparable accuracy.

Our data was derived from measurements taken in one area within a small temporal window; for practical purposes, it is important to understand how these results can be generalized to deal with greater variability in location and time. Given the high cost of establishing ground truth compared to the relatively modest cost of obtaining complete images, it is likely that semi-supervised techniques will play a large role in such extensions.

Acknowledgement

We acknowledge the Council for Scientific and Industrial Research (CSIR), South Africa, for providing the funding for this study. Hyperspectral and LiDAR imagery were supplied by the Carnegie Airborne Observatory, which is funded by the Andrew Mellon Foundation. The CAO system is further supported by the W.M. Keck Foundation, the Gordon and Betty Moore Foundation and William Hearst III. Data were pre-processed by T. Kennedy-Bowdoin, D. Knapp, J. Jacobson, and R. Emerson at the Carnegie Institution for Science. Field data were collected by M. Cho, J. Fisher, B. Majeke, R. Main, R. Mathieu, C. Munyati, L. Naidoo and A. Ramoelo. B. Erasmus provided research inputs in the project. I. Smith provided valuable logistic support and help in the Kruger National Park.

References

[1] M. A. Cho, P. Debba, R. Mathieu, L. Naidoo, J. van Aardt, and G. P. Asner, "Improving discrimination of savanna tree species through a multiple endmember spectral angle mapper (SAM) approach: canopy level analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4133–4142, 2010.

[2] G. Asner, R. Hughes, P. Vitousek, D. Knapp, T. Kennedy-Bowdoin, J. Boardman, R. Martin, M. Eastwood, and R. Green, "Invasive plants alter 3-D structure of rainforests," *Proceedings of the National Academy of Sciences (USA)*, vol. 105, pp. 4519–4523, 2008.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[4] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 145–163, 1993.