OPTIMAL INDIVIDUAL SUPERVISED HYPERSPECTRAL BAND SELECTION DISTINGUISHING SAVANNAH TREES AT LEAF LEVEL

P. Debba, M. Cho, R. Mathieu

The Council for Scientific and Industrial Research (CSIR)
P.O. Box 395, Pretoria, 0001, South Africa

ABSTRACT

This paper uses simulated annealing and focus on the spectral angle mapper (SAM), to demonstrate how the separability of two mean spectra from different species can be increased by choosing the bands that maximize the metric. It is known that classification performance is enhanced when the differences in mean spectra for each endmember species are maximized. Comparison was made using the selected bands derived from the proposed method, to all bands in the electromagnetic spectrum (EMS), only the bands in the visible, near infrared and short wave infrared regions of the EMS and selected bands using stepwise discriminant analysis. The bands from the proposed method often indicates a better choice of band selection as viewed by the summary statistics for (a) the SAM measurements, (b) the correlations between bands and (c) the spectral information divergence (SID), for each pair of species; and the classification accuracy of SAM and SID.

Index Terms— Band selection, simulated annealing (SA), stepwise discriminant analysis (SDA), hyperspectral, spectral angle mapper (SAM), spectral information divergence (SID)

1. INTRODUCTION

One of the most common issues in hyperspectral classification is to improve class separability by selecting appropriate spectral bands [1]. These bands are usually highly correlated, and using all bands in classification algorithms could reduce the classification performance. Improvement to the classification accuracy is important because remote sensing images can cover large areas and thereby reduce labor intensive, sample based in situ classification of a targeted area.

The most widely used approach for (1) unsupervised classification is the principal components analysis (PCA) [2], also known as a feature extraction algorithm and (2) supervised classification is the Fisher discriminant analysis, in particular, stepwise discriminate analysis (SDA) [3], also known as a feature selection algorithm. SDA uses the between-class and within-class variances to select bands that maximize the separability of the classes. PCA is rather difficult to interpret because of the linear combination of all bands into the eigenvectors.

The success to correctly classify spectrally similar vegetation species is limited and dependant on the set of bands used for the classification algorithm. Hence. the objective of the study is to maximize the similarity measures, namely SAM, by choosing the proper set of bands. This enhances classification performance when the angle between the mean spectra for two species is maximized.

P. Debba is with the CSIR Built Environment. M. Cho and R. Mathieu are with the CSIR Natural Resources of the Environment. Corresponding author: P. Debba, email: pdebba@csir.co.za, tel: +27 12 841 3421, fax: +27 12 841 3037

2. DATA DESCRIPTION

The Analytical Spectral Device (ASD) spectrometer (FieldSpec3 Pro FR) was used to recorded hyperspectral measurements of leaf samples taken from several different savannah trees in the Kruger National Park, in an attempt to assess tree species diversity in the park. The hyperspectral data consist of 2151 spectral bands and seven plant tree species. The seven tree species include Lonchocarpus capassa, Combretum apiculatum, Combretum heroense, Combretum zeyherrea, Gymnospora buxifolia, Gymnospora senegalensis, and Terminalia sericia. Each tree species has 10 measurements recorded (see Figure 1 for Combretum apriculatum) with the exception of Gymnospora Buxifolia, which has only seven. The total data set therefore had 67 observations for the species measurements.

Initially, the number of bands were reduced to 1552 due to typical regions that are affected by atmospheric distortion and water absorbtion region: bands in the region 1.350–1.599 μm and 1.800–2.099 μm.

3. MAXIMIZING SEPARABILITY

3.1. Spectral Angel Mapper (SAM)

SAM measures the similarity as the angle formed between two spectra [4]. In this paper, we used the mean spectra for each species (Figure 2), so that the angle between the mean

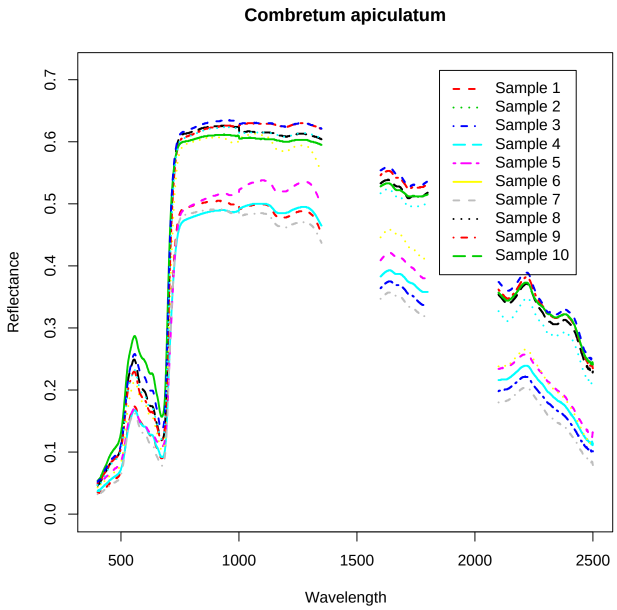

Fig. 1. Reflectance spectra of the 10 samples for Combretum apriculatum.

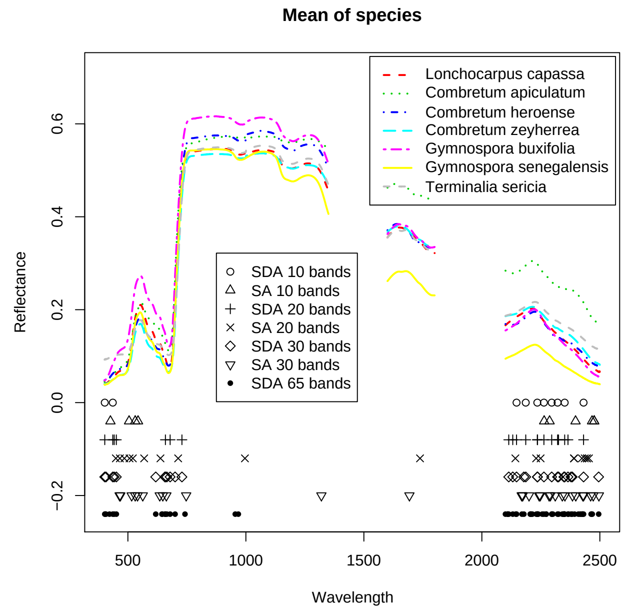

Fig. 2. Mean spectral reflectance for all seven species. Also, band selection using stepwise discriminant analysis and the proposed simulated annealing method. For SDA the results for the best 10, 20, 30 and 65 selected bands are shown, whereas for SA the results for the best 10, 20 and 30 selected bands are shown.

spectra is then used as a measure of discrimination [5] between species. SAM is thus defined as:

$$\mathrm{SAM}(\bar{s}_i, \bar{s}_j) = \cos^{-1}\left(\frac{\sum_{l=1}^{L} \bar{s}_{il}\bar{s}_{jl}}{\left[\sum_{l=1}^{L} \bar{s}_{il}\right]^{\frac{1}{2}} \left[\sum_{l=1}^{L} \bar{s}_{jl}\right]^{\frac{1}{2}}}\right), \quad (1)$$

where $\bar{s}_i$ and $\bar{s}_j$ are the mean spectral signatures for species i and j and L is the total number of bands considered. The smaller the angle between the mean spectra, the more similar are the two mean spectra. SAM is a useful deterministic similarity measure in species identification [5] and discrimination. One approach to finding the best bands is to determine which subset of bands maximizes Equation 1 for each combination of two mean spectra.

3.2. Spectral Information Divergence (SID)

SID computes the discrepancy between the probability distributions produced by spectral signatures of pixels [6] and useful in capturing subtle spectral variability [5]. SID is defined as:

$$\mathrm{SID}(r_i, r_j) = D\left(r_i \| r_j\right) + D\left(r_j \| r_i\right), \quad (2)$$

where $D\left(r_i \| r_j\right) = \sum_{l=1}^{L} p_i D_l\left(r_i \| r_j\right) = \sum_{l=1}^{L} p_l \log\left(p_l/q_l\right)$ and $D\left(r_j \| r_k\right) = \sum_{l=1}^{L} q_l D_l\left(r_j \| r_i\right) = \sum_{l=1}^{L} q_l \log\left(q_l/p_l\right)$ as derived from the probability vectors $p = (p_1, p_2, \ldots, p_L)^T$ and $q = (q_1, q_2, \ldots, q_L)^T$ for the spectral mean signatures $\bar{s}_i$ and $\bar{s}_j$, where $p_k = \bar{s}_{ik}/\sum_{i=1}^{L} \bar{s}_{il}$ and $q_k = \bar{s}_{jk}/\sum_{l=1}^{L} \bar{s}_{jl}$. The smaller the value of SID, the more similar are the two mean spectra. In this paper, we used SID as a stochastic similarity measure to evaluate the performance of the bands selected.

3.3. Fitness function

The fitness function is defined as

$$\mathrm{TSAM} = \sum_{i=1}^{L} \sum_{j=i+1}^{L} \mathrm{SAM}(\bar{s}_i, \bar{s}_j), \quad (3)$$

that is, the total accumulation of the spectral angles between pairwise mean spectra for all species. The larger the value of TSAM, the more likely the bands selected discriminate between the species.

3.4. Simulated annealing (SA)

Simulated annealing is a general optimization method used in finding the global optimum of an objective function called the fitness function $\phi(\omega)$. In our case, the fitness function TSAM depends on the selected bands ω that is to be maximized to increase separability between species. As such, simulated annealing [7] is a computer intensive search technique to find the bands optimizing the value of TSAM as a function of the bands, by continually updating this function at successive steps. Band selection by means of minimizing the total accumulated correlation was previously addressed by applying simulated annealing in [1].

Starting with a random selection of bands ω^0, $\phi(\omega^0)$ is calculated for TSAM. Let ω^i and ω^{i+1} represent two solutions with fitness $\phi(\omega^i)$ and $\phi(\omega^{i+1})$, respectively. Band selection ω^{i+1} is derived from ω^i by randomly replacing one band b_j of ω^i by a new band b_k in ω. Further to this, we also used a local procedure, where we considered all bands within a search radius of 10 bands to b_k. The band with the highest value of TSAM is then considered in the next step. A probabilistic acceptance criterion decides whether ω^{i+1} is accepted or not. This probability $P_c(\omega^i \to \omega^{i+1})$ of ω^{i+1} being accepted equals

$$P_c(\omega^i \to \omega^{i+1}) = \begin{cases} 1, & \text{if } \phi(\omega^i) \leq \phi(\omega^{i+1}) \\ \exp\left(\dfrac{\phi(\omega^{i+1}) - \phi(\omega^i)}{\mathbf{c}}\right), & \text{if } \phi(\omega^i) > \phi(\omega^{i+1}) \end{cases} \tag{4}$$

where $\mathbf{c}$ denotes a parameter. This parameter is reduced by a factor of 0.9 after several transitions are made, thereby decreasing the probability of accepting inferior moves. A transition takes place if ω^{i+1} is accepted. Next, a solution ω^{i+2} is derived from ω^{i+1}, and the probability $P_c(\omega^{i+1} \to \omega^{i+2})$ is calculated with a similar acceptance criterion as equation 4. The process terminates when it stabilizes.

4. RESULTS AND DISCUSSION

Tables 1, 2 and 3, respectively, contain the summary statistics (minimum, first quartile, medium, mean, third quartile and maximum) for the values of SAM, the correlation between bands and the values of SID, for each pair of species, using all spectral bands, using only the bands in the VIS, NIR and SWIR parts of the EMS, the 10, 20, 30 and 65 best bands selected by stepwise discriminant analysis and the 10, 20 and 30 best bands selected by the proposed optimization method.

The proposed method generally have higher SAM values (Table 1, especially when 10 bands were optimally selected. The values of SAM when using 20 and 30 bands from SDA were similar to that of the proposed method. Using all the bands, only the bands in the VIS, NIR and SWIR performed much worse than using the optimally selected 10, 20 and 30 bands.

Table 1. Summary statistics of SAM

	Min	Q1	Med	Mean	Q3	Max
ALL	0.04	0.06	0.10	0.10	0.13	0.24
VIS	0.04	0.07	0.10	0.12	0.16	0.24
NIR	0.01	0.02	0.02	0.02	0.03	0.05
SWIR	0.04	0.06	0.09	0.12	0.16	0.29
BANDS[1] 10	0.03	0.07	0.10	0.10	0.11	0.17
BANDS[1] 20	0.05	0.13	0.18	0.19	0.23	0.44
BANDS[1] 30	0.06	0.15	0.18	0.21	0.24	0.44
BANDS[1] 65	0.06	0.13	0.18	0.20	0.24	0.47
BANDS[2] 10	0.07	0.16	0.23	0.25	0.34	0.52
BANDS[2] 20	0.06	0.13	0.18	0.19	0.24	0.42
BANDS[2] 30	0.07	0.13	0.18	0.19	0.24	0.42

[1] using the selected bands from stepwise discriminant analysis.
[2] using the selected bands from the proposed simulated annealing method.

Table 2. Summary statistics of the correlation r^2 between bands

	Min	Q1	Med	Mean	Q3	Max
ALL	0.88	0.96	0.98	0.97	0.99	1.00
VIS	0.86	0.94	0.98	0.96	0.99	1.00
NIR	0.75	0.90	0.96	0.94	0.98	1.00
SWIR	0.97	0.99	0.99	0.99	1.00	1.00
BANDS[1] 10	0.79	0.92	0.96	0.94	0.99	1.00
BANDS[1] 20	0.59	0.86	0.93	0.89	0.98	0.99
BANDS[1] 30	0.54	0.81	0.90	0.85	0.95	0.99
BANDS[1] 65	0.65	0.86	0.93	0.90	0.98	0.99
BANDS[2] 10	0.00	0.24	0.66	0.49	0.85	0.99
BANDS[2] 20	0.74	0.92	0.96	0.93	0.99	1.00
BANDS[2] 30	0.80	0.93	0.96	0.94	0.98	0.99

[1] using the selected bands from stepwise discriminant analysis.
[2] using the selected bands from the proposed simulated annealing method.

The lowest pairwise correlations r^2 (Table 2) were for the 10 bands that were optimally selected. The pairwise correlations for the 20 and 30 bands that were selected using SDA were generally lower than that using the proposed method.

To demonstrate that although the proposed method selects bands that optimized SAM, these bands can also be used for other classification methods. We used SID to determine the separability between the different species and found that the bands selected from the proposed method result in greatest separability compared to SDA, using all bands, using only the VIS, NIR and SWIR bands.

The actual bands that were selected for SDA and the proposed method can be seen in Figure 2. Generally SDA selects more band in the SWIR region. The proposed method tends to select more bands in the VIS and NIR regions compare to the SDA method. A number of these bands are also ideally positioned in absorbtion areas of the EMS. The proposed method selects 30 bands distributed throughout the

Table 3. Summary statistics for SID

	Min	Q1	Med	Mean	Q3	Max
ALL	0.003	0.011	0.022	0.027	0.033	0.115
VIS	0.002	0.006	0.010	0.024	0.035	0.078
NIR	0.000	0.000	0.001	0.001	0.001	0.002
SWIR	0.003	0.005	0.014	0.025	0.038	0.114
B[1] 10	0.003	0.008	0.016	0.023	0.027	0.069
B[1] 20	0.004	0.023	0.037	0.047	0.061	0.160
B[1] 30	0.005	0.023	0.043	0.055	0.071	0.165
B[1] 65	0.004	0.025	0.045	0.051	0.059	0.186
B[2] 10	0.005	0.029	0.066	0.093	0.139	0.308
B[2] 20	0.006	0.027	0.046	0.070	0.089	0.249
B[2] 30	0.006	0.023	0.047	0.060	0.081	0.220

[1] using the selected bands from stepwise discriminant analysis.

[2] using the selected bands from the proposed simulated annealing method.

EMS while SDA method selects the 30 bands in the 0.4 μm–1.0 μm and 2.1 μm–2.5 μm.

We further accessed the proposed method by classifying the 67 observations according to SAM and SID using all the bands, bands in the VIS, NIR, SWIR, best 10, 20 and 30 bands identified by SDA and best 10, 20 and 30 bands identified using SA. The classification accuracy of these sets of bands was examined using the kappa coefficient as defined in [8].

Table 4 contains the kappa coefficient for each of the band regions using SAM and SID. The proposed method performs better for both classification methods when compared to using the bands in the entire EMS, bands in the VIS, NIR and SWIR regions. SAM performed best using SDA method with 30 bands (kappa coefficient of 72%) and the proposed method was closely followed when 10 bands were optimally selected (kappa coefficient of 70%). SID performed best using SDA method with 30 bands (kappa coefficient of 74%) and the proposed method was closely followed when 20 and 30 bands were optimally selected (kappa coefficient of 69%).

Table 4. Kappa coefficient

	ALL	VIS	NIR	SWIR		
SAM	0.56	0.63	0.44	0.44		
SID	0.60	0.58	0.44	0.41		
	Using SDA			Using SA		
	B[1] 10	B[1] 20	B[1] 30	B[2] 10	B[2] 20	B[2] 30
SAM	0.60	0.62	0.72	0.70	0.62	0.56
SID	0.63	0.70	0.74	0.67	0.69	0.69

[1] using the selected bands from stepwise discriminant analysis.

[2] using the selected bands from the proposed simulated annealing method.

5. CONCLUSIONS

In this paper, we have demonstrated that selecting bands by optimizing SAM through simulated annealing has some desired properties for improving classification accuracy. For pairwise comparison of bands, the proposed method of selecting bands (a) tend to have higher average SAM values thus implying greater species separability, (b) tend to have lower average correlations between the bands, and (c) results in higher average SID values thus demonstrating that the selected bands are not simply specific to SAM but can be used by other classification methods. The proposed method often resulted in better classification accuracy as viewed by the kappa coefficient for both SAM and SID.

Further investigation should address incorporating the within species variability since the proposed method uses the mean of the species and ignores the variability of the samples for each species. We are currently investigating other separability measures, that specifically takes into account the within- and between-class variability, with simulated annealing to improve on the classification accuracies.

6. REFERENCES

[1] J. P. Fang, Y. L. Chang, H. Ren, C. C. Lin, W. Y. Liang, and J. F. Fang, "A simulated annealing band selection approach for hyperspectral imagery," in *Proceeding of SPIE 6378*, 2006, pp. 1–10.

[2] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis, An Introduction*, Springer-Verlag, New York, 3rd edition, 1999.

[3] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[4] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, and K. B. Heidebrecht, "The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data," *Remote Sensing Of Environment*, vol. 44, pp. 157–159, 1993.

[5] Y. Du, C-I. Chang, H. Ren, C-C. Chang, J. O. Jensen, and F. M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, pp. 1777–1784, 2004.

[6] F. Van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 8, pp. 1–4, 2006.

[7] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines*, New York: John Wiley, 1989.

[8] J. Cohen, "A coeffiecient of agreement for norminal scales," *Educational Psychology Measurement*, vol. 20, pp. 37–46, 1960.