Fast, Interactive Analysis of AVIRIS Data Using nPDF

Haluk Cetin, Timothy Warner and Donald Levandowski
Department of Earth and Atmospheric Sciences
Purdue University
West Lafayette, Indiana

Abstract

The n-dimensional Probability Density Functions (nPDF) technique is a new approach to data analysis that is both user-interactive and extremely fast, making it attractive for the analysis of high dimensional data, such as AVIRIS. Both scene and training data are mapped into nPDF space, producing a graphical display of the spectral data distribution. By overlaying this data, informational classes may be identified in the nPDF spectral data, and a classification look-up table is generated by the analyst. Unlike the maximum likelihood, Mahalanobis, and minimum distance classifiers, nPDF classification run-time does not increase with number of classes, and increases very slowly with increasing number of bands. A test scene of agricultural fields in the Maricopa area of southern Arizona was selected to demonstrate the application of the nPDF approach for the analysis of AVIRIS data on a personal computer. Classification accuracy measurements of the test scene, using 15 AVIRIS bands, indicate the nPDF approach has an overall accuracy rate of 71%, compared to 53% for minimum distance, 67% for maximum likelihood, and 68% for Mahalanobis distance. Using training fields from the test area, the entire scene, covered by 512 lines by 614 pixels and 180 bands, representing over 56 megabytes of data, was classified in under 13 minutes on a personal computer.

I. Introduction

The commonly used classification strategies of maximum likelihood, Mahalanobis, and minimum distance all have a number of inherent limitations. These include:

A. The memory requirements of the computer routines tend to be very large for high dimensional data, and the run-times are very long.

B. The algorithms thus tend to be implemented in computer routines that allow for only a limited number of input bands.

C. The algorithms are relative classifiers, and thus training fields from all spectral classes need to be identified prior to classification.

D. Classes are described statistically, and thus it is impossible to check that the previous requirement (point C.) has been satisfied.

E. Class overlap can be shown for only two bands at a time.

F. Mahalanobis and maximum likelihood classifications assume the data is normally distributed.

The nPDF analysis and classification technique, described by Cetin (1990) and also by Cetin and Levandowski (1991b), addresses all the problems of the traditional classifiers discussed above. This new procedure has been shown to be an effective tool in the analysis of Thematic Mapper (TM) data for mineral exploration (Cetin and Levandowski, 1991a), as well as for vegetation classification (Warner *et al.*, 1991a). Furthermore, the nPDF classification may be used to analyze the spectral and informational classes in multi-spectral thermal data (TIMS) (Warner *et al.*, 1991b).

In this paper the advantages of an nPDF approach for analysis of AVIRIS data will be discussed, with emphasis on the rapid run-times and the potential for analyzing high dimensional data. A test area from within an AVIRIS scene over the Maricopa Research Farm, in southern Arizona, is used to demonstrate the ability to process high dimensional data on a personal computer.

II. Methodology

The nPDF approach overcomes the problems of the complex spectral distribution of high dimensional data by combining both statistical and graphical methods in the analysis technique.

In three-dimensional feature space, the feature vector is defined by

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Figure 1 shows a generalized distribution of highly correlated, remotely sensed data in three-dimensional feature space. Instead of using the original feature space, the nPDF technique relies on the calculation of the Euclidean distance from each data point to selected corners of the feature space.

With a large number of closely spaced spectral bands, AVIRIS data tends to be highly correlated. We therefore choose corners 1 and 5, at opposite ends of the feature space (see Figure 1), for calculation of the nPDF components. For three-dimensional data, the Euclidean distances to the corners 1 and 5 are:

$$\mathbf{D_1} = \left\{ x_1^2 + x_2^2 + x_3^2 \right\}^{1/2}$$

$$\mathbf{D_5} = \left\{ (\mathbf{MR} - x_1)^2 + (\mathbf{MR} - x_2)^2 + (\mathbf{MR} - x_3)^2 \right\}^{1/2}$$

where

MR = maximum range of input data. For example, for 8 bit data, MR = 255.

The general formula for the calculation of the nPDF components for a hyper-dimensional cube is described by Cetin and Levandowski (1991). For corners 1 and 5, the hyper-dimensional Euclidian distances are:

$$D_1 = \left[\sum_{j=1}^{n} x_j^2 \right]^{1/2}$$

$$D_5 = \left[\sum_{j=1}^{n} \left[MR - x_j^2 \right] \right]^{1/2}$$

where
n = number of input bands.

For displaying the nPDF components, a scaling function is applied to the distances calculated by the formulae given above, in order to ensure that the nPDF components will fall within a preselected range.

$$nPDF_1 = S * D_1 / (2^{BIT} * n^{1/2})$$

$$nPDF_5 = S * D_5 / (2^{BIT} * n^{1/2})$$

where
S = desired scale factor for the nPDF axes
BIT = number of bits of input data (8 bit for TM, etc.).

The $nPDF_1$ and $nPDF_5$ components are calculated on a pixel by pixel basis and entered in a frequency array to map the nPDF data distribution. The procedure is then repeated for the training fields, and a second frequency array is calculated. For supervised nPDF classification, the nPDF data and the training distributions are then overlaid. In this way the nPDF spectral distribution of the informational classes is identified, and boundaries may be drawn between classes by digitizing polygons on the screen. This polygon class information is converted to raster form, and is used as a look-up table to classify the input data on a pixel by pixel basis [see slides 11 and 12].

III. Mapping and classification of Arizona AVIRIS data

Figure 2 is an image of a single band (0.64 µm) of the AVIRIS scene of the Maricopa area of southern Arizona. This data was collected on October 4 1990. Statistics of the original AVIRIS data showed that 44 of the 224 bands had at least some negative DN values. These bands were excluded from further analysis, leaving 180 bands remaining. It was also found that the radiometric range was limited to the equivalent of 8 bits (i.e. 256 DN levels) in most bands, except for a number of channels in the visible wavelengths. Thus, to reduce data volumes for handling full scenes of AVIRIS data, all data was scaled to an 8 bit range.

A test area (see Figure 3) was selected to compare the classification strategies of the different approaches. This test area is the Maricopa Research Farm, owned by the University of Arizona, and is located approximately 40 kilometers south of Phoenix. This research farm has been the site of intensive, remote sensing investigations since the mid-eighties (Jackson, 1990). Cover types present on the farm include cotton, alfalfa, bare soil, grapes and wheat stubble.

For comparison between nPDF and traditional classifiers, we were limited to 15 input bands, as this is the maximum number allowed by the image processing software that we used for the Mahalanobis, minimum distance and maximum likelihood classifications (ERDAS, 1990). Therefore, a maximum 15 AVIRIS bands, selected arbitrarily throughout the spectrum, were used. At the end of this section we describe the results of when all 180 bands are used.

Table 1 shows the percentage accuracy of the various routines for classifying the selected 15 bands of AVIRIS data of the Maricopa test site. The nPDF technique had the highest overall accuracy rate, 71%, compared to 53% for minimum distance, 67% for maximum likelihood, and 68% for Mahalanobis distance.

The nPDF technique is also very fast. Figure 4 shows that the run-times of the traditional techniques are several times longer than nPDF, and that for the Mahalanobis distance and maximum likelihood classifiers, run-times increase approximately exponentially with the number of bands. Furthermore, nPDF classification is unaffected by the number of classes (Figure 5), whereas the minimum distance run-time increases linearly, and the remaining classifiers increase approximately exponentially with number of classes.

To demonstrate the power of the nPDF approach, the entire AVIRIS scene of 512 lines by 614 pixels, with a total of 180 bands, representing over 56 megabytes of data, was classified using the Maricopa training information. This entire classification took 12 minutes and 48 seconds using a Northgate 25 megahertz, 486 personal computer.

IV. Discussion and Conclusions

The nPDF procedure is a significant improvement over the traditional approaches of Mahalanobis, minimum distance and maximum likelihood classifications. It is extremely fast, and yet also produces improved accuracy. Run-time does not depend on the number of classes, and run-time increases only slightly with increased bands. Mahalanobis and maximum likelihood classifiers increase approximately exponentially with the number of classes and the number of bands. The nPDF routine has a low memory requirement, making it suitable for processing AVIRIS and other high dimensional data on a personal computer.

The commonly-used classifiers - maximum likelihood, Mahalanobis and minimum distance - require the selection of training fields that encompasses all the spectral classes in the data, yet this is very difficult to check. In contrast, nPDF classification does not have this limitation because it is an absolute classifier. Furthermore, with the commonly-used classifiers, the training distributions are specified statistically; whereas in nPDF, the spectral distribution of data, as well as training distribution, is mapped in a visual representation. This results in better understanding of the class distributions present in the data. Inter-class separability is visually depicted, and spectrally overlapping classes may be identified. If a significant amount of overlapping does occur, the user can try alternative nPDF transformations to achieve a better separation for those particular classes. Alternatively, because the user creates the classification look-up table, the overlapping area can be apportioned in some user-defined way between the particular classes.

V. References

Cetin, H., 1990. nPDF-An algorithm for mapping n-Dimensional Probability Density Functions for remotely sensed data. *Proceedings of the 10th Annual International Geoscience and Remote Sensing Symposium, IGARSS'90* Vol. 1: 353-356.

Cetin, H. and D. W. Levandowski, 1991a. An integrated study of remotely sensed and geophysical data for mineral exploration in Lincoln County, Nevada: A new approach to mapping hydrothermal alteration using Thematic Mapper data. *Proceedings of the Eighth Thematic Conference: Geologic Remote Sensing, April 29 - May 2 1991, Denver Colorado,* Vol 2: 1271 - 1278.

Cetin, H. and D. W. Levandowski, 1991b. Interactive classification and mapping of multi-dimensional remotely sensed data using n-dimensional probability density functions (nPDF). *Photogrammetric Engineering and Remote Sensing* (In press.)

ERDAS, 1990. *ERDAS Version 7.4 Users' Manual: Image Processing Module.* Atlanta, Georgia. 218pp.

Jackson, R. D., 1990. Foreword: The MAC experiments. *Remote Sensing of Environment* Vol. 32: 77 - 79.

Warner, T. A., D. J. Campagna, C. S. Evans, D. W. Levandowski and H. Cetin, 1991a. Analyzing remote sensing geobotanical trends in Quetico Provincial Park, Ontario, Canada, using digital elevation data. *Proceedings of the Eighth Thematic Conference: Geologic Remote Sensing, April 29 - May 2 1991, Denver Colorado,* Vol. 1: 349 - 360.

Warner, T. A., H. Cetin and D. W. Levandowski, 1991b. Analysis of TIMS data using nPDF. *Proceedings of the 3rd Thermal Infrared Multispectral Scanner (TIMS) Workshop,* 20 - 24 May 1991, JPL Pub. 91-29 (in press.)

VI. Acknowledgements

We are very grateful to Phil Slater and Ray Jackson for kindly providing the AVIRIS and ground reference data used in this study, as well as to Howell Johnson for his patient assistance. This study was supported by the Indiana Mining and Mineral Resources Research Institute (U.S. Bureau of Mines grant number G1104118) as well as a Purdue University David Ross Fellowship.

VII. Table

Table 1. Classification accuracy percentages

Cover Class	Maximum Likelihood	Mahalanobis	Minimum Distance	nPDF
Cotton 1	99.17	98.66	92.34	95.75
Cotton 2	41.34	44.99	53.20	59.61
Cotton 3	59.72	61.98	50.27	59.56
Cotton 4	98.34	97.97	56.93	98.97
Alfalfa	80.95	83.40	33.39	83.49
Grapes	79.53	77.56	65.02	85.27
Bare Soil	79.65	77.56	70.79	85.27
Wheat Stubble	64.59	66.43	55.41	65.88
Overall	**67.34**	**67.77**	**52.74**	**70.99**

VIII. Figures

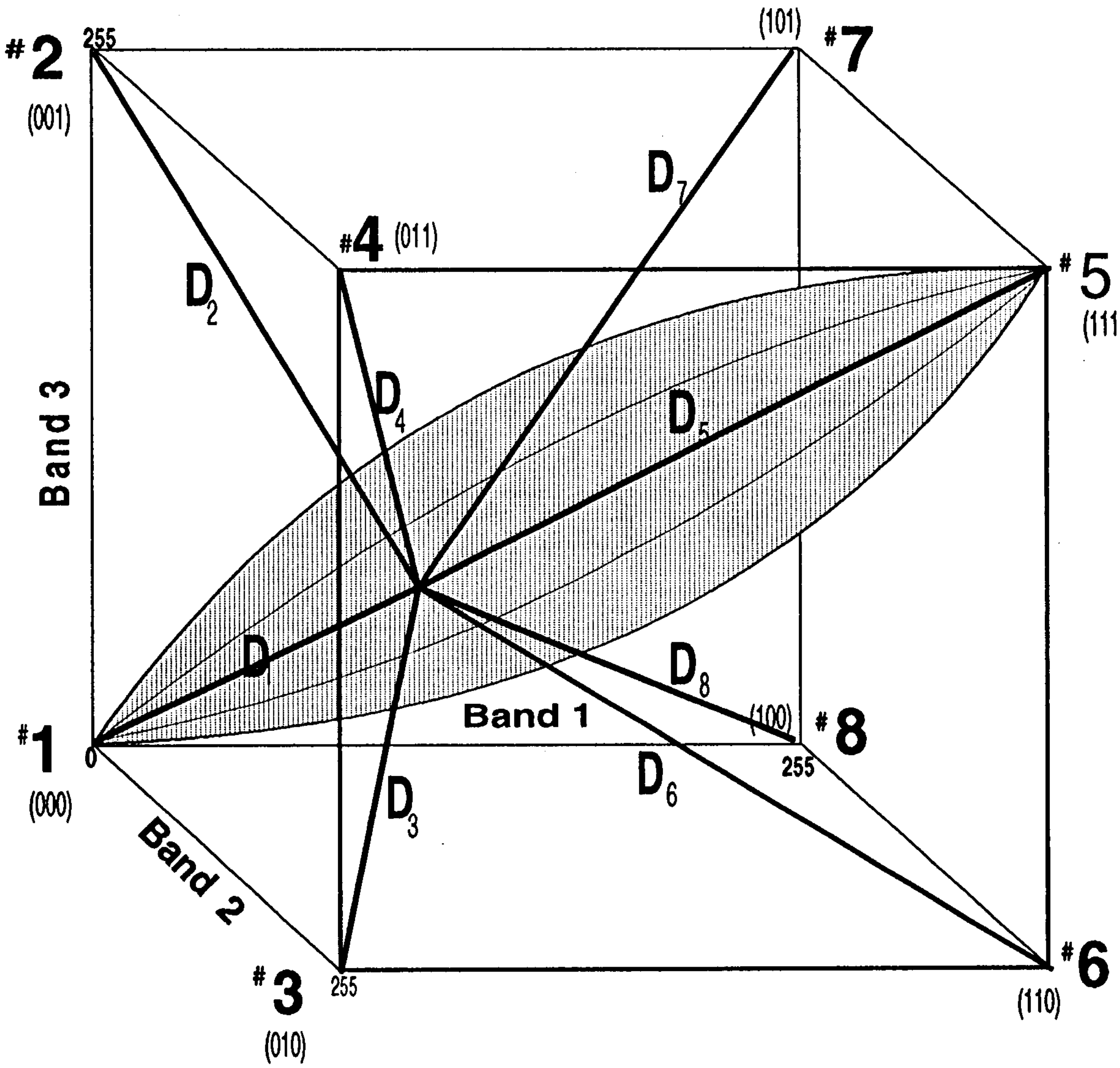

Figure 1. Three dimensional feature space showing schematic distribution of highly correlated data, and numbering convention for the corners of the data space.

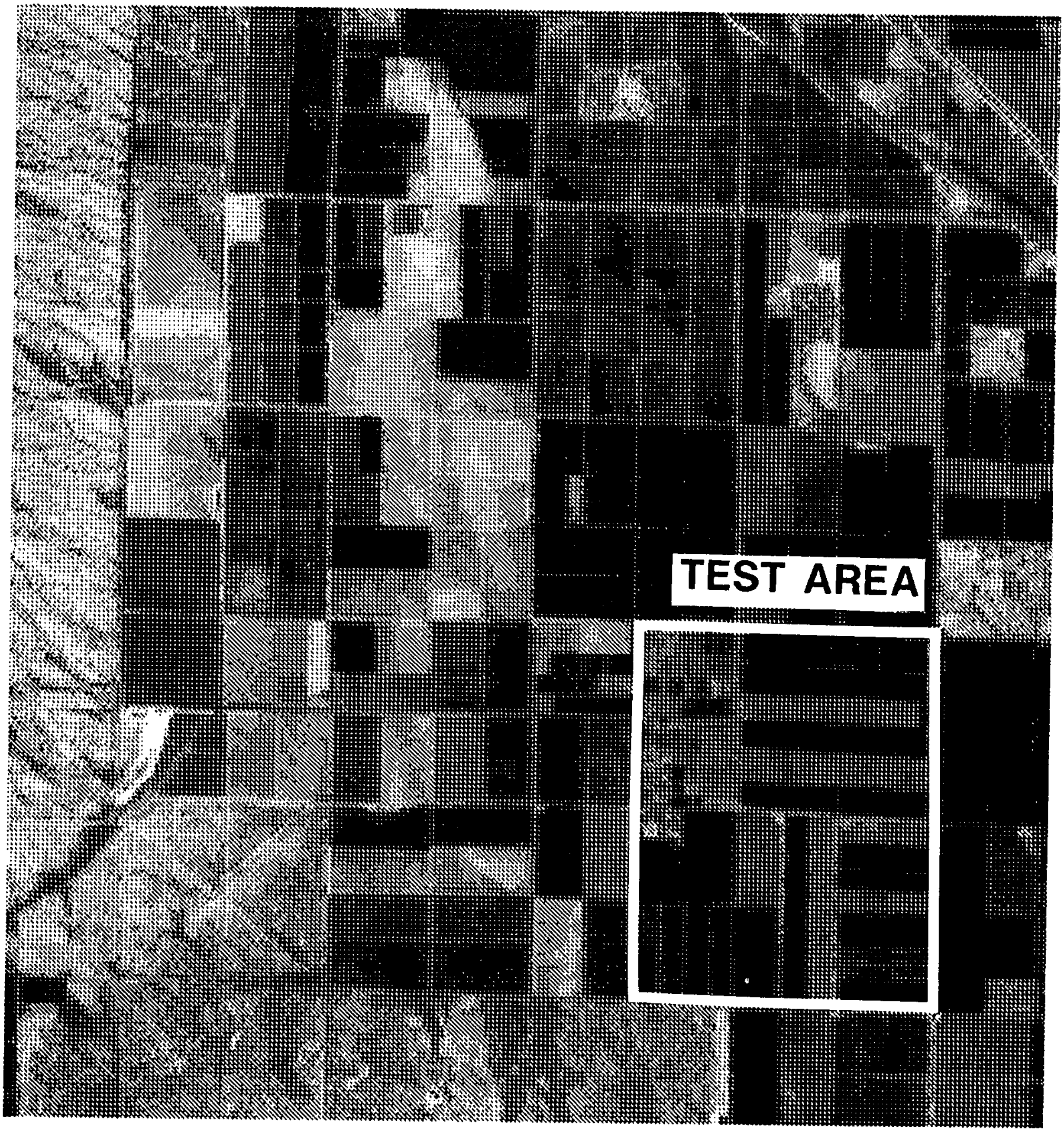

Figure 2. 0.64 μm image of Maricopa AVIRIS scene.

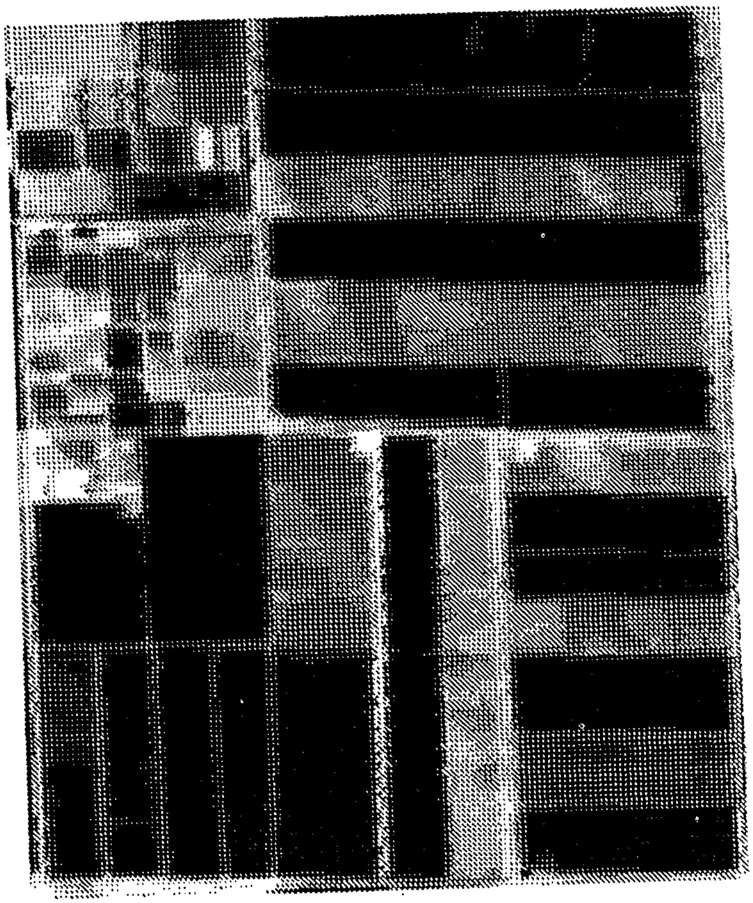

Figure 3. 0.64 μm image of Maricopa research farm.

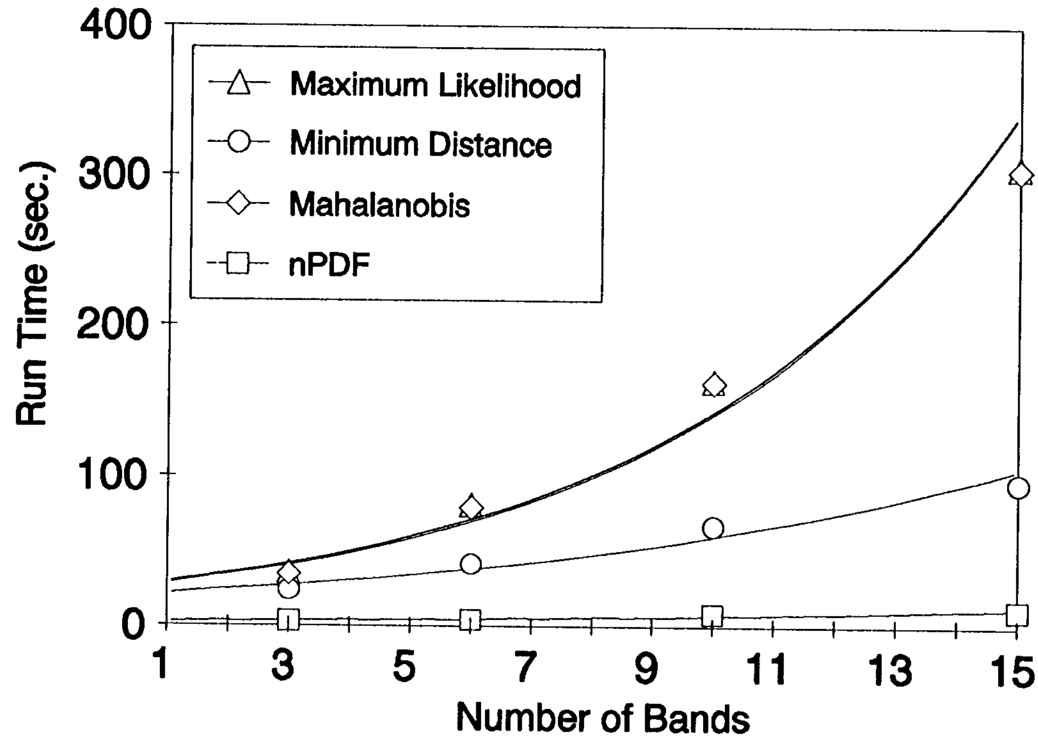

Figure 4. Graph of run-time versus number of bands.

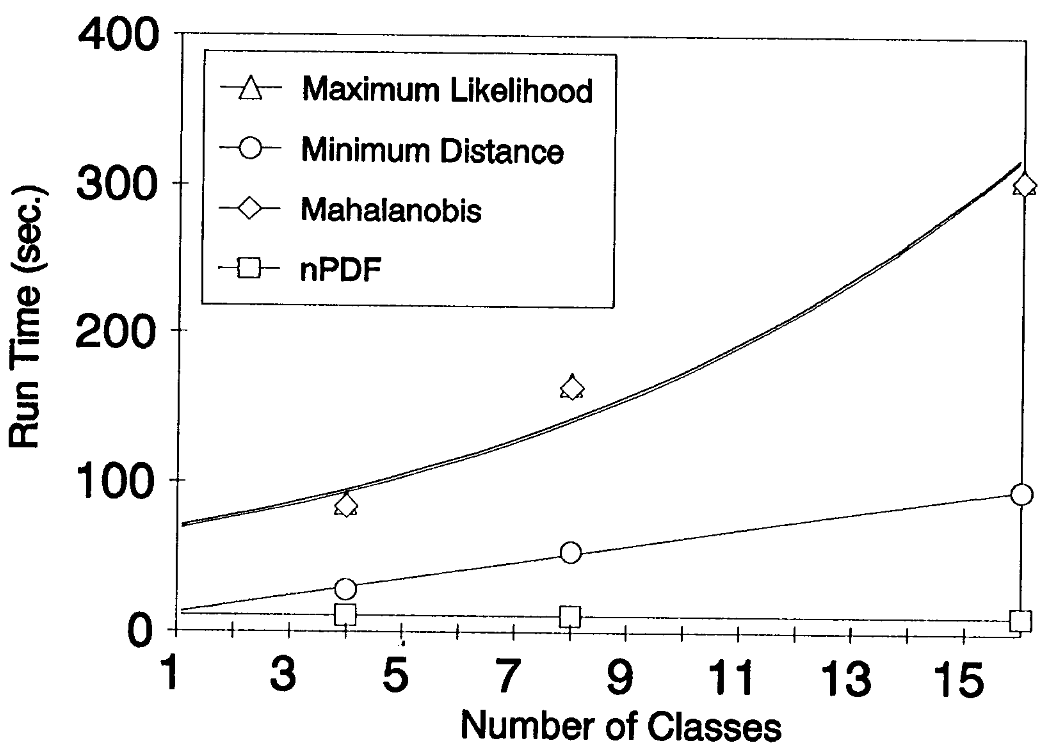

Figure 5. Graph of run-time versus number of classes.