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

IMPACTS OF HIGH RESOLUTION DATA ON AN OPERATIONAL REMOTE SENSING PROGRAM

J.A. MASLANIK, C.R. SMITH

Technicolor Government Services
Denver, Colorado

Abstract.--The increase in data volume associated with high resolution imagery such as TM and SPOT is a source of concern for managers of operational remote sensing programs. To assess the impact of this increased processing requirement on The Bureau of Land Management's remote sensing facility, simulated MSS, TM, and SPOT data were processed to provide system performance figures. In addition, spectral clustering measures for MSS and TM data were compared to estimate the effects of feature selection on cluster detail and variability. Results show that increased tape and disk storage requirements will be the most significant factor affecting BLM's processing system.

I. INTRODUCTION

The Bureau of Land Management (BLM) has, over the past four years, made considerable use of digital remote sensing techniques to develop, in part, geographic data bases designed to assist field personnel in management and planning decisions. Since the establishment of The Bureau`s Digital Image Analysis Laboratory in 1979, 58 Landsat Multispectral Scanner (MSS) scenes have been digitally classified and processed to provide land cover and land use information. Higher spatial and spectral resolution data such as Landsat Thematic Mapper imagery are expected to offer significant advantages over MSS data for resource mapping (Haas and Waltz, 1983, pp. 122-132). In particular, BLM anticipates using this improved spatial and spectral sensitivity to monitor range condition and to map small riparian areas important for wildlife management.

With the introduction of higher resolution data, and the associated increase in data volume, the Bureau is faced with the possibility of significantly greater processing requirements to maintain the present level of data base production. This study was performed to evaluate the effect of high resolution digital remote sensing data on the BLM's existing image processing facility and classification procedure.

II. BACKGROUND

A. PROCESSING FACILITY

The Bureau's Digital Image Analysis Laboratory (DIAL) consists of a Hewlett-Packard HP3000 minicomputer linked with an array processor, an International Imaging Systems (IIS) Model 70 display system, 800 byte/inch and two 6250 byte/inch tape drives, and 830 megabytes of disk storage. Image processing software resident on this system includes Electromagnetic Systems Laboratory's Interactive Digital Image Manipulation System (IDIMS) and International Imaging Systems' (IIS) S530 software. Output devices in use at BLM include two image displays, a Dunn camera, and several vector and raster hardcopy devices. The DIAL system operates 24 hours per day, seven days per week, and is used for both interactive and batch processing.[1]

B. CURRENT BLM PROJECT METHODOLOGY

To estimate the effects of high resolution data on BLM's digital image classification procedure, a typical mapping project which used digital classification of Landsat MSS data to delineate land cover types was chosen to serve as a baseline. This program, carried out for the Bureau's Washakie Resource Area in the Big Horn Basin of Wyoming, involved digital classification and mapping of 2.3 million hectares of range and forest land. Landsat MSS data for the Washakie area were classified using a guided clustering approach which has proven to be effective for wildland resource mapping (Rhode and Miller, 1981, pp. 74).

Prior to the classification procedure, raw imagery for the Washakie area was processed to minimize radiometric anomalies.

The resulting data were geometrically corrected and registered to a 50 meter Universal Transverse Mercator grid using a second-order polynomial transformation developed from ground control points. Following geometric rectification and classification of the Landsat MSS imagery, the resulting land cover image was used to produce color land cover maps at 1:24000 scale for each of the 106 U.S.G.S. quadrangles within the project area.

III. ANALYSIS

The impact of high resolution data on BLM`s processing system and the mapping procedure described above was assessed by quantifying the changes to be encountered in processing system workload. The measures chosen to estimate workload were on-line storage requirements, central processing unit (CPU) time, system throughput time, and system connect time.

The use of Thematic Mapper and SPOT data presents a number of feature selection options beyond those offered by MSS data for digital classification. TM data can be classified using one of a number of spectral band combinations (spectral data reduction). Both TM and SPOT data can be processed at reduced spatial resolutions which, though degraded, still exceed MSS's optimum resolution of 80 meters. As a preliminary quideline to assess the feasibility of alternate processing strategies incorporating data reduction techniques, clustering detail assessments were made using several data sets. These data sets included full spatial (30 meter) and spectral (seven band) resolution TM data to raw, four band MSS data and three band TM data registered to a 50 meter grid. System processing measurements were made for representative samples of each data type to estimate processing time and resources required for a mapping project with the areal extent of the Washakie Project Area.

A. STUDY AREA

The Washakie Project Area chosen as representative of typical BLM digital mapping projects covered 2.3 million hectares and measured 117 km in the east-west (sample direction) and 200 km in the north-south (line direction). The subset of this area used to assess the spectral clustering performance of the Landsat imagery covered the Thermopolis and Red Hole 7.5 minute quadrangles in the Big Horn Basin of Wyoming. Land cover in this area consisted of riparian species and irrigated pasture and cropland along the Big Horn River and tributary drainages, and semi-arid shrub and grassland types on the moderately dissected surrounding terrain.

B. DATA TYPES

The data types available for this study included Landsat 3 MSS data in raw and registered form and raw TM data. Acquisition dates for the MSS and TM imagery were July 19, 1981 and November 21, 1982, respectively.

In order to gauge the impact of high resolution data on the DIAL system in relation to the data quantities required to process a typical project area, data sets representing full-width coverage of the Washakie area were generated for each image type. Processing efficiency on the DIAL system for the most image processing procedures is proportional to the size of the data set in the sample direction, given band-by-band organized data. Processing of data in the line direction for these procedures is commonly handled sequentially; system performance parameters can thus be extrapolated from a subset of data in the line direction. The impact of processing the full Washakie area with high resolution data can therefore be reasonably approximated using data sets 117 km in the east-west direction and an arbitrary number of lines in the north-south direction. Data representing raw MSS (59 meter in the sample direction), raw TM (30 meter), and raw SPOT (20 meter) were generated to serve as initial preprocessing input. The number of samples required to cover a 117 km swath for each data type are as follows:

raw MSS - 2760 samples;
raw TM - 3900 samples;
raw SPOT - 5850 samples;
MSS and TM 50 meter pixels - 2340 samples.

The evaluation of clustering performance was made using MSS four-band data registered to a 50 meter grid, and TM three, four, and seven-band data registered to 30 and 50 meter grids. The choice of band subsets (bands 1, 3, 4 and bands 2, 4, 5, 7) were made based on optimum TM band combinations for vegetation mapping suggested by previous research (Teillet et al, 1981, pp. 54; Dean and Hoffer, 1982, pp. 304). Both the MSS and TM imagery were subsectioned to provide coverage of the Thermopolis/Red Hole study site.

C. SYSTEM PERFORMANCE

The sample data sets were processed through the major image processing steps carried out for the original Washakie project. This processing consisted of runs through representative functions on the IDIMS and IIS systems for the categories of preprocessing, classification, and output product generation. Functions used to represent preprocessing steps were a convolution filter to approximate banding removal and a second-order warping function to register the data to a UTM grid. Classification functions included subsampling to generate training sites, nearest-neighbor clustering, and maximum-likelihood classification. The generation of output products was represented by a scale-change function, a data transfer procedure, and three raster data plotting routines. The simulated SPOT data were not processed past the registration step. As the SPOT data sets were larger than 4096 bytes in the sample direction, the data sets would have required subsectioning to avoid hardware and software limitations on the IDIMS and IIS systems. Thus, a consistant comparison with TM and MSS data could not be made. The key system performance parameters of CPU time, system throughput, system connect time, and storage requirements for the sample processing runs are listed in Tables 1 through 4.

1.Results Preprocessing, classification, and output product generation show, in general, the expected trend of increased processing time with increased data volume. However, several deviations from this pattern were apparent. Throughput rates varied widely by function and by data type. In addition, the 30 meter to 30 meter registration of TM data showed a faster processing time than any other data set, even though the same order transformation was used and the 30 meter output required 1560 more samples than that required to represent 117 km with the 50 meter grid. Additional processing of this data set indicated that the increased processing efficiency for 30 to 30 meter registration was a function of transformation equation coefficients maintaining the same scale factor between input and output data sets.

The increases in processing time for the different spatial and spectral resolution data types, as shown in Tables 1 and 4, are relatively large. These increases, however, can be accomodated on the DIAL system as presently configured, given that the number of ongoing classification projects does not significantly increase. The most critical impact of these data types on the DIAL system occurs as a result of the 452% increase in online storage required to handle a 30 meter, four band raw TM scene compared to the storage required for a raw four band MSS scene. This added volume will greatly increase the time required for transfer to tape, which is the slowest link in the processing procedure.

Table 1. Storage Requirements by Data Type for a Landsat Scene

Data Types	Data Set Size	% increase
Raw MSS, 4 bands	31 mbytes	--
Raw TM, 3 bands	128 mbytes	312
Raw TM, 4 bands	171 mbytes	452
Raw TM, 7 bands	299 mbytes	865

Table 2. Throughput Rates in Bytes per Second

Processing Step	Throughput Rates			
Data Type	59m	50m	30m	20m
Convolution (1 band)*	16235	17123	16250	12103
Data Type	59m to 50m	30m to 50m	30m to 30m	20m to 50m
Registration (1 band)*	756	663	1548	478
Data Type	50m,4 bands	50m,7 bands	30m,4 bands	30m,7 bands
Subsampling	4255	--	7090	10500
Clustering *	6424	3532	9176	2150
Max. Likelihood Class. *	7290	3185	1695	2966
Data Type		50m,1 band	30m,1 band	
Output Products (7.5 min. quad., average of all functions)		273	598	

* array processor functions

Table 3. CPU and Connect Times in Seconds

Processing Step	CPU Time	Connect Time	CPU Time	Connect Time	CPU Time	Connect Time	CPU Time	Connect Time
Data Type	59m		50m		30m		20m	
Convolution (1 band)*	13	51	13	41	13	728	12	145
Data Type	59m to 50m		30m to 50m		30m to 30m		20m to 50m	
Registration (1 band)*	16	1568	14	1792	12	1271	14	2426
Data Type	50m,4 bands		50m,7 bands		30m,4 bands		30m,7 bands	
Subsampling	121	668	--		164	668	278	780
Clustering*	25	518	61	1391	24	714	68	4444
Max. Likelihood Class.*	15	386	17	1543	8	533	16	2790
Data Type			50m,1 band		30m,1 band			
Output Products (7.5 min. quad., average of all fuctions)			145	243	171	307		

* array processor functions

Table 4. Processing Summary (Connect Time) Extrapolated for Full Project Area

	Preprocessing		Classification		Output Products		Total
Data Types	hrs.	% of total	hrs.	% of total	hrs.	% of total	hrs.
30m, 4 bands	55.7	56.3	1.5	1.5	41.7	42.2	98.91
30m, 7 bands	97.4	66.0	8.4	5.7	41.7	28.3	147.5
50m, 4 bands	26.8	44.6	1.1	1.9	32.2	53.5	60.19

Other results show that the use of TM data with four bands requires a 64% increase in throughput compared to a 145% increase for TM with seven bands. Conversion from 30 meter data to 50 meter data during the geometric registration process reduces TM processing time by about 30%. As shown by Table 4, the machine processing time for classification is negligible compared to overall processing time. The figures in Table 4 also indicate that the use of high resolution data causes a shift in processing time emphasis from output product generation to preprocessing. Processing time for output product generation, shown in Tables 3 and 4, reveals only a slight increase between 30 and 50 meter data for a 7.5 minute quadrangle-sized area.

The inability of the DIAL system to reliably process images greater than 4096 bytes in the sample direction would result in additional system processing time required to subset and transfer the smaller data sets. Overhead required to initiate functions and to locate the subsets on storage media can be approximated as a multiplication of throughput overhead for a single image by the number of subsets required to remain within the 4096 byte limitation.

D. CLASSIFICATION PERFORMANCE

In order to better estimate the overall impact of high resolution data on system and project performance, it was necessary to assess the differences in clustering performance between full and reduced spatial and spectral resolution imagery in light of possible processing advantages incurred by using data subsets.

Due to the relatively poor acquisition date of the TM data for vegetation mapping , direct assessment and comparison of MSS and TM range land classifications using numerical accuracy assessment means were not considered appropriate. An evaluation of spectral clustering results to assess more basic performance characteristics for the MSS and TM full and reduced resolution data was, however, judged to be acceptable. Preliminary work on TM and TMS data has indicated that improved spatial resolution may reduce the number of mixed pixels between cover types (Pitts and Badhwar,1980, pp. 209). Other research, however, has shown that classification accuracy may be decreased due to an increase in the number of clusters generated from within-cover type variation (Badhwar, et.al.,1982, pp. 143). In order to evaluate the effect of varying resolution on these factors for the data types available, a set of cluster diversity measures was derived which provides an indication of cluster variability. These measures included the following: 1). the number of spectral classes generated for a sample plot; 2). the number of uniform pixel groups (three or more pixels of same class) in each plot; 3). the average number of pixels contained in the uniform pixel groups, and; 4). the number of isolated pixels in the sample plot. Items 1,2, and 3 give an indication of clustering detail, and Item 4 provides an approximation of variability due to within-cover type variation.

The Thermopolis/Red Hole study area, which measured 15.6 km by 14.8 km, was gridded into 1.2 km square plots. Five plots were randomly selected to be used as sampling sites. The following data sets were clustered using a nearest-neighbor clustering algoritm: four band MSS data registered to a 50 meter grid, registered 50 meter three, four, and seven band TM data, and registered 30 meter registered three, four, and seven band TM data. Each data set was clustered using an identical set of clustering parameters.

The diversity measures for the resulting clustered images were obtained by locating the same five sample plots on each image. Through a combination of manual and digital methods, the number of unique spectral class occurences, number of isolated pixels, number of uniform groups, and number of pixels per group within each sample plot were determined. The image identities were masked throughout this process. Figure 1 represents raw averages from the five plots for the diversity measures over the sample plots for the seven data sets. In Figure 2, the totals were normalized to account for the greater number of pixels in the 1.2 km plots for the 30 meter data. MSS results were: number of classes = 13, number of isolated pixels = 10, number of groups = 50, and number of pixels per group = seven.

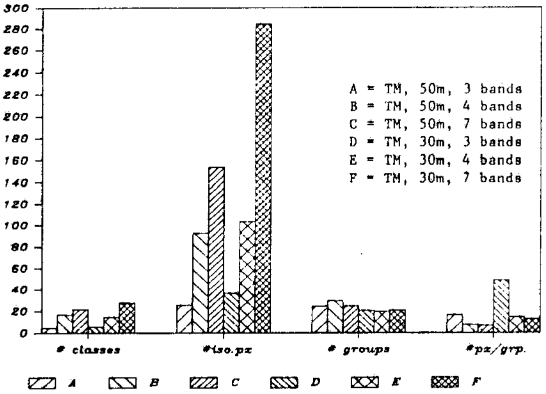

Figure 1. Average Raw Diversity Values

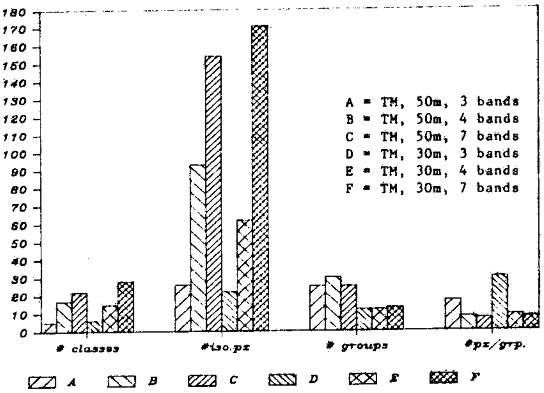

Figure 2. Average Normalized Diversity Values

Comparisons of significance of normalized mean differences between 30 and 50 meter data and between three, four, and seven band data were carried out using a Wilcoxon paired observation procedure (Walpole and Myers, 1972, pp. 254).

1. Results The results from the cluster diversity measures show that the number of classes and number of isolated pixels increased as more spectral information was included. The effect of increased spatial resolution on clustering detail was less pronounced. The number of occurences of uniform groups was not statistically different between the TM three, four, and seven band data sets, although the average size of these groups decreased significantly from the three band to the four and seven band data sets.

The comparison of diversity measures for the 30 meter versus 50 meter resolution TM data showed that the change in isolated pixels was not statistically significant. However, the number of spectral classes increased in the 30 meter data, while the number of uniform groups and size of these groups decreased compared to the 50 meter data.

The MSS results suggest a greater clustering uniformity than that shown be the TM data, as indicated by the lower number of spectral classes and the lower number of isolated pixels identified. The larger number of relatively small groups may be a result of mixed pixels and, therefore, may represent a larger-scale version of the within-class noise exhibited as isolated pixels in the higher resolution TM data. More precise analysis of these comparisons between TM and MSS data requires that the variation due to surface condition be minimized by using data sets collected at more similar dates.

D. DISCUSSION

The results of both the system processing comparisons and the clustering diversity measures show that the use of higher resolution data will have a significant impact on the planning and operation of digital land cover mapping projects at BLM's Digital Image Analysis Laboratory. The greatest impact will occur due to significant increases in data storage and transfer requirements. Additional processing time required for image preprocessing and classification can be handled using the system as currently configured. The distribution of processing emphasis and concern for the three major areas (preprocessing, classification, output products) will be shifted from output products to preprocessing.

The analysis of cluster diversity using sample plot analysis revealed no significant difference in cluster detail or uniformity between 30 and 50 meter TM data. Differences were apparent, however, between data sets of different spectral resolution. The

MSS data showed strong deviations from the patterns exhibited by the TM imagery.

The system evaluation shows that some flexibility exists in the ability of the DIAL system to process different data types. Significant savings can be made in data storage and transfer time by using reduced spatial and/or spectral resolution imagery; however, the system has the capacity to process smaller project areas or project area subsets at maximum resolution. The analysis of clustering performance for data of various resolutions did not conclusively identify an optimum TM data set for rangeland mapping. Rather, the results indicated that different combinations of spectral and spatial resolution should be considered based on the detail of the land cover classification desired. This conclusion, combined with the system performance evaluation, suggests that the project manager using a minicomputer-based image processing system will have to weigh the tradeoffs between the classification detail best suited for his project area and the disadvantages of increased data storage and transfer requirements.

Although research has indicated that increased resolution imagery such as TM data offer advantages over MSS for rangeland mapping, additional study is required to identify the utility of TM and other high resolution imagery for specific applications requiring different degrees of spatial and spectral resolution. The effects of varying degrees and types of feature selection on classification performance requires further study. In particular, the optimum feature selection schemes for different mapping applications must be defined to permit analysts to make the best use of limited computing power. These data reduction schemes, such as band selection and principal components, should be evaluated in terms of computing time required to reduce the data volume versus the time saved in subsequent processing. The utility of data reduction in classification methods also be considered in terms of large-area operational projects.

IV. REFERENCES

Badhwar, G.D., K.E. Henderson, D. E. Pitts, W. R. Johnson, M. L. Sestak, T. Woolford, and J. Carnes, 1982. A Comparison of Simulated Thematic Mapper Data and Multispectral Scanner Data for Kingsbury County, South Dakota, Machine Processing of Remotely Sensed Data Symposium. pp. 143.

Dean, M. E. and R. M. Hoffer, 1982. An Evaluation of Thematic Mapper Simulator Data for Mapping Forest Cover, Machine Processing of Remotely Sensed Data Symposium. pp. 304.

Haas, R. H. and F. A. Waltz, 1983. Evaluation of Thematic Mapper Data for Natural Resource Assessment, Pecora VIII, Sioux Falls, SD. pp. 122-132,

Pitts, D. E. and G. D. Badhwar. 1980. Field Size, Length and Width Distribution Based on LACIE Ground Truth Data. Remote Sensing of the Environment, Vol. 10, pp. 209.

Rohde, Wayne G. and Wayne A. Miller, 1981, Arizona Vegetation Resource Inventory (AVRI): Final Report U.S.G.S. Eros Data Center, Sioux Falls SD, pp. 74.

Teillet, P. M., B. Guindon, D. G. Goodenough Forest Classification Using Simulated Landsat-D hematic Mapper Data. Canadian Journal of Remote Sensing, Vol 7,#1, pp. 54.

Walpole, Ronald E., Raymond H. Myers, 1972, Probability and Statistics for Engineers and Scientists, MacMillan Co., pp 254.

---

1 Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the Bureau of Land Management, U.S. Department of Interior, or by Technicolor Government Services, Inc.

AUTHOR BIOGRAPHICAL DATA

James Maslanik is an Applications Scientist currently involved in rangeland mapping using satellite imagery. He received his B.S. in Forest Science and Masters in Environmental Pollution Control from The Pennsylvania State University and is pursuing a Ph.D. in geography at The University of Colorado.

Charles Smith is a Systems Analyst and specializes in software engineering and systems design in the areas of digital image analysis and geographic information systems. He received his B.S. degree in Civil Engineering from Purdue University and is currently pursuing a masters degree in data processing from the University of Denver. Before joining Technicolor Government Services he was a systems engineer with the Laboratory for Applications of Remote Sensing (LARS).