7.9 - 10.28.6
CR-162273

FEATURE SELECTION VIA ENTROPY MINIMIZATION:
AN EXAMPLE USING LANDSAT SATELLITE DATA

(E79-10286) FEATURE SELECTION VIA ENTROPY MINIMIZATION: AN EXAMPLE USING LANDSAT SATELLITE DATA (Telespazio) 12 p HC A02/MF A01 CSCL 08F

N79-32610

Unclas
G3/43 00286

Angelo Zandonella[*] & Buzz Sellman[**]

[*] Telespazio S.p.a. per le Comunicazioni Spaziali

[**] Consultant, Environmental Research Institute of Michigan, Ann Arbor, Michigan

Presented to the:
"Secondo Convegno sulle Metodologie di Trattamento dell'Informazione", Trieste, February 1979.

ABSTRACT

A methodology for enhancing the significant spectral features in Landsat data is introduced.
The process, by which significant spectral features are determined, uses a minimum entropy model to guide subsequent analysis efforts.
Classification results using traditional and minimum entropy method are presented and discussed.

1. INTRODUCTION

The Landsat satellite system was designed to provide a global earth observation system capable of delivering digital tape data and photography to a wide audience of researchers as well as practicing resource managers. The widespread adoption of Landsat data as a useful source of information for on-going earth resources programs has made the original concept a high success.
This very success has created a large and growing community of users who need sophisticated pattern recognition techniques, but who themselves are not prepared to personally develope and refine the required techniques. This new "breed" of user, often a practicing environmental resource manager or an individual of similar training and experiences, generally has limited time and funds to support computer aided analysis of Landsat data.
Nevertheless, he wants sophisticated analysis that can be done quickly and cheaply. Thus there is a need, as well as an opportunity, to join technique developers with users to improve those techniques that effectively transform the raw data into useable information at the lowest possible speed and cost. This study reports, on the efforts of just such a partneship, bringing together two individuals with differing yet complementary interests to efficiently translate Landsat data into information readily useable by the varied user community of the Landsat system. The problem addressed in this study, at least in an introductory way, was how to improve the process of spectral feature selection, that is, locating the spectral training sets which are to comprise the prototypes for classification of the entire data

set.

2. THE MINIMUM-ENTROPY CONCEPT

Entropy is a statistical measure of uncertainty. At the onset of a Landsat data analysis effort, one is confronted with considerable uncertainty as to the number and kind of spectral classes that are contained on the data tape. In fact, consider an ensemble of potential spectral classes, the optimum feature selection mode will be that which chooses features which minimize the entropy of this ensemble. Since this is equivalent to minimizing the dispension of the various pattern populations, it is reasonable to expect that such a feature selection mode will have clustering properties. This concept can be effectively used in the design of an optimum feature selection process within a pattern recognition system (refer to Tou and Heydorn, 1967).

Consider a pattern recognition system which is designed to recognize K pattern classes.

For each pattern class, the feature selection process within the system will determine the set of discriminating features which are necessary for a correct recognition of these classes.

Assume that each of the K pattern population is characterized by a normal probability density function and the covariance matrices, describing the statistics of the K pattern classes, are equal .

Let, for Landsat satellite data:

$\underset{(n,m)}{X}$: a matrix of "n" pixels in "m" spectral bands. This matrix is a pattern of ℓ-th training set, where $\ell \in K$.

$\underset{(n,p)}{Y}$: a matrix of "n" pixels in $p \leq m$ images transformed.

$\underset{(m,p)}{A}$: a transformation matrix of the spectral bands. The columns of this matrix are the feature vectors of the pattern classes.

The method employed generates a linear transformation matrix $\underset{(m,p)}{A}$ (or "p" feature vectors) which operates on $\underset{(n,m)}{X}$ to yield a new matrix $Y_{(n,p)}$, so that the intraset dispersion (entropy) of $Y_{(n,p)}$ is minimized. This transformation may be written as:

$$Y_{(n,p)} = X_{(n,m)} \cdot A_{(m,p)}$$

If one assumes a multivariate normal distribution for each pattern population, this function is characterized by its mean vector and covariance matrix which is, in turn, characterized by its eigenvalues and eigenvectors. These eigenvectors carry the information describing the properties of the patterns under consideration. However, some of these eigenvectors bear less information, in a pattern recognition sense, than others, and may be ignored. In fact it would be desireable to use a method which provides for the selection of only the most significant feature vectors. Such a method is possible since the entropy function of $Y_{(n,p)}$ is minimized when we select "p" eigenvectors associated with "p" smallest eigenvalues by forming the transformation matrix $A_{(m,p)}$, (refer again to Tou + Heydorn, 1967, or Watanabe, 1969).

3. PROPERTIES OF THE MINIMUM-ENTROPY TRANSFORMATION

The main properties of this method are:

a. The reduction in dimensionality of the patterns.
In fact, the minimization of the entropy function implies the mathematical idea of information compression over the coordinate system so that most of the random patterns are concentrated on a few coordinates instead of widely distributed among all of them.

b. The orthonormality of the features and the transformed image. This is due to the fact that the primary vectors are the eigenvectors of a real symmetric matrix (covariance matrix). The orthonormality implies that the images transformed are uncorrelated. Note, however, that for Landsat satellite data there is some redundancy in the information content between contiguous bands.

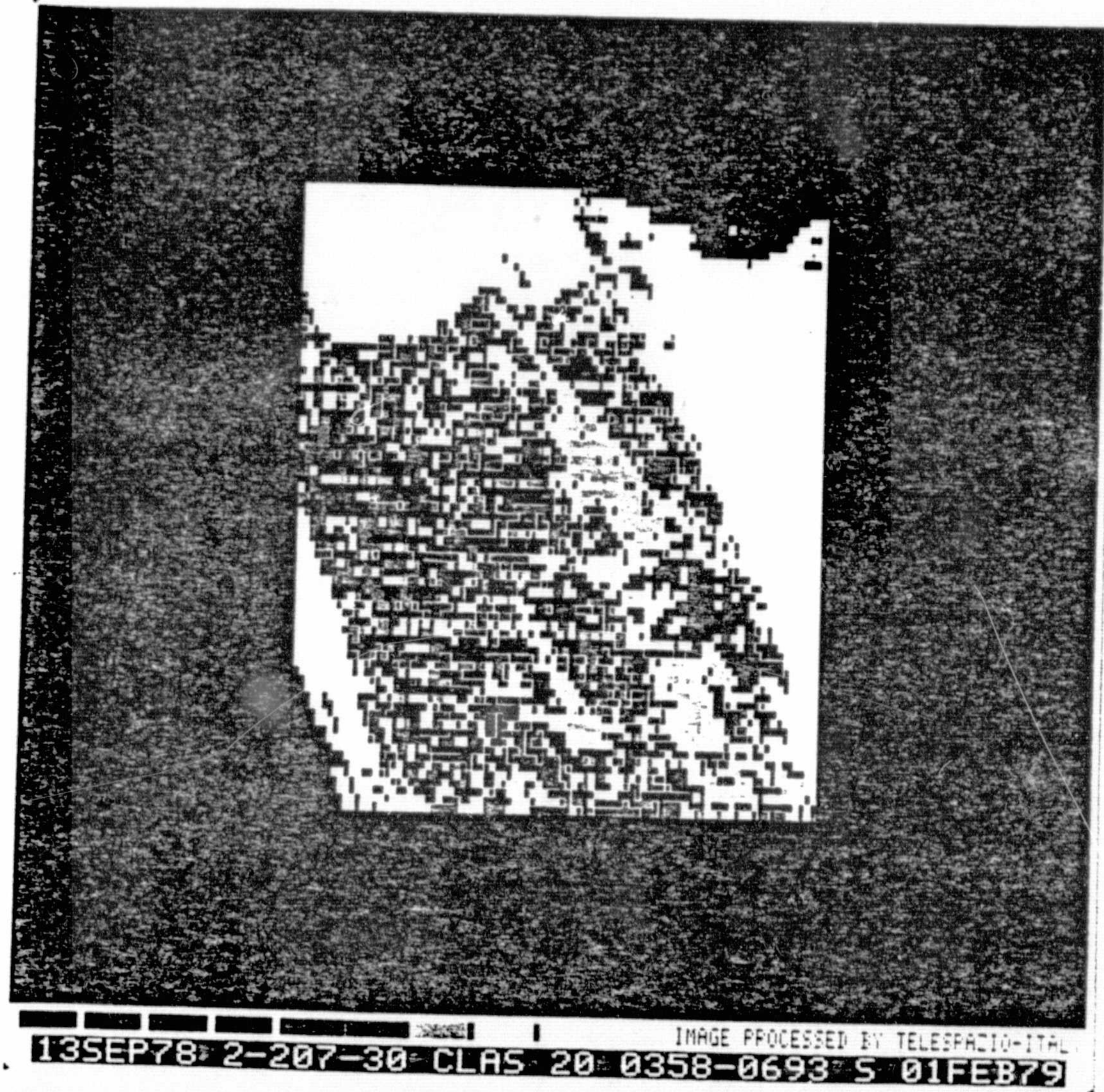
c. Rank ordering of the features as a function of their relative discriminant importance.
In fact this rank is made according to the descending order of the associated eigenvalues. Since it is possible to demonstrate, see i.e. Kendall (1972), that there exists an equivalence between the values of the eigenvalues and the variances of the features, the resulting features will contain, for its low variance, the maximum possible discriminating information concerning the pattern classes.
In relation to the properties discussed above and the physical characteristics of the Landsat spectral bands, it's possible to define a vector of features formed by the 1st eigenvector, associated with the smallest eigenvalues, for the pair contiguous spectral bands:

4 (0.5 - 0.6 μm) and 5 (0.6 - 0.7 μm)
5 (0.6 - 0.7 μm) and 6 (0.7 - 0.8 μm)
6 (0.7 - 0.8 μm) and 7 (0.8 - 1.1 μm)

This vector is used for training set selection.

4. AN EXAMPLE OF THE MINIMUM ENTROPY MODEL AS APPLIED TO LANDSAT DATA

Landsat data for the border area between Lucca and Pisa province, the forested shoreline near lake Massaciuccole, were processed using "standard" pattern recognition methods, that is, commonly used digital techniques for conducting an unsupervised classification, leading to a land use type map. Figure 1 shows a black and white print of the final color classification image (map). Water at the lower left (Tyrrhenina Sea) and upper right (lake Massaciuccole) are very dark, while unclassified areas (primarily agricultural land types) are depicted as off-white.
Three classes of woodland are depicted in the center of the scene as light, medium and dark grey. An assessment of the correctness of this classification map into three woodland classes was verified using black and white aerial photography at a scale of 1:13,000.
Thus, this classification correctly discerns three forest types that vary in their density (of trees), age (tree height and width),



(Original in color)

FIGURE 1 - Final classification image (woodland, water) derived from traditional technique.

and/or composition (e.g. ratio of trees to brush).
What is most important to note, moreover., is the fact that in many cases the "accuracy" of the classification as determined by the user will very often be subjectively determined after viewing the final classification map (image). With the more traditional approach to processing and analysing Landsat data, as was used to make this classification, such an image is not available until the very end of the processing program.
The use of the minimum entropy model to prepare a transformed image, that is made available early in the processing program, is an initial advantage over traditional procedures. Figure 2 (also a black and white print of the color original) show just such an image for a sligthly larger area than that depicted in Figure 1.
Figure 2 shows more of the Tyrrhenian Sea on the left and lake Massaciuccole in the upper center of the image. In the center, grey levels indicate the varying same spectral diversity amongst the woodedshoreline as was shown in Figure 1. However, this image is made available to the user early in the processing program (refer to Figure 3), and can be used immediately as a guide to training set selection .
Employment of the minimum entropy model also provides the user with a very powerful analytic tool to complement his subjective reactions to the 1st transformed image.
As was suggested earlier, it is very important to proceed in a recognition program with training sets that are different and representative of the entire scene.
The hierarchical classification method, applied to the feature vectors (refer, again, to Figure 3), provide the analyst/user with a convenient means to satisfy these conditions, that is, different yet representative.
The hierarchical classification is represented diagrammaticaly in Figure 4. This figure summarizes the relationships between every pair of groups (training sets, entire scene, and cumulative behaviourof training sets) in the form of a dendrogram. These relationships are expressed in terms of correlation coefficents so that a statistical threshold, set by the user, can be employed.
What can be seen in Figure 4(a) is that the five spectral classes (3 woodland, 2 water) are different, but they are not representative of the entire scene. Why?
Refer back to Figure 1. Note that this scene includes unclassi-

(Original in color)

FIGURE 2 - Minimum-entropy transformation color-composite image.

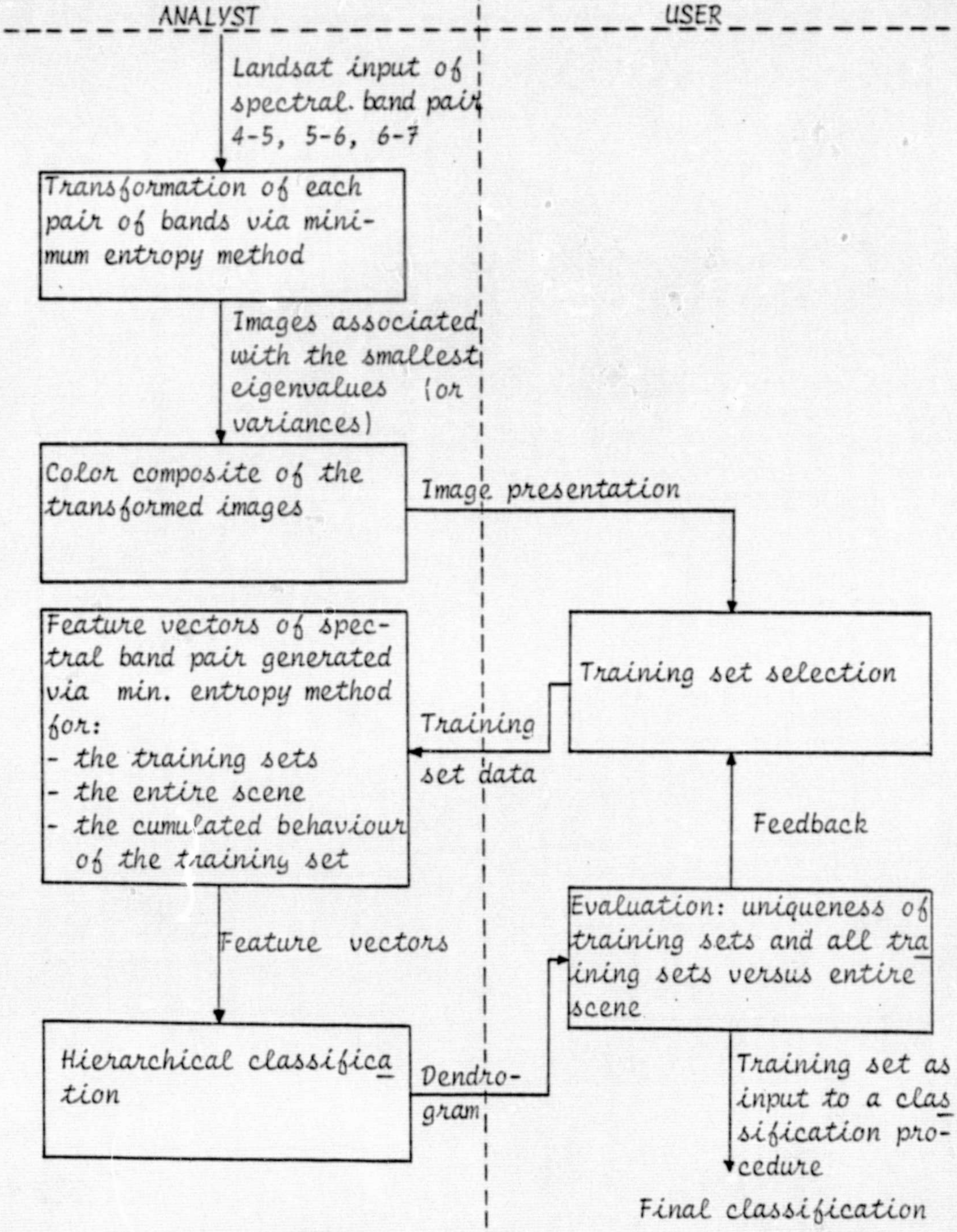


Figure 3 - Scheme of the procedure for training set selection.

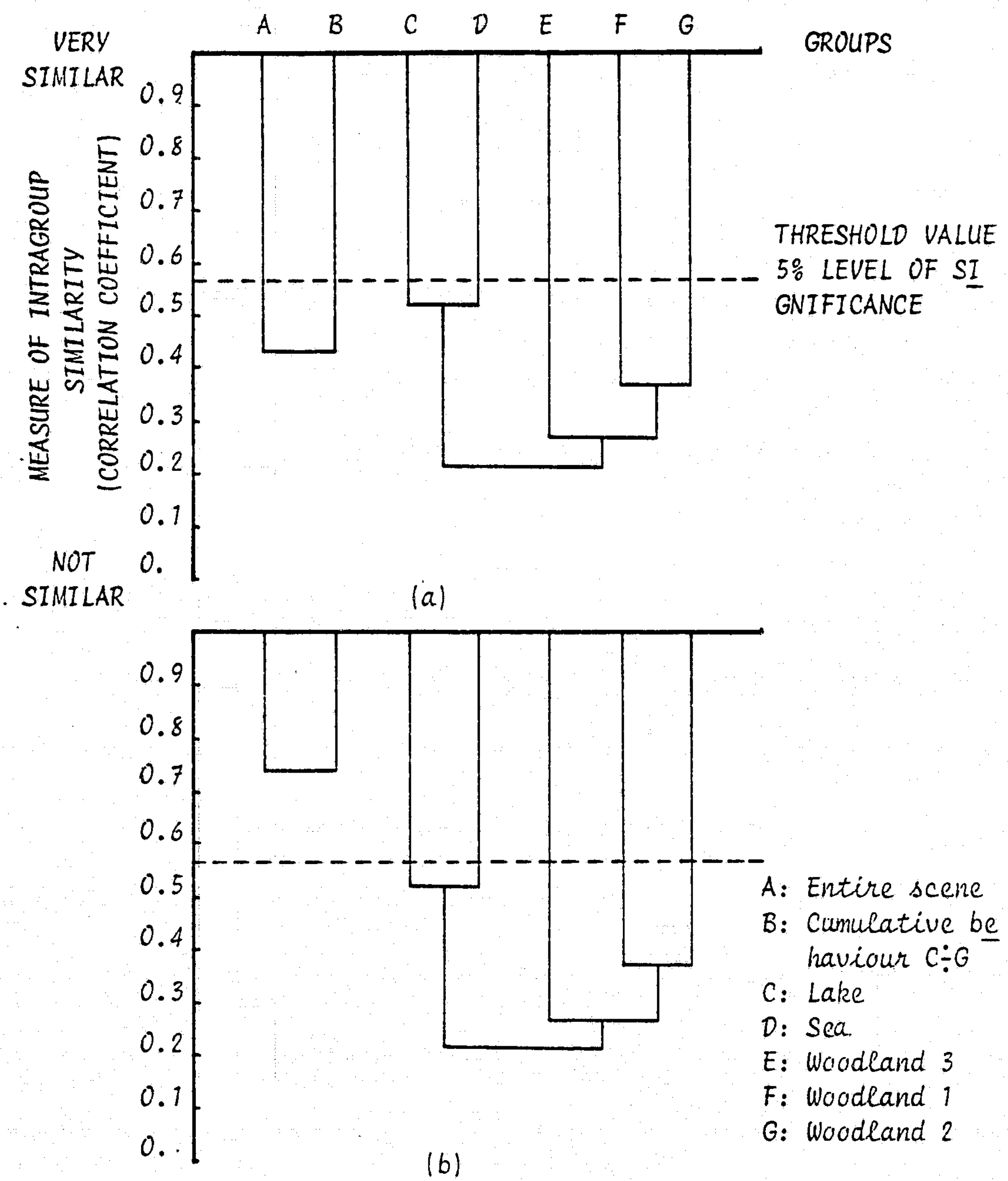


FIGURE 4 - Esemple of dendrograms obtained after the hierarchical classification.
(a) All training sets are different but not representative of the entire scene.
(b) All training sets are different and representative of the entire scene.

fied areas which are depicted in an off-white color. Thus the hierarchical classifier is reporting correctly, that the three woodland and two water classes, while different, are not representative of the entire scene. In Figure 4(b), where the dimension of the scene have been reduced to exclude unclassified areas (off-white), we find that the classes are now both different and representative of the entire scene. The feedback loop characterizing this process, which provides sets of features for review by the user, as shown in Figure 3, is easily and rapidly performed by the computer.

5. CONCLUSION

A proven pattern recognition procedure, the minimum entropy model, has been employed for processing a small portion of Landsat data. This trial was conducted to investigate the impact of the model upon the speed, clarity and accuracy of the final results when compared with a more traditional approach. The Authors have found that the minimum entropy model may provide several useful advantages over tradition techniques. First, total computer time to conduct a complete pattern recognition process is reduced. Second, subjective (transformed image), as well as statistically derived, information are made available to the analyst/user much earlier in the analysis process. A rapid feedback loop in which numerous training set combination can be tested for difference and representativeness is available.

Additional tests of Landsat data processing using the minimum entropy model are clearly justified. Data sets of differing spectral composition, drawn from other areas and other season, should be evaluated before general conclusion and recommendations.

REFERENCES

Kendall, M.G., 1972, A course in multivariate analysis. Griffin, London.

Tou, J.T., and Heydorn, R.P., 1967, Some Approaches to Optimum Feature Extraction, Computer and Information Sciences - II

Watanabe, S., 1969, Feature Selection for Pattern Recognition System, Methodologies of Pattern Recognition, Academic Press, 495-498.