Reprinted from

# Tenth International Symposium

# Machine Processing of

# Remotely Sensed Data

with special emphasis on

# Thematic Mapper Data and

# Geographic Information Systems

**June 12 - 14, 1984**

# Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

# COMPARISON OF CLASSIFICATION SCHEMES FOR MSS AND TM DATA

P.E. ANUTA, L.A. BARTOLUCCI,
D.F. LOZANO-GARCIA, J.A. VALDES,
C.R. VALENZUELA

Purdue University/Laboratory for
the Applications of Remote Sensing
West Lafayette, Indiana

ABSTRACT*

The launch of the Landsat-4 satellite in July 1982 provided the first full coverage from space of the .4-12 μm spectrum of the earth scene. In addition to the green, red, and near IR bands of the MSS, the TM provides a band in the blue, two in the middle IR, and one thermal IR. The paper describes spectral class analysis of coincident MSS and TM data to evaluate the contribution of the additional TM bands. In addition, various classifiers are available which were applied to the TM data. In the spectral class analysis, twice the number of separable classes was found in the TM data compared to the MSS data.

## I. INTRODUCTION

The Landsat-4 satellite provides 7-band spectral coverage of the .4 to 12 μm region of the spectrum. The experiment described in this paper used spectral data analysis techniques, including clustering, and separability measure calculation to determine the number of separable groups that exist in the data.

A detailed analysis procedure was applied to both MSS and TM data for a small area in Iowa. In addition, different classifiers were applied to the data to evaluate performance of the classifiers as well as to compare the results from MSS and TM.

* This work was sponsored by the NASA Goddard Space Flight Center, Greenbelt, MD, under Contract NAS5-26859.

## II. SPECTRAL CLASS ANALYSIS OF TM AND MSS DATA

The detailed spectral analysis was conducted of Thematic Mapper and MSS data for a 1,000 by 1,000 pixel area near Des Moines, IA from the September 3, 1982 data set. The surface dimension of the area is 28.5 Km square. Data were utilized from 7 blocks distributed throughout the area, which included agricultural, forest, suburban, urban, and water scene types. The blocks were processed using a clustering algorithm to produce up to 18 cluster groupings for each block. Each cluster class was then identified with a ground-cover class using aerial photography and maps of the area. The clusters from each of the 7 blocks were inspected with regard to separability, means, and variances and were either deleted or pooled with spectrally similar clusters.

The separability measure used in the transformed divergence function or processor[1] measures the statistical distance between classes based on class means and covariance matrices. The measure has a maximum value of 2,000 and the minimum of 0. Spectrally, very close classes will typically have values as low as 50 to 500.

For the TM data, initially 94 classes were defined and the pooling and deleting process reduced these to 42 finally spectrally separable classes. Table 1 lists these classes.

The MSS data were then analyzed using the same clustering and merging sequence. The number of separable classes in the MSS is 21, half of the TM result. This result is considered to be a very significant indicator of the dimensionality of TM relative to MSS. The MSS class occur-

Table 1. Spectrally Separable Classes in TM and MSS Data of Des Moines, Iowa Area.

| CLASS NO. | TM CLASS NAME | EXISTS in MSS |
|---|---|---|
| 1 | Forest1 | x |
| 2 | Forest2 | x |
| 3 | Corn1 | x |
| 4 | Corn2 | |
| 5 | Soy1 | x |
| 6 | Soy2 | x |
| 7 | Soy3 | x |
| 8 | Soy4 | x |
| 9 | Soy5 | x |
| 10 | Soy6 | x |
| 11 | Wheat residue | x |
| 12 | Grass1 | x |
| 13 | Grass2 | |
| 14 | Grass3 | |
| 15 | Soil/Veg1 | |
| 16 | Soil/Veg2 | |
| 17 | Soil/Veg3 | |
| 18 | Farm/Grass | |
| 19 | Road/Farm | |
| 20 | Baresoil1 | x |
| 21 | Baresoil2 | |
| 22 | Substation | |
| 23 | Quarry | |
| 24 | Concrete | x |
| 25 | Sludge | x |
| 26 | Industrial1 | |
| 27 | Industrial2 | |
| 28 | Urban/Hiway | x |
| 29 | Soil/Hiway | |
| 30 | Residential1 | x |
| 31 | Residential2 | x |
| 32 | Beach1 | |
| 33 | Beach2 | |
| 34 | Beach3 | |
| 35 | Soilwet1 | |
| 36 | Soilwet2 | |
| 37 | Marsh | |
| 38 | Water1 | x |
| 39 | Water2 | x |
| 40 | Water3 | x |
| 41 | Water4 | x |
| 42 | Water5 | |

rences are indicated in Table 1 by Xs in the TM class listing. The maximum divergence values of any one class with respect to all others also were much less for the MSS classes relative to TM.

Table 2 contains the minimum and average transformed divergence values for the 42 spectral classes and for the best subsets of TM spectral bands. It should be noted that the best spectral band for any combination of Bands 1 through 7 is the first middle IR band (1.55-1.75 μm). The next best band is the near IR (0.76-0.90 μm), followed by the red band and then the thermal IR. The best combination of 4 bands includes one from each of the 4 regions of the spectrum (visible, near IR, middle IR, and thermal IR).

Table 3 contains the minimum and average transformed divergence values for the best combination of MSS bands.

The high average divergence indicates that the 21 spectral classes found in the MSS were about as separable as the TM classes. However, there was twice the number of equally separable TM classes. This is considered to be the most significant result of the spectral analysis.

## III. MULTISPECTRAL CLASSIFICATION COMPARISON

A final test using a preliminary set of test data was carried out using the small amount of ground truth available. A set of 5,615 TM and 1,376 MSS pixels containing forest, corn, soybean, soil, water, and urban classes was extracted from the TM and MSS data where the cover classes were known or could be inferred from aerial photography.

The overall correct recognition was 95.7% for the TM classification using all 7 bands, 92.6% for the TM using a subset of the best 4 bands, and 67.4% for the MSS classification using all 4 bands and a per-point, Gaussian maximum likelihood (GML) classifier. The results are listed in Table 4, along with the amount of CPU time required to classify each data set.

An additional classification using all 7 TM bands was performed on the test data using a contextual algorithm, SECHO (Supervised Extraction and Classification of Homogeneous Objects), to demonstrate the effectiveness of such contextual algorithms over per-point algorithms for use with the 30-meter resolution TM data. The contextual or per-field algorithm SECHO first divides the scene to be classified into homogeneous fields and then classifies these fields using an extension of the GML algorithm[2].

Table 2. Separability (Transformed Divergence) for 42 Classes in TM Data.

| CHANNEL COMBINATIONS | DIVERGENCE MIN. | DIVERGENCE AVER. | BEST CHANNELS |
|---|---|---|---|
| 1 | 1 | 1574 | 5 |
| 2 | 210 | 1880 | 4 5 |
| 3 | 522 | 1949 | 3 4 5 |
| 4 | 1090 | 1973 | 3 4 5 7 |
| 5 | 1356 | 1979 | 3 4 5 6 7 |
| 6 | 1405 | 1983 | 2 3 4 5 6 7 |
| 7 | 1553 | 1986 | 1 2 3 4 5 6 7 |

SECHO incorporates the fact that since cover classes are more likely to occur in homogeneous areas larger than one pixel in size (i.e., larger than 30 meters), adjacent pixels are highly correlated, with the degree of correlation diminishing with an increasing distance between the pixels[2]. Thus SECHO assigns an analyst-specified threshold value below which adjacent pixels will be grouped into a homogeneous field. Statistics for these fields are calculated and compared to the original cover class statistics and a "homogeneous field" is classified as a unit into that which it most closely resembles.

The principal components transformation can be used to concentrate the variance and equivalently the information in a multidimensional data set on a minimum number of axes[3]. This enables feature selection without selecting a subset of the original dimensions and may give better results than with an equivalent size subset of original dimensions. This transformation was applied to the TM data for the test site and classifier training was repeated.

Table 3. Separability for 21 Classes in MSS Data.

| CHANNEL COMBINATIONS | DIVERGENCE MINIMUM | DIVERGENCE AVERAGE |
|---|---|---|
| 3 | 32 | 1842 |
| 2 3 | 730 | 1957 |
| 2 3 4 | 1032 | 1968 |
| 1 2 3 4 | 1112 | 1973 |

The first four principal components were classified for the 32 classes which were obtained from the training. The results are presented in the fourth column of Table 4. The classification accuracies are not as high as expected. The corn class accuracy is extremely low and grass and urban fall short of the best 4-band case. Further analysis is planned to determine the reason for this inconsistent result.

These "test" fields are limited in number of pixels and so are not really evaluating how representative the final spectral classes are of the entire scene but rather how separable the classes are. Deletion of certain spectral classes (e.g., Corn2), due to low separability, resulted in much confusion of corn with trees in the MSS but not the TM for both the best 4 and for all 7 TM bands. Also, the resolution of TM actually allowed "purer" cluster classes to be defined since smaller areas (e.g., beaches, roads) were distinct.

These results, along with the listed relative CPU time for the classifications, indicate that a subset of the best 4 TM bands incorporates the advantages of the higher spatial resolution (e.g., "purer" cluster classes) over the equivalent MSS data without simultaneously incurring as substantial an increase in computer time required for classification as with all 7 bands. In addition, these test results support the use of contextual algorithms, such as SECHO, over per-point algorithms for use with TM data. An earlier study using simulated TM data[4] demonstrated that higher classification performances could be expected with the SECHO classifier for the 30-meter TM data, especially in those areas where the cover class field sizes are relatively large relative to the scanner FOV, e.g., most agricultural areas.

## IV. CONCLUSIONS

A detailed spectral analysis of Landsat Thematic Mapper and MSS data was conducted for a limited data set. This was a single first look at these new data and the results must be considered preliminary. In general, it is clear that the very high classification accuracies achieved in the early 1970s with 12-band

aircraft multispectral scanner data are now achievable from satellite data. Actually, a 10% difference was observed between MSS results and full spectrum (i.e., TM) results in the past. A very encouraging result was obtained with the contextual classifier (SECHO) which gave the best result of all cases. This gives real evidence that the combination of spatial relationship with spectral features can have a beneficial effect on the classification process. Again, these results are from a single test at a time relatively late in the season and more analysis must be done to verify performance achievable with the TM and the MSS data.

## V. REFERENCES


1. Swain, P.H. T.V. Robertson, and A.G. Wacker. 1971. Comparison of the Divergence and B-Distance in Feature Selection. Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, IN 47906-1399. LARS Technical Report 020871

2. Kettig, R.L., and D.A. Landgrebe. 1975. Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects. LARS/Purdue. LARS Technical Report 062375.

3. Ready, P.J, and P.A. Wintz. 1972. Multispectral Data Compression Through Transform Coding and Block Quantization. LARS/Purdue. LARS Technical Report 050572.

4. Latty, R.S. 1981. Computer-Based Forest Cover Classification Using Multispectral Scanner Data of Different Spatial Resolutions. LARS/Purdue. LARS Technical Report 052081.


Table 4. Classification Accuracy and CPU Time Comparison on Test Data in Des Moines Area.

Scene ID: 40049-16264

| CLASS | TM GML Per-Point Classifier (All 7 Bands) % Correct | TM GML Per-Point Classifier (Best 4 Bands) % Correct | TM SECHO Classifier (All 7 Bands) % Correct | TM Prin. Comp. GML Per-Point Classifier (PC 1-4) % Correct | MSS GML Per-Point Classifier (All 4 Bands) % Correct |
|---|---|---|---|---|---|
| Forest | 99.0 | 97.1 | 100.0 | 98.7 | 91.2 |
| Corn | 92.0 | 76.8 | 97.7 | 46.8 | 30.8 |
| Soybeans | 100.0 | 99.8 | 100.0 | 99.8 | 99.3 |
| Bare Soil | 99.7 | 99.0 | 100.0 | 91.9 | 55.6 |
| Grass | 96.8 | 87.6 | 98.1 | 83.3 | 1.9 |
| Water | 100.0 | 96.8 | 100.0 | 100.0 | 98.9 |
| Urban | 91.7 | 99.9 | 95.8 | 86.3 | 50.2 |
| Overall | 95.7 | 92.6 | 97.9 | 88.8 | 67.4 |

| | No. of Classes | No. of Pixels Classified | No. of Bands | IBM 370/158 CPU Time Ratio |
|---|---|---|---|---|
| MSS | 21 | 250,000 | 4 | 1.00 |
| TM | 42 | 1,000,000 | 4 | 6.75 |
| TM | 42 | 1,000,000 | 7 | 19.75 |
| TM | 42 | 1,000,000 | 7 | .125* |

* on Cyber 205

AUTHOR BIOGRAPHICAL DATA

Paul E.Anuta is Associate Program Leader for Data Handling Research at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University. He received a B.S. in electrical engineering from Purdue University in 1957, M.S.E.E. from the University of Connecticut in 1962, M.S. in Computer Science from Purdue in 1967, and is a doctoral candidate at Purdue. Mr. Anuta joined the LARS staff in 1967 and has researched data handling systems for a multispectral aircraft scanner system, interferometer spectrometer, and other sensors. He is responsible for research and evaluation of remote sensor data preprocessing techniques. Key data handling research areas are image registration, geometric correction, and resolution enhancement of satellite multispectral imagery. His current interests are in the area of multitype data integration and preprocessing and analysis methods. He is a member of Tau Beta Pi, Eta Kappa Nu, the Institute of Electrical and Electronics Engineers, and the American Society of Photogrammetry.

Luis A.Bartolucci is Technical Director for Technology Transfer at LARS/Purdue University. He received his B.S., M.S., and Ph.D. in Geophysics from Purdue University. He has been involved in remote sensing research since 1969. He has played an active role in the development of remote sensing technology for applications in the area of water resources and also has made outstanding contributions in the field of thermal infrared radiation for remote sensing applications. In addition, Dr. Bartolucci has served as consultant to the U.S. Information Agency, the U.S. Agency for International Development, the Inter-American Development Bank, the U.S. Defense Mapping Agency (DMA), Control Data Corporation (CDC), International Atomic Energy Agency (IAEA), United Nations, and to several Latin American development agencies. He has been principal investigator and project director of several domestic and international research and training programs involving computer-aided processing and analysis of remotely sensed data for earth resources inventories. Dr. Bartolucci is currently involved in the development of Geographic Information Systems (GIS).

D.Fabian Lozano is a graduate research assistant at the Laboratory for Applications of Remote Sensing (LARS), Purdue University. He received his degree in biological sciences, with emphasis on remote sensing of vegetation in arid regions, from Universidad Nacional Autonoma de Mexico in 1979. He was a research staff member of the Instituto Nacional de Investigaciones sobre Recursos Bioticos in Mexico until 1981. Mr. Lozano is currently pursuing an M.S. degree in the Department of Forestry and Natural Resources at Purdue.

Jose A.Valdes A. is a graduate research assistant at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University. He had been a staff member of the Remote Sensing Department of the Direccion General de Geografia del Territorio Nacional in Mexico. As an agronomist at the Instituto Tecnologico y de Estudios Superiores de Monterrey, campus Queretaro, Mexico, he presented a Bachelor's thesis on the applications of remote sensing to agrononmic surveys. Presently he is planning to obtain an M.S. degree in agronomical remote sensing at Purdue. He is a member of the Sociedad Mexicana de Fotogrametria, Fotointerpretacion y Geodesia.

Carlos R.Valenzuela is research instructor at the Laboratory for Applications of Remote Sensing at Purdue University. He holds a B.S. degree from the Agricultural State College, Deventer, The Netherlands; M.S. from the International Institute for Aerial Survey and Earth Sciences (ITC), Enschede, The Netherlands; and is a Ph.D. candidate at Purdue University. He was Soils Investigator at the ERTS/GEOBOL Bolivian Remote Sensing Program, 1977-80, in charge of developing soil survey methodologies using Landsat MSS data. At LARS/Purdue, Mr. Valenzuela has participated as Visiting Scientist in developing a Geographic Information System for Bolivia. He also has taught a remote sensing course at graduate and undergraduate levels at Ball State University in Indiana.