© EARSeL and Warsaw University, Warsaw 2005. *Proceedings of 4th EARSeL Workshop on Imaging Spectroscopy. New quality in environmental studies*. Zagajewski B., Sobczak M., Wrzesień M., (eds)

IMPROVING WITHINCLASS SEPARATION FOR VEGETATATION CLASSIFICATION

Hristo Stoianov Nikolov[1], Nina Nikolova Jeliazkova[2]

1. Solar-terrestrial Influences Laboratory, Remote Sensing Dept., Sofia, Bulgaria, hristo@stil.bas.bg
2. Institute of Parallel Processing, Distributed Computing Dept., Sofia, Bulgaria, nina@acad.bg

ABSTRACT

In the recent years some serious improvements in increasing spatial resolution of orbital remote-sensing systems has been made. In many articles the ETM+ instrument onboard Landsat is considered to have modest spatial resolution. Combining the high spatial and high spectral resolution (more than 30 spectral channels), new possibilities for obtaining precise information from remotely sensed data about the land cover types are presented. The single element of the data represents more adequately the specific type of land cover. In case of vegetation one application is precision agriculture, where the information obtained can be used to improve crop quality, monitor the irrigation and fertilization process, and enhance weed management. Another factor that could improve the state assessment is the textural information.

In this study we propose novel technique for classification of vegetation types combining spectral, spatial and textural data for better within class separation. The method applied is based on Bayesian decision rule with preliminary data processing for feature selection. The results show that of crucial importance is the parameters selection procedure.

INTRODUCTION

The vegetation maps are used in the process of inventory, monitoring, and managing the ecological resources across multiple spatial and temporal scales. Current vegetation maps usually only describe one vegetation characteristic, such as cover types, across the landscape.

Vegetation can be classified by any number of criteria, including height, canopy density, seasonal leaf area index (LAI). However, when using remote sensing the patterns of vegetation must be classified according to their spectral properties, whether from a single instantaneous image, a time series of images, or a selection from the bands available in the combined image sets. On the concept of data processing with complex arrays of spectral and temporal components, it should be cautioned that a combination of all possible approaches to classification might quickly become unmanageable. Instead, the great majority of successful vegetation classifications with NOAA/AVHRR, Landsat and SPOT sensors have relied on relatively simple combinations of spectral reflectance properties. New information gained from multitemporal radar coverage will extend the possibilities offered by the passive sensors. (iv)

The paradigm of land cover classification using remote sensing is assumes that different types of land cover have distinct reflectance properties. The unique spectral properties of a land cover class derive from its combinations of canopy geometry, leaf densities, colors, optical properties and moisture content, shadow components, transpiration rates, and non-vegetated reflectances. These factors aggregate within each pixel, and the population of pixels within the class offers a set of mean values and variances about those means. Together the two class attributes of central tendency and spread are defined within those image bands collected for the analysis, and are known as the class "spectral signature." When all signatures are collected and compared, it is optimal if they are well separated according to their means, at distances which are large compared with their internal variances. Accepted sets of signatures are presented to statistical classifiers, to assign each pixel of the image to one of the classes according to some form of best-match algorithm.

It is important to note that these reflectance properties are not constant through time. In fact, they vary greatly with the position of the solar light source (time of day and time of year) and with the position of the sensor view angle relative to the imaged area. The same process can be observed first hand by viewing a grassy area from different angles at different times of the day.

Classification of vegetation using spectral reflectance properties in different spectral bands is a reasonably straightforward procedure, but for handling the statistical calculations carried out on each and every pixel of an image, in each of the spectral wavelength bands available, specialized software is required.

The array of values at each pixel must be compared with the statistics generated from the training classes (in the case of supervised classification) or with multiple clusters in an iterative fashion (in the case of unsupervised classification). These computationally intensive calculations quickly result in many billions of computer operations and hours of computation time, depending on how many classes are desired and the type of classification algorithm used. In many programs past and future the computer processing cycle is daily for the entire globe.

Vegetation indices condense the data from two (or more) spectral bands into one level of information, they can reduce the amount of computation time required for image classification. However, this benefit comes at a cost of information loss (most usually in the distinctions associated with overall brightness), and the classification results may be degraded. Vegetation indices derived from band combinations might better be reserved for monitoring, while classifications in support of monitoring would gain from use of full multiband original imagery.

METHOD AND DATA

Bayes' decision theory is a fundamental statistical approach to the problem of classification (iii, xi). In this study the Bayesian approach of classification, which assumes that the observation vector **x** is a random vector whose conditional probability density depends on its class, has been chosen. This rule is the theoretically optimal decision rule and guarantees lowest classification error.

In this research the PCA method (ix) has been applied first for dimensionality reduction and next for comparison of the performance of the nonparametric Bayesian classifier for different number of principal components. As an alternative to the crisp classification methods, where each pixel is classified to exactly one class, the soft classification methods assign multiple class memberships to a pixel. Soft classification methods provide more realistic interpretation of the real world, where land covers intergrade gradually and boundaries between classes are sometimes vague. The accuracy of a crisp classification is usually summarized in an error (confusion) matrix, where rows represent classes as observed (ground truth) and columns represent predicted classes (i). The cell *(i,j)* contains the number of pixels from *class i*, predicted as *class j*.

The nonparametric technique of choice in this paper is the kernel density estimation. It is known to approximate the true density of the data if enough data points are observed (xvii). The idea of kernel density estimation is to model the density as a sum of the influences of the data points. The influence of a data point is given by a kernel function, which is symmetric and has maximum at the data point. Examples for kernel functions are Gaussian bell curve, square wave function, etc (ii). The density function takes higher values in regions, where some kernel functions have a significant overlap. Given univariate random sample $X_1,X_2,\ldots X_n$ from an unknown density f, the kernel density estimator of f at the point $x \in R$ is

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad (1)$$

where h is the *smoothing parameter*. The equation 1 has been calculated using *Very Fast Multivariate Kernel Density Estimation algorithm* (iii), which achieves several orders of speed improvement by using computational geometry to organize the data not decreasing the accuracy. The implementation uses kind of "kd-trees", a hierarchical representation for

point sets which caches sufficient statistics about point locations in order to achieve potential speedups in computation. For kernels with infinite support (like Gaussian) it provides an approximation tolerance level, which allows tradeoffs between evaluation quality and computation speed. The implementation of this algorithm is available as a Matlab toolbox, distributed by GNU LGPL license (0).

A Matlab script was developed for the vegetation recognition, making use of this toolbox for the probability density estimation. The classification procedure proposed is as follows:

1. The Principal Component Analysis (Kahrunen-Loeve Transform) is performed over all training data (all classes) in order to obtain orthogonal variables. The data is then projected into principal components;

Multivariate conditional probability densities for each of the *C* classes are estimated via Very Fast Multivariate Kernel Density Estimation algorithm (vii);

Posterior probability for each pixel of training data set is calculated using the estimated densities assuming equal prior probabilities for all classes. Thus for a pixel we obtain *C* probability values representing the degree of class membership. $p(w_i \mid x), i=1..C$.

2. Pixel-by-pixel classification. A pixel is assigned to the class with maximum posterior probability. The accuracy of classification is summarized in an error matrix
3. Object classification. In land cover type recognition we are more interested in recognizing homogeneous regions than single points. Here we propose two approaches in order to make decision for a each homogeneous region of specific cover type, which take advantage of the "soft classification":
    - Calculate mean class probability values over all points in a region. Allocate current region to the class with highest mean probability. Results are presented in Table 1: Kullback Leibler Distance for the training set..
    - Measure similarity between estimated conditional density of a class and estimated conditional density of a query region and allocate the region to the class with highest similarity. This can be accomplished by a probability distance as Kullback-Leibler (KL) distance:

$$KL(p,q) = \int \log\left(\frac{p(x)}{q(x)}\right) p(x)dx \,, \tag{2}$$

KL distance is always nonnegative, has a minimum at zero when the two probability densities are the same and becomes higher as the two distributions differ (there is no maximum value). In this context highest similarity between probability distributions means lowest KL distance. The object is classified to the class, for which KL distance between the probability densities of the object and class is a minimum.

We assess the classification quality using test data. The test data is also projected into principal components space, obtained in the step 1 above, and the posterior probability for each pixel of test data set is calculated using densities estimated from the training data (step 2). The same statistics as for the training data (confusion matrix, mean class probability, KL distance) are calculated and presented in corresponding tables and figures.

The input data consisted of sub scene of North Bulgaria in a form of 7-channel satellite image (667 x 663 pixels), obtained by Landsat Thematic Mapper instrument in the mid simmer of 1992. A modern topographic map at scale 1:25000 of the same region served as a ground truth. Shape files in ESRI format with correct subclasses were introduced from the *Corine Land Cover 1994* project for Bulgaria. An automated analysis was performed by the authors and the 161 objects for the 9 vegetation subclasses of interest, namely arable land, permanent crops, pastures, heterogeneous agricultural areas, were identified. In this manner the whole data set was produced. This data set was further divided into training and a test sets in a way that the test data set for each

class consists of a single object with largest number of pixels. Here should be mentioned that the number of pixels vary noticeably for the subclasses present in this sub scene – from 10000 for subclass 211 to 58 for subclass 244.

In our scenario we considered all 9 vegetation subclasses from CLC classification (x) and the suggested by (xvi) approach during the STATLOG project (xii, xiv), taking into account the texture. In this case each pixel is represented by 7 values for each spectral band and additional 7 values per each of its eight immediate neighbors, thus assigning (1+8)x7=63 features to every of the pixels.

RESULTS

On the preprocessing step PCA analyses was carried out and after it 11 out of 63 principal components were selected. After the running the PCA procedure for the training data a Kullback-Leibler (KL) distance was calculated. As it can be seen from Table 1 the KL distance provided excellent classification of the test set since training and test data are the same. This illustrates the fact that the training data is faultlessly recognized. This result is encouraging one since even for one dataset this not always true.

Table 1: Kullback Leibler Distance for the training set.

	211	212	221	222	231	241	242	243	244
211	0,00	50,15	15,72	Inf	15,22	20,78	14,39	10,36	18,18
212	9,65	0,00	7,64	33,86	9,62	7,45	7,40	8,25	15,32
221	Inf	Inf	0,00	36,35	9,23	10,34	7,36	7,23	14,60
222	11,72	33,00	15,80	0,00	9,56	12,79	10,69	10,60	18,25
231	11,13	30,00	10,94	Inf	0,00	12,07	7,72	6,31	13,39
241	9,99	14,32	8,31	45,49	8,62	0,00	6,52	7,10	13,83
242	11,06	24,74	9,88	Inf	8,13	10,92	0,00	6,16	11,31
243	20,59	39,22	17,66	50,81	12,74	20,04	13,03	0,00	16,29
244	25,37	76,14	27,26	77,38	15,48	27,07	16,16	9,07	0,00

Table 2: Kullback-Leibler distance between probability distributions of training (columns) and test (rows) data sets (9 classes, 63 features per pixel)

Kullback Leibler distance	211	212	221	222	231	241	242	243	244
211	7,87	28,00	10,93	50,15	10,30	11,65	8,43	8,40	14,67
212	17,45	69,64	22,22	33,86	12,54	23,16	10,91	10,56	16,98
221	8,44	14,47	7,74	45,56	9,58	9,92	7,42	10,27	15,74
222	14,95	24,69	15,73	8,77	11,62	14,80	13,08	12,20	20,98
231	11,05	22,77	10,13	22,42	6,80	9,63	8,73	8,14	15,84
241	8,92	27,58	10,13	44,35	7,84	11,39	6,31	7,66	11,08
242	9,14	9,75	7,29	34,78	6,82	6,31	6,22	7,00	14,36
243	30,33	Inf	28,34	63,35	20,26	31,92	17,91	9,12	19,60
244	21,85	28,61	15,78	48,67	15,23	17,04	15,18	8,89	15,02

Figure 1 shows an example of estimated conditional densities for the first principal component (PC) of 9 subclasses (top) and corresponding posterior probability (bottom). The densities of the rest of the PC's are similar. From this graph evident that it is hard to perform classification based only on spectral properties of single pixels. For this reason we took into account the values of surrounding pixels.

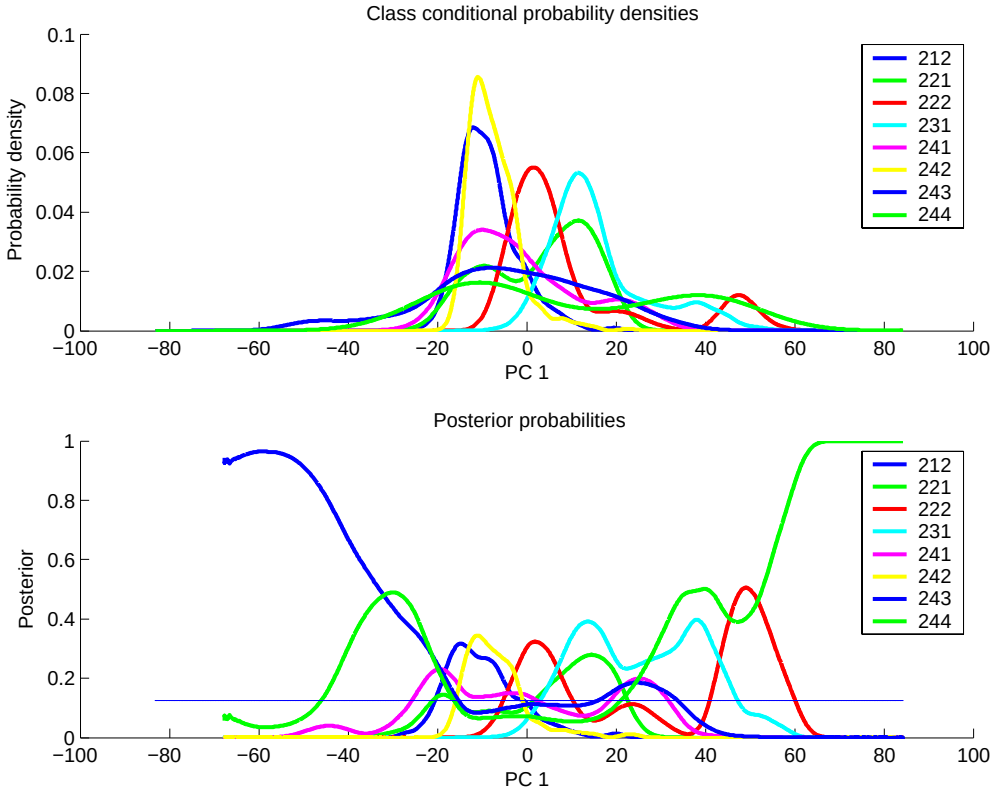

Figure 1 Conditional and posteriori densities of 9 subclasses

The results in table 2 confirm that even the number of features is increased the accuracy of classification will remain under the level of acceptance. One possible reason for this is the fact that in subclass 211 the largest number of pixels in concentrated and for the rest subclasses the number of pixels is much smaller.

CONCLUSION

From the results presented here our conclusion is that the spectral signatures, even after PCA and taking into account some textural properties, could not provide reliable basis for vegetation subclasses classification. One possible approach to overcome this is to introduce more features based on the textural properties of the vegetation subclasses. This we consider as future research activity and our anticipation is that better results will be achieved.

ACKNOWLEDGMENT

This paper was supported by the NSFB under contracts B-1306/03, NZ-1410/04, MU-NZ-1201/02.

References

Congalton R. & Green K., A practical look at the sources of confusion in error matrix generation. Photogrammetric Engineering and Remote Sensing 59, 641-644, 1993

Devroye L., L. Gyorfi, G. Lugosi. A probabilistic Theory of Pattern Recognition, Springer, 1996.

Duda R., P. Hart, D. Stork. Pattern Classification, 2nd ed., John Wiley & Sons, 2000.

Foody, G.M., 1999. The continuum of classification fuzziness in thematic mapping, Photogrammetric Engineering and Remote Sensing, 65, 443-451.

Fukunaga, K., Introduction to Statistical Pattern Recognition, San Diego, California, Academic Press Inc., 1990.

Gray A. and A. Moore, Nonparametric Density Estimation:Toward Computational Tractability, Proc. SIAM International Conference on Data Mining , San Francisco, USA, 2003.

Gray A. and A. Moore, Very Fast Multivariate Kernel Density Estimation using via Computational Geometry, in Proceedings, Joint Stat. Meeting 2003.

Hand, D. (1982), Kernel Discriminant Analyis, Research Studies, New York.

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Signal_Theory.pdf

http://europa.eu.int/comm/eurostat/ramon/

http://www1.ics.uci.edu/~mlearn/MLRepository.html

http://www1.ics.uci.edu/pub/mlearn/public_ftp/databases/statlog/

K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press,1972.

King, R.D., Feng, C., & Sutherland, A. (1995) StatLog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence.* 9, 289-333

MATLAB KDE class, http://ssg.mit.edu/~ihler/code/kde.shtml.

Mitchie, D., Spiegelhalter, D. and Taylor C. (eds) (1994). Machine learning, Neural and Statistical Classification, Ellis Horwood Series in Artificial Intelligence, Ellis Horwood.

Silverman B. W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986