

HIERARCHICAL CLASSIFICATION OF MULTITEMPORAL/MULTISPECTRAL SCANNER DATA

D. F. LOZANO-GARCIA, ROGER M. HOFFER

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

ABSTRACT

Most techniques for computer classification of multispectral scanner data involve a "single-stage" approach in which every pixel in the data is classified in a single step, using a single set of training statistics and a single set of wavelength bands. Hierarchical classifiers, on the other hand, involve a sequence of classification steps, each of which can involve a different wavelength band or combination of wavelength bands. In addition, at each step in the classification process only one spectral class or a specific group of spectral classes are separated from all other classes in the data. Since a relatively small number of wavelength bands are involved at each step, and (after the initial step) only a portion of the data is being classified at each step, such hierarchical classifiers are computationally very efficient. However, as compared to single stage classifiers, the effectiveness of hierarchical classifiers in terms of classification accuracy is not clear, especially when dealing with multi-temporal data sets.

In this study, Landsat-1 MSS data sets obtained in June 1973 and February 1974 over the Monroe Reservoir and Hoosier National Forest in central Indiana were used. After digitally registering the two data sets, four classification procedures were compared. The first consisted of a standard single stage maximum likelihood classification using an eight channel training statistics deck (four wavelength bands from each two dates). The second utilized the 4 best channels of the 8 available. The third involved the Layered hierarchical classifier and the same eight channel training statistics. The fourth approach utilized the Layered classifier again, but the data from the two dates were treated independently for the purpose of developing training statistics.

The results indicate that the Layered classifier is a more effective and efficient approach for classification of multitemporal/multispectral scanner data. The classification accuracies were relatively high for all four classifications, but the Layered classifier required only one third of the CPU time used in the single stage classification.

I. INTRODUCTION

One of the most important characteristics of the Landsat system is the repetitive coverage over the same portion of the earth --18 days for Landsat 1,2 and 3, or 16 days for Landsat 4 and 5. These satellites also provide a synoptic view of the earth's surface and digital multispectral scanner data availability. These characteristics provide the resource manager with the potential for monitoring the areas where ground cover is subjected to changes either by nature or man-made activities, and provide the capability to analyze the phenology of natural and cultivated vegetation.

The characteristics of the Landsat system are effective for multitemporal analysis because:

- a) Data is obtained at the same time of the day, minimizing sun angle effects.
- b) The spectral response of cover types is recorded in a consistent way, i.e. in the same wavelength bands, and for resolution elements of the same size, and also having minimal difference in look angle.
- c) The format of the data provides the capability for both visual and digital analysis.

Multivariate classification of MSS has been performed traditionally with single stage classification algorithms (Gaussian Maximum Likelihood, Minimum Distance, etc.), which involve the categorization of the data using only one of the many available sets of spectral channels. The large number of computations required (in the single stage GML classification procedure) or the sensitivity of the classifier (Minimum Distance and Parellelepiped) to category variance are the primary limitations of these approaches.

Another approach for performing a classification is the layered or decision tree approach, in which multispectral scanner data is classified through a hierarchical decision procedure. In this approach, the analyst goes through a series of decision nodes, and at each node he/she selects the best set of spectral channels that separate a class or relatively small group of classes from certain other cover types of from everything else. This separation is made step by step, until all classes have been separated, in the layered "decision tree". The decision tree simply defines which combination of wavelength bands are to be used at each node, and which cover types will be separated from which other cover types at that node. An example of this approach is shown in Figure 1. The actual classification of each pixel is performed using any of the available per-point algorithms (GML, Minimum Distance, etc.).

The layered classifier is designed to solve some of the limitations of the single stage classifiers (GML, Minimum Distance, etc.). These limitations include:

- 1) Only one of the many possible combinations of wavebands (features) is used in the single stage classification. Very often, analysts simply utilize all available wavelength bands in order to keep from inadvertently omitting an important band. However, as pointed out by Duda and Hart (1973), "...it has been frequently observed in practice that, beyond a certain point, the inclusion of additional features lead to worse rather than to better performance". Latty and Hoffer (1981), working with simulated Thematic Mapper data, found that no more than four spectral channels were required to satisfactorily classify the data set, and that additional channels did not improve the classification performance.

Thus, there may be some subsets which are more effective than the entire set of spectral channels, or some subsets will be more effective for the discrimination of particular spectral classes.

- 2) In the standard single stage GML classification technique, every data sample (pixel) is tested against all classes. This characteristic leads to very low efficiency because all wavelength bands being utilized and every spectral class present in the data is involved in the classification of each pixel. For example, each classification decision might involve 8 wavelength bands and 25-35 or more spectral classes. However, by using a sequential decision making process as is used in the layered technique, each classification decision often involves only 1 to 3 wavelength bands and only 3 to 5 spectral classes. Therefore, each classification decision is much faster, thus providing relatively high computational efficiency.

II. BACKGROUND

Multitemporal/Multispectral classification involves a single classification of a data set consisting of data obtained on two or more dates. In the case of a two date set of Landsat MSS data, the combined dates produce a data set with eight channels. The data from the two dates are digitally registered, and the combined data set is analyzed using the standard pattern recognition principles that are normally applied to single date data sets. The training statistics are generated either by the supervised or unsupervised method. An advantage of the multitemporal/multispectral classification approach is that it can be accomplished in only a single classification step. Conversely, the derivation of training statistics can be an intricate task, considering the spectral changes that may have occurred between the two dates, as well as the complexities of the study area. Also, when using a combination of images from the same sensor, a channel redundancy can be found and this can increase the processing cost without an increase in the accuracy of the results.

Williams and Yates (1977) used a single stage classification approach with multitemporal/multispectral data. A supervised method was used for developing training

statistics, and the data were classified with a Euclidean distance classifier. Their results show an improvement in the classification of residential, woodland, and lawns (as compared with single date classifications); however, unacceptable levels of agricultural/residential classification errors remained. Using two Landsat MSS images, representing summer and winter conditions in North Carolina, Williams (1979) also employed the single stage technique with multitemporal/multispectral data for classifying an area of hardwood and pine forest. His results show a good agreement for both types of forest. To study the area of Tokyo bay, Hang and Itsaka (1982), employed three Landsat MSS images for assessing land-use changes, using supervised training statistics and a maximum likelihood classifier. Their results show the percentage of reclaimed land in the period 1972-1980.

Hierarchical classifications have not been used commonly in the analysis of remotely sensed data. One of the earliest applications of this scheme was the work of Bartolucci et al. (1973) in mapping water temperatures. The authors used this approach to separate water from all other cover types in the first stage of the decision tree, using the best set of channels to accomplish this separation. The second step was to classify the water temperatures using the available thermal channel. They found this approach to be superior to the use of the single stage classification procedure. Hoffer et al. (1979) used the layered classifier in a study of combined multispectral scanner and digital topographic data. Their approach was to separate cover types using the spectral data in the first stage. The next stage of the classification involved the utilization of topographic data to divide the major cover types into individual forest cover types, and to remove some misclassification errors due to shadow effects. The results showed an improvement of approximately sixteen percent as compared to the use of spectral data alone using a single stage classification.

Application of the layered classifier to a multitemporal data set has also been accomplished by Landgrebe (1976), classifying a set of agricultural areas. Weismiller et al (1977) applied this technique in a change detection procedure. Hixson et al. (1980) compared this classifier against 6 others. Their conclusions indicate that the time required by the analyst

to design the decision tree can be significant. Also they conclude that this scheme is well suited to handle multitemporal MSS data sets.

III. STUDY AREA AND DATA USED

The test site for this study is located in the south-central portion of the state of Indiana, about 50 miles south-southwest of Indianapolis. It is located in the Interior Low Plateau province, in the unglaciated portion of Indiana. The area is well drained by a medium-fine dendritic drainage system (Lindsay et al. 1969). The forest of the area is classified according to Petty and Jackson (1966) as:

WESTERN MESOPHYTIC ASSOCIATION.- In this community, frequently 10 to 20 species share dominance in the crown cover and exert their controlling influence on the forest community.

The mixed forest usually occurs in ravines and on the cooler slopes, whereas oak or oak-hickory forests cover the drier slopes and ridges.

OAK-HICKORY ASSOCIATION (Quercus-Carya).- The oak-hickory forests are usually found occupying south-facing and west-facing slopes. In general, moisture content of the soil is consistently lower than in the oak-hickory type than the mesophytic mixed forest.

The study area is largely forested, mostly by tulip poplar, oak, maple, hickory, ash, walnut and sycamore. Small stands of pine are scattered in the study area. The steep slopes and heavily dissected topography have discouraged the extensive clearing of this area for agriculture, although selective logging has altered the composition of most stands. On the valley bottoms in the western, south-western and east central portion of the area the forest has been substituted for crops and pasture. Some portions of suburban Bloomington occur in the north-western corner of the area. Monroe Reservoir, Lemon Lake, Yellowood Lake and Grandview Lake are the major water bodies in the area.

Seven dates of registered multispectral scanner images of the Landsat-1 satellite were available for a portion of the Hoosier National Forest including the Monroe Reservoir and surrounding areas.

Scene ID	Date	Season
1285-16001	May 4, 1973	Spring
1320-15541	June 8, 1973	Late spring
1392-15531	August 19, 1973	Summer
1411-15584	September 7, 1973	Late summer
1482-15514	November 17, 1973	Late fall
1572-15493	February 15, 1974	Winter
1591-15550	March 8, 1974	Late winter

The data sets of June 1973 and February 1974 were selected due to their high quality and minimum cloud cover.

IV. ANALYSIS PROCEDURES USED

Table 1 shows the four different classification techniques utilized in this work. They can be divided into single stage vs layered classifiers. In the single stage approach two methods were tested --one with all 8 channels and the other with the four best channels. The layered approach was first tested using the same training statistics generated for the 8 channel, single stage approach. The second generating two sets of independent training statistics --one for each date.

Table 1 Classification Techniques Used.

Classification Technique	Single Stage		Layered	
	All 8 A	"Best" 4 A	Combined Dates (All 8 A)	Separate Dates (4 A June, 4 Feb.)
Statistics used for training	All 8 A	"Best" 4 A	Combined Dates (All 8 A)	Separate Dates (4 A June, 4 Feb.)
Algorithm used for classification	GML	GML	GML	GML

The first analysis conducted in this study involved classifications of a two date, eight channel data set. The selected dates were June of 1973 and February of 1974. These dates were selected on the basis of the reference data (primarily used for the interpretation and evaluation procedures).

Training statistics were generated

using a "multi-cluster blocks" approach, similar to the one described by Fleming et al (1975). The available tools for this type of analysis are more restricted in comparison with those available for the single date classification since only separability values (transformed divergence) between class pairs and a coincident spectral plot can be used with this type of data.

Two different classifications were performed with this data set. The first used all the channels of both dates (Multitemporal/Multispectral 8 channels). For the second classification, four channels were selected based upon the minimum and average transformed divergence value, calculated by the separability processor of LARSYS. The selected channels were:

Waveband μm	Date	Spectral Region
0.6-0.7	June	Visible (red)
0.7-0.8	June	Reflective IR
0.6-0.7	February	Visible (red)
0.7-0.8	February	Reflective IR

It is important to point out that one channel in each major portion of the spectrum (i.e. visible and reflective infrared) covered by Landsat-MSS was selected for this classification.

One of the characteristics of the layered classifier is that it permits the analyst to optimize the decisions (use of certain spectral bands) in the separation of a class or group of classes. In a multitemporal/multispectral scanner classification, this algorithm also permits the use of the best season (represented by a set of spectral channels) for the identification and separation of cover types.

The first approach used with this classifier involved using the same 8 channel training statistics developed for the single stage Multitemporal/Multispectral classification. Next, a set of training statistics was developed independently for each of the two dates and used as part of the input to the layered classifier.

The selection of the classes that will constitute a particular node, and the set of spectral channels to be used to separate this node were based on the separability information. This was obtained by

calculating the transformed divergence values of the training classes for all possible combinations of spectral channels. The best set of features to be used in each particular node was also defined using the separability information, based on a threshold of $Dt=1750$.

To determine the accuracy of the classification of MSS data, a set of statistically valid test fields was developed. The standard color IR composite was displayed on the Comtal Vision One/20 and a test grid with dimensions of 50 lines by 50 columns was selected as the basis for the sampling procedure, over the area where aerial photographs were available. The cell to be analyzed was selected at random. Each cell of 2,500 pixels was subdivided into four quadrants of 25 lines by 25 columns (625 pixels). One quadrant was selected at random, and the biggest, most homogeneous field corresponding to each specific cover type present in that block was identified. The minimum number of observations (pixels) that were used for the evaluation of any particular cover type was 100 pixels (Landgrebe 1976).

Following information provided by Anderson (1972), the statistical evaluation was done with the arcsine transformation of the performance values due to the nature of the results -- a proportion dealing with binomial data (pixels are identified correctly or incorrectly). The tests were done using one-factor analysis of variance.

To determine if there were significant differences between the performance values of cover types or classifications, a Newman-Keuls Range test was performed at an alpha level of 0.1. This test allows the analyst to distinguish differences between means (performance values) in a sequential manner, thereby achieving a ranking of the classification results.

The criterion used for determining the cost effectiveness of the classification results was based on the amount of computer CPU time (Central Processing Unit) used to perform each classification. This was considered the most objective and accurate way to compare and evaluate the cost of each classification scheme. Because the analyst become increasingly familiar with the characteristics of the data during the sequence of analysis, it was believed that the "analyst time" required to develop the training statistics would be biased.

V. RESULTS

The overall, average and per-class performances were obtained using the *PRINTRESULTS processor of LARSYS. Total CPU time required for each classification and the overall and average performances are shown in Table 1.

Table 2.- CPU Time and Overall Classification Performance.

	CPU TIME	NO CLA SS	OVER- ALL PERFOR MANCE	AVER- AGE PERFOR MANCE
MULTITEMPORAL/ MULTISPECTRAL 8 Channels	77.3 min	14	91.0 %	83.0 %
MULTITEMPORAL/ MULTISPECTRAL 4 channels	25.4 min	14	90.8 %	83.2 %
LAYERED 1 set of 8 Channels	16.5 min	14	90.8 %	82.3 %
LAYERED 2 SETS OF 4 CHANNELS	9.1 Min	20	91.8 %	89.9 %

The Multitemporal/Multispectral single stage classifications (both eight channels and the four best channels) provided results showing detailed informational classes. For the deciduous forest, a class representing forest in shadow was identified. Bare soils were differentiated into two groups: those that are subject to flooding and those that not. Two distinctive classes of water (deep water and shallow water) were recognized in Monroe lake.

Figure 2 shows the the classification performance by class For the single stage classifications, both overall performance and the performances by class were very good, except for the class "Pasture" which had an accuracy of 31.5 % for the 8 channel classification and 32 % for the 4 channel classification, due to confusion between the pasture and soil classes. 58 % of the pasture test pixels were classified as soils in the 8 channel classification, and 50 % of the pasture test pixels of the 4 channel classification were assigned to the soils classes.

In the layered classifier the primary concern in the design of the decision tree was to obtain an adequate separation of the coniferous forest from all other classes, since the June data resulted in low accuracy for this class but was good for all other, and the February data had good accuracy for the coniferous class.

The two date layered classification had the best overall performance of all four classifications, with 91.8 %. Both Deciduous and Coniferous forest cover types had over 90 % correct classification. Most of the misclassifications in each of these forest classes were actually due to confusion occurring between them rather than between forest and non-forest categories.

Also, a more consistent classification was obtained for the other classes --over 80 % as can be seen on Figure 2. The statistical comparisons of the percent correct classification for the five classes show three groups (Water, Forest and Soils-Grasslands) in which there were no significant differences in the classification accuracy.

It was clear that the classification was improved due to the capabilities of the layered classifier. Using this classification processor, the analyst can select the best set of features to separate a class or group of classes.

The second classification using the layered technique (i.e. training statistics based on one set of 8 channels) showed no significant difference from the results obtained in the single stage Multitemporal/Multispectral classifications. A small decrease in the percent accuracy for the pasture class was found, but this difference was not statistically significant. The primary difficulty in classification of pasture was again due to confusion with the soil class. However, the CPU time required was only 60 % (10 minutes less) of that required for the single stage Multitemporal/Multispectral 4 channel classification.

As shown in Table 2, the overall classification performance varied only from 90.8 % to 91.8 %. However, in the first three classification results shown in Table 2, the average performance values indicate an important variability in relation to the overall performance values.

Of the four classifications, the Multitemporal/Multispectral 8 channel classifi-

cation required the highest amount of CPU time, followed by the Multitemporal/Multispectral 4 channel classification, then the Layered 1 set of 8 channels, and with the lowest CPU time of all four, the Layered 2 sets of 4 channels classification. Thus, based on both accuracy and CPU time, the Layered 2 sets of 4 channels approach was the best method.

VI. CONCLUSIONS

The results of this research show the advantage of the Layered classification approach over the Multitemporal/Multispectral classification approach in the analysis of Multitemporal MSS data.

The single stage Multitemporal/Multispectral approach provided an accuracy of over 98 % in the differentiation of forest versus non-forest classes. In addition, an accuracy of over 95 % was obtained in the separation between Coniferous and Deciduous Forest. The weak points of this scheme are: The complexity in the development and interpretation of the training statistics and the CPU time required to perform the classification. The eight channel classification required 8.5 times more CPU time than the best layered classification, and even the four channel Multitemporal/Multispectral classification required 2.8 times more CPU time than the best of the layered classification. The selection of the four best channels of this data set indicates that one channel of each of the available regions of the electromagnetic spectrum in the MSS for each date are required to perform an effective multitemporal classification.

The layered Classification procedure proved to be the best in terms of classification accuracy, with 99 % for the forest classes combined and 90 % for the non-forest classes (excluding water), for both the Layered 2 sets of 4 channels and Layered 1 set of 8 channels. Although percent accuracies in the Layered 2 sets of 4 channels for the individual forest cover types were slightly lower in relation to the Multitemporal/Multispectral Classifications, these differences were not statistically significant. Pasture classes showed an improvement, using the same comparison between techniques. The layered technique also provided more consistent results, since all accuracies were over 80 %. The design of the decision tree for the classification is one of the most

important and difficult tasks in this approach.

In summary, the combination of high classification accuracy, low CPU time required and the flexibility in handling multitemporal data sets makes the Layered classifier a very effective, efficient and useful tool in multitemporal analysis of remotely sensed data.

VII. REFERENCES

- Anderson, J.R., E.E. Hardy, J.T. Roach & R.E. Witmer. 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Geological Survey Professional Paper 964. Washington D.C.
- Anderson, V.L. & R.A. McLean. 1974. Design of Experiments: A Realistic Approach. Marcel Dekker, Inc. New York. 418 pp.
- Bartolucci-Castedo, L.A., R.M. Hoffer & T.R. West. 1973. Computer-aided Processing of Remotely Sensed Data for Temperature Mapping of Surface Water from Aircraft Altitudes. LARS Information Note 042373. LARS-Purdue University, West Lafayette, Indiana 47906. 143 pp.
- Bartolucci, L.A., P.H. Swain & C. Wu. 1976. Selective Radiant Temperature Mapping Using Layered Classifier. IEEE Trans on Geoscience Electronics GE 14(2):101-106.
- Duda, R.O. & P.E. Hart. 1973. Pattern Classification and Scene Analysis. John Wiley & Sons. New York. 482 pp.
- Fleming, M.D. J.S. Berkebile & R.M. Hoffer. 1975. Computer-aided Analysis of Landsat-1 MSS Data: A Comparison of Three Approaches, Including a "Modified Clustering" Approach. LARS Information Note 072475. LARS-Purdue University, West Lafayette, Indiana 47906.
- Hixson, M., D. Scholz, N. Fulls & T. Akiyama. 1980. Evaluation of Several Schemes for Classification of Remotely Sensed Data. Photogrammetric Engineering and Remote Sensing 46(12):1547-1553.
- Hoffer, R.M., 1979. Computer-aided Analysis Techniques for Mapping Earth Surface. LARS Technical Report 020179. LARS-Purdue University, West Lafayette, Indiana 47906. 20 pp.
- Hong, J-K. & J. Lisaka. 1982. Coastal Environment Change Analysis by Landsat MSS Data. Remote Sensing of Environment 12(2):107-116.
- Jensen, J.R. 1981. Urban Change Detection Mapping Using Landsat Digital Data. The American Cartographer 8(2):127-147.
- Latty, R.S. & R.M. Hoffer. 1981. Waveband Evaluation of Proposed Thematic Mapper in Forest Cover Classification. LARS Technical Report 041581. LARS-Purdue University, West Lafayette, Indiana 47906. 12 pp.
- Lindsay, A.A., D.V. Schmelz & S.A. Nichols. 1969. Natural Areas of Indiana and Their Preservation. Indiana Natural Areas Survey. Lafayette, Indiana 594 pp.
- Petty, R.O & M.T. Jackson. 1966. Plant Communities. p:264-296 in: Lindsey A. A. (editor) Natural Features of Indiana. Indiana Academy of Science.
- Swain, P.H. 1976. Layered Classifier Adapted to Multitemporal Data sets. LARS Final Report, NASA Contract NASA-14016. LARS-Purdue University, West Lafayette, Indiana 47906.
- Swain, P.H. & H. Hauska. 1977 The Decision Tree Classifier: Design and Potential. IEEE Transactions on Geoscience Electronics GE 15(3):142-147.
- Swain, P.H., C.L. Wu, D.A. Landgrebe & H. Hauska. 1975 Layered Classification Techniques for Remote Sensing Applications. LARS Information Note 061275. LARS-Purdue University, West Lafayette, Indiana 47906. 11 pp.
- Weismiller, R.A., S.J. Kristof, D.K. Scholz, P.E. Anuta & S.A. Momin. 1977. Change Detection in Coastal Zone Environments. Photogrammetric Engineering and Remote Sensing 43(12):1533-1539.
- Williams, D.L. 1979. Canopy Closure in the Pine Forest of North Carolina, p:19-22 in: Williams, D.L. & L.D. Miller. Monitoring Forest Canopy Alterations around the World with Digital Analysis of Landsat Imagery. NASA, Goddard Space Flight Center.

Williams, D.L. & F. Yates. 1977. A Reduction in Ag./Residential Signature Conflict Using Principal Components Analysis of Landsat Temporal Data. Proceedings American Society of Photogrammetry Fall Technical Meeting. Little Rock, Arkansas. p 230-238.

Wu, C.L., D.A. Landgrebe & P.H. Swain. 1974. The Decision Tree Approach to Classification. LARS Information Note 090174. LARS-Purdue University, West Lafayette, Indiana 47906. 174 pp.

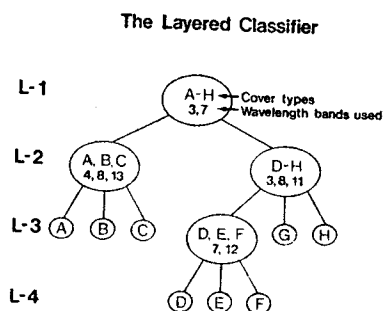


Figure 1. Hypothetic Decision Tree.

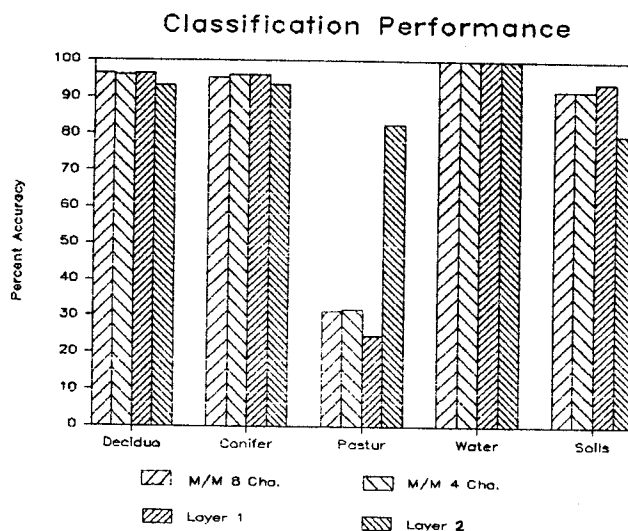


Figure 2. Classification Performance by Cover Type.