(E82-10033) APPLICATION OF REMOTE SENSING TO STATE AND REGIONAL PROBLEMS Semiannual Progress Report, 1 Nov. 1980 - 30 Apr. 1981 (Mississippi State Univ., Mississippi State.) 106 p HC A06/MF A01 CSCL 05B G3/43

N82-19609

Unclas 00033

SEMI-ANNUAL PROGRESS REPORT NO. 15

November 1, 1980 to April 30, 1981

NASA Grant NGL 25-001-054

Submitted To

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Office of Space and Terrestrial Application
Technology Transfer Division
Washington, D. C.

Submitted By

W. Frank Miller*
John S. Powers
Jon R. Clark
Jimmy L. Solomon
Sidney G. Williams

MISSISSIPPI STATE UNIVERSITY
P. O. Drawer FD
Mississippi State, MS 39762

April 30, 1981

*Program Coordinator

SEMI-ANNUAL PROGRESS REPORT NO. 15

November 1, 1980 to April 30, 1981

NASA Grant NGL 25-001-054

Submitted To

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Office of Space and Terrestrial Application
Technology Transfer Division
Washington, D. C.

Submitted By

W. Frank Miller*
John S. Powers
Jon R. Clark
Jimmy L. Solomon
Sidney G. Williams

MISSISSIPPI STATE UNIVERSITY
P. O. Drawer FD
Mississippi State, MS 39762

April 30, 1981

*Program Coordinator

TABLE OF CONTENTS

SEMI-ANNUAL PROGRESS REPORT NO. 15

November 30, 1981 to April 30, 1981

APPLICATION OF REMOTE SENSING TO STATE AND REGIONAL PROBLEMS

I. INTRODUCTION

The major purpose of the Remote Sensing Applications Program is to interact with units of local, state, and federal government and to utilize Landsat data to develop methodology and provide data which will be used in a fashion such that a concrete, specific action will be taken by the cooperating agency. The attainment of this goal is dependent upon identification of agency problems which are immediate in nature, and subject to at least partial solution through the use of remotely sensed data.

Other subsidiary objectives include the development of a trained staff from the faculty of Mississippi State University who are capable of attacking the varied problems presented by the respective state agencies; the training of students in various University academic courses at both the undergraduate and graduate levels; the dissemination of information and knowledge through workshops, seminars, and short courses; and the development of a center of expertise and an operational laboratory for training and assistance to cooperating agencies.

II. GENERAL PROGRAM PROGRESS

The past six months have been spent largely on activities with the U. S. Forest Service, U. S. Corps of Engineers, the State Tax Commission, the U. S. Fish and Wildlife Service, and the Mississippi Department of Wildlife Conservation. The activity has been largely educational, i.e., informal meetings to explore possibilities of employing remote sensing technology in agency problems. Appreciable staff time has also been devoted to software development in support of the proposed new projects, and in training sessions. Two short courses have been held primarily for local, state, and federal agency personnel. Participants included representatives of the State Tax Commission, the Corps of Engineers, forest industry, the Mississippi Department of Wildlife Conservation, the U. S. Fish and Wildlife Cooperation Unit, and the Mississippi Mineral Resources Institute.

Proposals which have been prepared are as follows:

Agency	Native of Project	Funding
U. S. Forest Service So. For. Expt. Sta.	Vegetative Mapping of Puerto Rico	Cooperative
U. S. Fish & Wildlife Serv.	Comparison of Landsat and Aircraft Data in Wetland Discrimination	Demonstration
Miss. Tax Commission	Development of a Data Base for Taxation	Demonstration

Aside from individual project responsibility, the Center has given three items highest priority for the next reporting period: working with State and Federal agencies to develop small application projects, documenting present methodologies, and reorganizing administrative structure for increased service function.

One weak point in the Center's operation has been the lack of formalization and documentation of computer support. Only by having easily accessable, user-oriented computer software/hardware will we be able to properly support the local user community. As such, the documentation of all existing software has been given highest priority in our computer support efforts. Also given very high priority is the integration of our image analysis and data base software into a single package.

In concert with our increased service role, we are administratively reorganizing to permit proper service for users, plan for the future work, and continue with our educational/research functions. Our future general structure will likely be as follows:

Univ. Admin.

MRSC
Director

Univ.
Admin.

Computer Support.

Project Leader

Project Leader

Project

________ : Control

___ _ ___ : Coordination

Establishment of the Center's program priorities and proposals for new work will be the responsibilities of the MRSC Director. One individual will be assigned responsibility for Computer Support and one each for the various Projects. Individuals may have dual project leadership roles for smaller efforts. These individuals will have primary control of their project, it's organization, progress, and day-to-day operation. Project leaders will be responsible to the Director. All requests for personnel, administrative support, and other resources will be submitted by the Project Leaders to the Director.

III. PROJECT PROGRESS REPORTS

A. Remote Sensing Applications in Land Use Planning-Lowndes County

Objective

To develop a Landsat-based data management system that will provide variables and data which can be used by the County Tax Assessor, the Civil Defense Director, and the Lowndes County Board of Supervisors, and for employment in the land use planning function by the Golden Triangle Planning and Development District and the Mississippi Research and Development Center.

Accomplishments

The primary use of the Lowndes County Information System has been in demonstrations to various individuals and groups (Section IV). One very productive demonstration was to a group of businessmen, bankers, and legislators from Lowndes County which contained individuals with opposing viewpoints on selection of sites in the County for hazardous waste disposal. Through negotiation, the group arrived at a suitability model which was run as they input the data. The model indicated the relative suitability not only of areas already under consideration, but also of areas previously unidentified. The exercise clearly pointed out the high value of an information system, not only in decision making, but also in the area of resolving public interest conflicts.

Current Status and Plans

The county is being considered for use in demonstrating the data base technologies applicability to store and manipulate tax records. Land ownership data, both map and tabular, will be obtained from sample areas within Lowndes County. The map data (ownership boundaries) will be digitized for select areas and integrated into the existing 48 variable data base. The tabular-type tax records will be loaded into the computer files and cross references (by coordinates) against the geo-referenced data base. This will be accomplished using the programs currently under development at the MRSC (see Sect. IIID). The development of this linkage between a geo-referenced data base and a tabular data base is being given high priority in anticipation of state-wide needs for tax-reassessment currently under discussion in the Mississippi Legislature.

B. Applications of Landsat Data to Strip Mine Inventory and Reclamation

Objective

The objective of this project is to provide the Alabama Surface Mining Reclamation Commission and the Geological Survey of Alabama with the software and interpretative techniques for monitoring strip mine occurrence and reclamation activities. The results will also be provided to the Mississippi Geological, Economic, and Topographic Survey, the State agency which is responsible for administering the surface mining law in Mississippi.

Accomplishments

Only recently has a frame of data become archived which may serve to test the validity of the Decision-Tree Classifier developed in 1979. Searches have been made periodically, but data of even reasonable quality have been unavailable. Band 5 and 7 prints have been ordered for a Feb. 23, 1981 date, but the cloud cover is 10% and Band 7 quality is 5.

Current Status and Future Plans

After preliminary evaluation of imagery on order, CCT's will be ordered if it is suitable and temporal extension of the Decision-Tree Classifier will be attempted. The project is inactive at this time.

C. White-Tailed Deer Habitat Evaluation Using Landsat Data

Introduction

In order to provide a basis for sound natural resource management in Mississippi, the Mississippi Department of Wildlife Conservation has initiated the development of a state-wide data base system which will be used to describe various components of Mississippi's ecosystems. The high priority of the white-tailed deer (*Odocoelius virginiana*) in the Department's management policies dictates that various types of deer "habitat" be mapped and evaluated on a state-wide basis. These "habitats" will be delineated on the basis of several biophysical variables.

Because of its synoptic and temporal characteristics, Landsat multispectral scanner (MSS) data will be used as the basis for vegetative classification. Both supervised and unsupervised classification of the data will be performed to determine the most accurate and the most cost-effective means of mapping vegetation. Other variables used to evaluate deer habitat will be compiled from existing sources. All data will be configured in a computer-assisted data base to facilitate rapid and accurate habitat evaluation.

Objectives

The project's objectives, in order of planned completion, are:

1. To determine those types of vegetative associations which are of significance in managing Mississippi's white-tailed deer.
2. To determine which of several analytical procedures are most effective in detecting these vegetation types using Landsat MSS data.

3. To configure this vegetation data, as well as other data pertinent for habitat evaluation, in a computer-assisted data base which will permit habitat description and evaluation.

Accomplishments

Computer-assisted digital image analysis for land cover at the Leaf River study area has been completed. The classified image is presently stored on magnetic tape awaiting injection into the area's digital habitat data base. Image analysis efforts for the Choctaw and Tallahala study areas are 30 percent and 60 percent completed, respectively. All image analysis are being conducted using the EOD-LARSYS program package. Specifically, an unsupervised approach is being employed using the ISOCLS and CLASSIFY processors.

Digital, geo-referenced, habitat data bases have been under construction for each study area. These three data bases have been given high priority during the past six months, and numerous terrain features have been digitized from map overlays. The data base for the Tallahala study area is complete except for Landsat-derived land cover and contains the following 15 variables:

Political Boundaries

Section-Township lines

Roads - Trails

Surface Water

Soils

Land Cover (from 1974 air photos)

Proximity to Roads

Proximity to Streams

Elevation

Ownership

Prescribed Burns - 1976

Prescribed Burns - 1977

Prescribed Burns - 1978

Prescribed Burns - 1979

Prescribed Burns - 1980

Data bases for the Choctaw and Leaf River study areas are 20 percent and 70 percent completed, respectively. Each contains terrain information similar to that for the Tallahala data base.

Computer programs were acquired from the Defense Mapping Agency's St. Louis, MO office to analyze digital topographic data. These programs were reviewed, and it was decided that extensive modifications are necessary to use them for generating slope and aspect from NCIC data tapes. Thus, it was decided to write a more simple set of programs in-house. Approximately two weeks of programming effort has been expended in this regard.

Working with Mr. Bob Griffin of the Mississippi Department Wildlife Conservation (DWC) a model was designed for Wild Turkey habitat evaluation at the Tallahala Wildlife Management Area. This model was used to interrogate the Tallahala habitat data base. The result of this procedure was a map of the study area depicting zones of different (relative) habitat quality. This map, along with tabulations of acreages in each habitat quality group have been forwarded to the DWC for review.

Current Status

Computer-assisted land cover mapping for the Tallahala and Choctaw study areas is being conducted as the highest priority item. Additional field checks for the Choctaw study area are being considered to maintain high accuracy in the complex land cover pattern in that area.

Digitizing of terrain features is underway for data from the Choctaw and Leaf River areas. This is a high priority item but is not a high expense task as all work is being conducted by technicians.

User-oriented documentation on digital image analysis and data base construction and use has been initiated. This effort is parallel to that for documenting the program-level functions but is intended for use by resource planners using the remote sensing/data base technologies.

Future Plans

The following tasks have been established as goals for the next reporting period:

1. Habitat data bases will be completed for the Choctaw and the Leaf River study areas.
2. Computer-assisted land cover mapping for all three study areas will be completed and loaded in the habitat data bases.
3. The Wild Turkey habitat model which has been forwarded to the Mississippi DWC will, upon receipt, be reviewed and revised if necessary. The DWC will also be solicited for specifications of models for other species across all study areas.
4. A first draft of user-oriented documentation for digital image analysis and digital data bases will be completed.
5. A technology transfer effort will be initiated to involve the Mississippi DWC, the Mississippi Forestry Commission, the U. S. Fish and Wildlife Service, and the U. S. Forest Service in the digital data base technology. We feel we are at the point where we can demonstrate the methods, show examples of user pro-

ducts, and justify the costs of specific projects. A seminar will be planned for summer 1981 to present the techniques and to solicit initial contracted projects with these agencies.

D. Remote Sensing Data Analysis Support Systems

Objectives

Concurrently with development of new software capabilities at the Center, much existing software for the creation and management of geographic information systems has been enhanced, documented, and prepared for "downloading" to a micro-computer system. This means that algorithms to perform specific tasks have been modified to make minimal demands on core storage and I/O sophistication, in order to keep the cost of a fully implemented micro-system within a range accessible to county level governments. User fault tolerance has also been improved in much of this software.

New software developed, at the Center, is best described in terms of individual projects. Development of new software has generally been in response to a known need for a given capability.

Accomplishments and Current Status

Much work done at the Center involves classification and utilization of Landsat imagery in forest and ecosystem-oriented research, generally (though not always) in conjunction with the Center's geographic information system (G.I.S.) Capability. New software developed addresses the problem of enhancing classifier performance and accuracy. To this end, programs now exist which will effectively destripe Landsat imagery (including georeferenced imagery), and generate pseudochannels of data which have been quantitatively demonstrated to enhance maximum likelihood classifier sensitivity as well as accuracy. The results of this study appeared as a master's thesis in computer science by a staff member (Appendix I).

Software developed at the Center now permits integration of classified Landsat imagery into a G.I.S. on a routine basis. The software also performs a "transect" resampling procedure, at the option of the user, for instances when a scale conversion of the data is also required.

Software has been developed in cooperation with the Department of Wildlife Management at the University, which permits the tracking by radio of animal locations within a G.I.S. This has opened up an entirely new dimension to wildlife research, since the capability now exists to monitor an animal's movements through time within a logically structured ecosystem rather than simply across the surface of a map.

A software system is now being designed and coded which will directly link a G.I.S. to a county's files on land ownership. In this manner, management information based on ecological monitoring utilizing the G.I.S. can be directed immediately to interactions involving the individual landowners affected. Such a system when completed, will be a very powerful and positive tool for agricultural extension services and the concentration of resources where they can be applied with the greatest impact. The formal linkage of the legal land ownership system recently proposed by the Mississippi legislature to an accurate ecological information base is an exciting prospect.

Presently, three programs exist on the Data General S130 computer for the purpose of manipulating images on the Lexidata 6400 image display system. The three programs are DEMON, CALUP, and IDBM. A brief discription of each program follows.

DEMON works from a data base file to display various images on the image display system. A data base consists of up to 40 scenes (called variables) each describing a different attribute of a common geographical area. The primary utility of DEMON has been to display the various data base variables and to overlay linear feature variables onto terrain feature variables.

The CALUP (Computer Aided Land Use Planning) program works from the same data base fill as does DEMON. The user inputs a model consisting of the data base variables to be used and weights to be associated with each value of each variable. The different variables described by the model are then combined in a linear fashion, one sixth at a time, to form a new scene or image. The new scene is immediately loaded in the Lexidata 6400 refresh memory for viewing on the color monitor. Both the model and the resulting image can be saved for further viewing and manipulation.

The IDBM program has been used primarily for drawing and manipulating images from MSS tapes and images produced by the LARSYS classifier routine. The program allows the user to input an expression containing both arithmetic and logical operations. Operanda in the expressions use previously saved images. The program then combines said images one pixel at a time in the manner described by the input expressions to produce a new image. The new image is then displayed on the color monitor.

Future Plans

Outside of the fact that DEMON and CALUP use a common data base, the only connection among the three programs lies in both program ability to save and restore images in a common format. This makes it awkward (and sometimes frustrating) for the user who needs to use more than one of the programs to obtain a desired image. In order to remedy this situation an effort is currently underway to combine the features of CALUP, DEMON, and IDBM into one program.

The new Image Operating System program will be modeled somewhat after the NASA/ELAS program. That is, the program will be very modular in structure. It will consist of a root segment and a single overlay area. The function of the root segment will be to input commands, load the necessary module into the overlay area, and pass control to the module to carry out the command. Utility routines will be provided to add modular and (commands) to delete modules from the program. It is expected that the effort will require approximately three months to complete.

A situation which has existed for some time at the Center is the "Division of Labor" between two different computer systems; a UNIVAC 1100/80 Mainframe, and a Data-General S-130 Minicomputer (DG). The recent acquisition of the DG has made this redundancy no longer a viable approach since system support costs to this system have increased. It is therefore considered absolutely necessary that all existing capabilities and services be transferred as quickly as possible to the minicomputer system. The existing mainframe capability will be archived on tape as a backup for times when the DG is being serviced.

It is the goal of the Center to tailor the DG into a sophisticated tool for research as well as service in the areas of forestry, land management, ecological science, and image processing. The system design strategy will be to make interactive use fault tolerant and self-explanatory in order to encourage use of the system by professionals in the field of bio-science who have limited computer background. In order to make the system more accessible to such users, plans are underway to relocate the system in the Department of Forestry on the campus of Mississippi State University. It is felt that widening the professional user base of the system will insure a progressive evolution of system capability as well as future support participation by user departments.

E. Landsat Change Discrimination in Gravel Operations

Objectives

The major objectives of this project are: (1) the development of computerized concepts and methodologies that will allow the user to effect temporal change detection in the extent of gravel mining operations using Landsat MSS data; (2) to apply these concepts in a change detection analysis on a portion of the Loessial Bluffs in north central Mississippi.

Accomplishments

The development of conceptual software necessary for the project has been achieved. A complete digital image classification analysis has been performed for the study area from satellite data exposed on March 22, 1978, a date just prior to the enactment of the Mississippi Reclamation Act. The project is being funded by the Mississippi Mineral Resources Institute (MMRI).

Current Status

Due to line start anomalies and hardware malfunctions of Landsat 2 and 3, it was not possible to obtain satellite data of a sufficient quality for the spring season of 1980 as originally planned. Currently, data are available and on order that meet the following specifications:

Exposure Date:	02JAN81
Image Quality (Bands 4, 5, 6 and 7):	8888
Cloud Cover:	00%

After receiving a 1:1,000,000, Band 5 positive print of the area at this exposure date and a preliminary evaluation of it to insure the quality and visibility of the training fields to be utilized in the study, band sequential CCTs will be obtained for this exposure and a complete digital analysis will be performed that will serve as the actual change detection analysis. Methods and procedures will be accomplished as follows:

(a) Existing 1:24,000 9" X 9" CIR positive prints and field checks to all training fields being analyzed will be utilized as reference data and comparison guides for the development of spectral signatures from the Landsat data.

(b) A complete digital analysis (using all four spectral bands) will be performed by unsupervised classification methodologies for a portion of the Loessial Bluffs from just north of Greenwood, MS, to just south of the Mississippi-Tennessee State Line. Change detection in the last 3 years for 12 gravel operations falling along this area will also be performed. Spectral signatures for the following land use/cover types will be differentiated:

(1) Water

(2) Pine

(3) Hardwoods

(4) Pastures/Vegetated fields

(5) Gravel

(6) Reclamation areas

(7) Other inert (i.e. bare solid, concrete etc.)

(8) Edge effects

(c) Upon completion of the digital analysis and acreage estimates a supplemental report will be submitted to MMRI.

F. Discrimination of Unique Forest Habitats in Potential Lignite Areas In Mississippi

Objectives

The principal objectives of this project are: (1) to develop practical and cost-effective methodologies using Landsat and aircraft data to discriminate areas of relatively undisturbed old growth hardwoods within Mississippi's lignite belt; (2) to identify and map such areas and provide this information to botanists employed by the Mississippi Natural Heritage Program (MNHP); and (3) to incorporate various attributes, i.e., hydrology, soils, and land cover of the lignite belt into a digital geographic information system (GIS) that will facilitate the management of the accumulated data of the study area.

Accomplishments

A preliminary report was written and is now in the process of being reviewed by personnel of the MNHP for final publication. All computer software needed for maintenance and updating of the GIS has been written and is in full operation. Problems encountered in the integration of the Landsat land cover classification for the area have been resolved and the maximum root mean square error (RMS) for the entire data base now averages well below 2 for the fifty acre cells. The list of current variables for the data base is as follows:

```
CARTER•CALUP(1).NATHERIT-DB
11   NATHERIT PIXEL AREA = 50 ACRE
 1   ROUGH TOPOGRAPHY
 0   NO ROUGH TOPOGRAPHY
 1   ROUGH TOPOGRAPHY
◆
 2   SOIL ORDERS
 0   VOID
 1   ALFISOLS(WET)
 2   ALFISOLS(MOIST)
 3   ENTISOLS(WET)
 4   ENTISOLS(MOIST)
 5   INCEPTISOLS(WET)
 6   ULTISOLS(MOIST)
◆
 3   SOIL ASSOCIATIONS
 0   WATER OR VOID
 1   NEARLY LEVEL,MWD AND SPD,SILTY:ALLUVIUM,DELTA AND LOESS
 2   ALLUVIAL,FLOOD PLAINS,SPD AND PD,SILTY AND CLAYEY
 3   DEEP LOESS
 4   SANDY CLAY HILLS
 5   THIN LOESS AND COASTAL PLAINS
 6   ALLUVIAL TERRACE WITH PAN
 7   THIN LOESS
 8   SLACKWATER FLATS
 9   SILTY TERRACE
10   DELTA-MEDIUM TEXTURE WET
◆
 4   HARDWOOD FORESTS- LANDSAT IMAGERY
 0   VOID
 1   FORESTS
◆
 5   CULTURAL
 0   VOID
 1   ROADS
 2   CITIES
 3   RAILROADS
 4   AIRPORTS
◆
 6   HARDWOOD FORESTS - COUNTY FORESTERS
 0   VOID
 1   HARDWOODS
 2   PINE
 3   MIXED
 4   OTHER - NOT IDENTIFIED
 5   CONTRADICTORY
 6   UNKNOWN
◆
 7   WATER
 0   VOID
 1   LAKES
 2   RIVERS
 3   MAJOR STREAMS
◆
 8   HARDWOOD SITES (LOW ALTITUDE COLOR I.R.)
 0   VOID
```

```
  1  HARDWOOD SITE
 ◆
  9  COUNTIES(NORTH PART OF DB)
  0  VOID
  1  TUNICA
  2  DESOTO
  3  TATE
  4  QUITMAN
  5  PANOLA
  6  LAFAYETTE
  7  TALLAHATCHIE
  8  YALOBUSHA
  9  CALHOUN
 10  GRENADA
 11  CHICKASAW
 ◆
 10  COUNTIES(SOUTH PART OF DB)
  0  VOID
  1  MONTOGOMERY
  2  WEBSTER
  3  CLAY
  4  CHOCTAW
  5  OKTIBBEHA
  6  WINSTON
  7  NOXUBEE
  8  KEMPER
  9  LAUDERDALE
 ◆
 11 LANDCOVER FROM LANDSAT IMAGERY
  0 VOID
  1 UNDEFINED
  2 HARDWOOD PREDOMINATES
  3 INERT, HIGHLY REFLECTIVE
  4 WATER
```

Current Status and Future Plans

Of the 54 stands originally designated as having at least moderate potential as unique biological communities in Mississippi's lignite belt, 34 stands were selected to be reflown to produce CIR imagery at an approximate scale of 1:8,000 for a more intense analysis, i.e., species identification, size, and refined site characteristics; the flight was made on April 27, 1981. The Center will await development of the film for further analysis.

Additive function modeling procedures using data gathered from various sources as to location of hardwood sites were performed. The variable utilized are as follows:

(1) Hardwood forests identified by County Foresters;

(2) Hardwood forests identified from low altitude CIR imagery:

(3) Hardwood forests identified from Landsat imagery;

(4) Hardwood predominates identified by digital analysis of Landsat imagery;

This modeling resulted in 10 areas selected that were reported in all four categories and 935 areas which were selected in three of four sources. Further maintenance and variable input for this data base is not anticipated although the system has capabilities for expansion.

G. Discrimination of Freshwater Wetlands for Inventory and Monitoring

Introduction

Concern over the destruction of the nation's wetland resources is currently in the forefront of environmental issues. Their value as a protection to other ecosystems has only recently been recognized, as well as their concomitant value in supporting unique plant and animal species themselves. Although many studies have been conducted, laws passed, and classifications systems established, there are still many large gaps in knowledge which need to be filled.

The official wetlands classification system currently being used by U. S. Government agencies is the U. S. Fish and Wildlife Service's Classification of Wetlands and Deepwater Habitats of the United States. This system was designed for use on a nationwide basis, and is therefore necessarily very general in character; hence it is of limited value on an area or local scale. The Mobile Corps of Engineer District has expressed a need for, and interest in developing a system that would be more closely correlated to unique local and area conditions.

Objective

The objective of this proposed study is to develop remote sensing techniques, utilizing aerial imagery and satellite data, for delineating freshwater wetland types with increased accuracy and decreased intensity of on-site inspection.

Methods and Procedure

Okatibbee Reservoir, located in Jasper County, Mississippi, will be used as the model or primary study area, since it is a hardwood bottomland wetland ecosystem typical of the southeastern United States. Due to the presence of the reservoir itself, a number of other wetland ecosystems and subsystems have developed adjacent to the area. In addition, aerial imagery and Landsat data are already on hand.

Training sites will be selected on Okatibbee Reservoir that are representative of classifications under the Fish and Wildlife system. These training sites will then be ISOCLAS'ed and the study area classified utilizing multi-temporal Landsat data and the EOD/LARSYS software package. Aerial imagery and field data will provide ground truth for refining the computer classification; i.e., assigning signatures to specific wetland types.

The refined system will then be tested for accuracy and functionality in several different locations using an unsupervised classification method, and the ground truthing by on-site inspection.

Accomplishments

The Okatibbee Reservoir area within the Corps of Engineer boundary (10,950 acres) has been ground truthed, manually analyzed and mapped on 1/24,000 color infrared imagery, and transferred by sketchmaster for geo-correction to U.S.G.S. 7.5 minute quad sheets. The mapping has been ground checked, and permanent mylar master copies have been made. Thirty-seven (37) land cover categories were delineated, and acreages for each category determined by polar planimeter.

The same general process has been completed for the supplementary study area on Tallahala Creek; however, little ground truth data were collected due to adverse land-owner reaction. This problem may result in the deletion of the Tallahala area from the study, even though several training sites have already been selected from a Landsat gray-scale computer map. There is, however, one large beaver impoundment which remains a possibility as it is located on property owned by an oil company from whom access permission has been obtained.

An ISOCLS and unsupervised classification of the entire Okatibbee Lake area utilizing May 1978 Landsat CCT digital data and the EOD/LARSYS software package was run prior to completion of imagery analysis to develop general land cover signatures.

Seven (7) major delineations were made:

1) Inert
2) Edge
3) Hardwood
4) Field
5) Pine
6) Open Water
7) Marsh

In addition, four subclasses were delineated for marsh areas.

Future Plans

Within the next three months, specific wetlands training sites will be selected at Okatibbee Lake, ISOCLSed, and intensively field checked in order to identify the four marsh subclasses and determine whether they are truly unique. Also, a decision will be made whether to utilize the oil company site on Tallahala Creek as a training site; additional test sites will be selected to test the classification system.

IV. LIST OF SPECIAL ASSISTANCE OFFERED

Information Supplied or Publications Supplied

Mr. John Soileau, Tennessee Valley Authority, Norris, TN

Mr. Paul Tessar, Nat'l Cont. of State Legislatures, Denver, CO

Ms. Lucy Wayne, Dept. of Archeology, Univ. of Florida, Gainesville,FL

Dr. C. Arnaiz, Dept. of Botany, University of Madrid, Madrid, Spain

Associated Planning Group, Inc., Canton, MS

Mr. Richard Birdsay, U. S. Forest Service, So. Forest. Exp. Sta. New Orleans, LA

Mr. Loyd Huckaby, Equalization Div. Chief, State Tax Comm., Jackson, MS

Dvorkovitz and Associates, Ormand Beach, FL

Demonstration and Educational Activities

Dr. Guido Kasten, Vice President, md. Dr. Flavis Rocha, Professor, Federal University of Pelotas, Brazil.

Miss. Senator Perrin Purvis; Representative Tommy Brooke, James Loyd, Bruce Hanson, and Tim Ford.

Mr. Clyde Whitaker, Mayer of Tupelo, MS.

Mr. J. W. Collins, President, Bank of Mississippi, Tupelo, MS.

Mr. Charlie Greer, Vice Pres., Bank of Mississippi, Tupelo, MS.

Mr. Bill Dunlap, Chairman of the Board, Peoples Bank and Trust Co., Tupelo, MS.

Mr. Jerry Clayton, Chancery Clerk, Lee County, MS.

Lee County Supervisors, Jack McKinney, Thomas Kennedy, Al Rice, James Mancey, and Billy Davis.

Gene Peterson, Director, Chamber of Commerce, Columbus, MS.

Kathy Webb, Office Manager, Chamber of Commerce, Columbus, MS.

Bill McClure, Owner of McClure Furniture, Columbus, MS.

Larry Tyner, Public Relations Director, Weyerhaeuser Corp.

Frances Jutman, National Bank of Commerce, Columbus, MS.

Bill Walker, Mississippi Employment Service, Columbus, MS.

Jack Colbert, Colbert Real Estate, Columbus, MS.

Stan Murry, National Bank of Commerce, Columbus, MS.

Hyde Powers, Swoope Insurance and Real Estate Agency, Columbus, MS.

Jim High, President Seminole Manufacturing Company, Columbus, MS.

Tony Pierson, Commander of 14th Flying Training Wing, Columbus, MS.

Stanley Hayes, Commercial Dispatch, Columbus, MS.

Eddie Cox, First Columbus National Bank, Columbus, MS.

Johnnie Murphree, Boad of Co-Operatives N.O., Treblock, MS. 38875

Darrell Smith, Manager, Sunshine Grain, Tupelo, MS. 38801

Robert E. Herndon, Owner-Operator, Herndon Well & Supply, Shannon, MS.

Frank Riley, Attorney, Tupelo, MS 38801.

Harry Martin, Executive Vice President & Secretary, Community Development Foundation, Tupelo, MS 38801.

Russell Bryson, Vice President, First Citizens National Bank, Tupelo, MS 38801.

Dan Wilford, Adminstrator, North Mississippi Medical Center, Tupelod, MS.

Loren Zimmerman, President, Federal Land Bank, Tupelo, MS.

Marion Eifling, Vice President, N. E. Mississippi Production Credit Association, Tupelo, MS 38801.

Billy Haygood, Accountant for L. D. Hancock & Co., Tupelo, MS 38801.

Bill Rieves, Assistant Superintendent, Tupelo Separate School District Tupelo, MS 38801.

Pete Albritton, Tour Guide

Invited Papers

Clark, Jon R. and W. F. Miller. 1980. Wildlife Habitat Eval. Using a Computerized Digital Terrain Data Base. Annual Meeting Mid-South Region, Amer. Soc. Photo. Member TN October

Turnipseed, D. P. and W. F. Miller. 1980. Discrimination of Active Gravel Operations from Landsat MSS Data. Annual meeting Mid-South Region, Amer. Soc. Photo. Memphis, TN. October.

Powers, John S. 1980. A Global Image Statistical Transform - Preliminary Investigation. Ann. Meeting, Mid-South Region, Amer. Soc. Photo. Memphis, TN. October.

Miller, W. Frank. 1981. The use of remote sensing techniques for archeological investigations in humid regions. Fourteenth Annual Meeting, Soc. for Historical Archeology, New Orleans, LA. January 5-7.

Miller, W. Frank. 1981. Imagery Analysis of the Barton Ferry Recreation Area prior to archeological, documentary, and oral historical investigations. Same reference as previous.

APPENDIX I

AN INFORMATION THEORIC SPATIAL TRANSFORM FOR DIGITAL, MULTISPECTRAL SATELLITE IMAGERY

BY

John Stephenson Powers

A Thesis Submitted to the Faculty
of Mississippi State University in
Partial Fullment of the Requirements
for the Degree of Master of
Science in the Department
of Computer Science

Mississippi State, Mississippi

December 1980

AN INFORMATION THEORIC SPATIAL TRANSFORM

FOR DIGITAL, MULTISPECTRAL

SATELLITE IMAGERY

by

John Stephenson Powers

APPROVED:

Bradley D. Carter
Professor of Computer Science
(Director of Thesis and
Chairman of Supervisory
Committee)

Bradley D. Carter
Head of the Department of
Computer Science

Sidney M. Williams
Associate Professor of
Computer Science
(Member of Committee)

Fred Davis
Director of the Division of
Computing Affairs

Associate Professor of
Forestry
(Member of Committee)

Vice President for Graduate
Studies and Research

Jimmy R. Solomon
Associate Professor of
Mathematics
(Member of Committee)

ACKNOWLEDGEMENTS

This research could not have been conducted without the support of certain individuals and the work already completed by others. A special debt of gratitude is felt to W. Frank Miller, Director of the M.S.U. Remote Sensing Center and Professor of Forestry, for his unceasing support and encouragement. This study could not have been undertaken without his assistance.

While this research resulted in the development of a sizable body of software, it made extensive use of Landsat image display software and geobase creation and display software written by others. Dr. Bradley Carter, Chairman of the M.S.U. Department of Computer Science, is largely responsible for the CALUP (Computer Assisted Land Use Planning) geobased software system implemented at the M.S.U. Remote Sensing Center and used in this study.

Special thanks go to Mr. Jonathan Clark, Wildlife Specialist at the Center, for his skilled management of data resources, and for undertaking the task of mapping ground truth in the study area from low altitude photography and field checks. His contribution to this study was significant and of uncompromising accuracy.

This thesis represents the culmination of an idea which first entered my mind five years ago, in 1975. To my parents, who have supported and encouraged me unceasingly through the years, I express my deepest gratitude.

ABSTRACT

John Stephenson Powers, Master of Science, December 1980

Major: Computer Science, Department of Computer Science

Title of Thesis: An Information Theoric Spatial Transform for Digital, Multispectral Satellite Imagery

Directed by: Dr. Bradley D. Carter

Pages in Thesis: 73 Words in Abstract: 237

ABSTRACT

This research addresses the problem of computer classification of regional landscape characteristics by utilizing multispectral image data acquired by satellite. Computer classified satellite imagery is gaining increasing acceptance as an integral component of landuse management information systems. Such geobased information systems require accurate and recent landcover information in order to facilitate the planning and decision making process regarding the environmental impact of various classes of economic activity.

Economic factors are of extreme importance in computing. This study is based on the premise that significant and cost-effective gains in image classification performance by existing classification algorithms may be effected by a fast, and relatively simple spatial transform applied to the image data set. The transform utilized is based on the findings of information theory, and is analogous to the spatial transform applied at the retinal level in human visual perception.

This study synthesizes channels of image information by use of an information-theoric spatial transform on the raw data set, and compares image classification results utilizing the transform channels, to classification results utilizing the unaltered multi-spectral data set. Visual inspection of the results reveals that the utilization of transform channels results in a striking enhancement of the sensitivity of the classification algorithm utilized. The enhanced sensitivity as well as structural coherence of landscape features discerned, indicates the economic viability of the new approach in conjunction with existing classification systems.

KEY WORDS

Pattern Recognition, Image Classification,
Channel Synthesis, Information Theory

TABLE OF CONTENTS

PAGE

APPENDIX

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

Chapter I

INTRODUCTION

An important problem in computing and pattern recognition is the discrimination and reliable classification of patterns on the earth's surface. This problem addresses unresolved theoretical issues in perception and pattern recognition, and has consequence as an important tool for the monitoring and management of earth resources. As world population trends continue, increasing demands will continue to be placed on all earth resources as developing nations compete to industrialize and the economies of industrialized nations adjust and continue to grow. Regional ecological and economic monitoring and the planned design of resource utilization (or non-utilization) will become increasingly necessary if costly ecological and economic blunders are to be averted at the local as well as the national level. For this reason, research into areas concerning the acquisition and utilization of information pertinent to the goals of accurate resource inventory, environmental monitoring, and economic management is justified.

Information systems exist which accomplish these tasks with an acceptable degree of accuracy. The major problem issue is the complexity of such systems and the large capital investment required to acquire computer equipment and personnel to accomodate the large volume of necessary software. This situation has been changing, however, since a major trend in computing is the movement towards smaller and more specialized systems which address specific levels of problem complexity and data volume. While the cost of computing equipment has been steadily decreasing over the last decade, recent trends indicate that a threshold has been reached based on rising costs of software and raw materials for electronic components, and it is unlikely that this continuing decrease in system cost will continue as it has over the previous decade. Medium to high powered data processing systems beyond the capacity of existing microprocessors will continue to require large investments of capital. Since the cost of a machine is directly related to its speed, utility software and data capacity, geobased information systems which require large and highly sophisticated computer systems to operate will be at a marked economic disadvantage. There is also reason to believe that widespread use of computerized geobased data systems in the private and public sectors will not become a reality until relatively low cost systems with a high degree of accuracy are developed for use on smaller, less expensive systems. In order for smaller equipment to accomplish the performance goals necessary, the recognition and classification process must be reduced to a number of discrete and computationally manageable steps.

Exploiting economies which may be achieved by relatively simple pre-processing of data to "filter" useful catagories of image information is directly analogous to the way the human brain receives and processes visual information. The recognition of an object, or class of objects, is not an immediate or "single step" process involving vast computational resources, but involves sequential coding, flow and transformation of information in discrete processing stages among highly specialized brain centers.[1] The final integration of visual information in the visual cortex is possible only after several stages of spatial and edge-detection transformations have been accomplished in hierarchically subordinate neural structures. The findings of recent research into these areas[2] indicate that human visual perception is a directional flow of information through interconnected processing stages which begins at the eye's retina. At the retinal level, visual information is not simply coded into an intensity matrix, but undergoes complex spatial preprocessing as well, which is based on the interactions of mutually inhibitory and excitatory receptive fields within the retinal structure.[3] Without this initial "primitive" level of information preprocessing, there is reason to believe that subsequent, "higher" processing levels would cease to function or would be greatly impaired, and coherent, reliable perception and classification of the visual world would become impossible. The point of this

---

[1] Bela Julez, Foundations of Cyclopean Perception, (The University of Chicago Press: Chicago, 1971), pp 54-94.

[2] David H. Hubel and Torsen N. Wiesel, "Brain Mechanisms of Vision", Scientific American, (August 1979), pp 150-163.

[3] Tom N. Cornsweet, Visual Perception, (Academic Press: New York, 1970), pp 268-310.

discussion is that each relatively simple level of processing is a necessary step in the final recognition and classification solution, and there is good reason to suspect that this may also prove to be the case in machine classification of features on the earth's surface. Existing classifiction algorithms which classify raw multispectral data by utilizing the maximum likelihood statistical clustering approach possibly may be greatly enhanced in accuracy by spatial preprocessing of such data. The synthesis of transform channels of data which are to be utilized in place of, or in addition to, raw multispectral data in the existing classification process is the purpose of this thesis. The procedure is not to mimic the processes of the human retina, but to approach the problem of spatial pattern classification at an analogous level within the total classification process.

Thesis Statement and Research Objectives

The features distinguishable on the earth's surface are amazingly diverse and it is therefore necessary that the study be concentrated on a specific problem area which will yield results of immediate practical benefit as well as illuminating a portion of the problem as a whole. For this reason, this research is centered on the spatial preprocessing of Landsat multispectral scanner (MSS) data for the purpose of enhancing the accuracy of existing forest classification procedures from hyper-altitude, small scale satellite imagery.

The scale of such imagery is an important and hotly debated factor in regional forest resource inventory. The scale of the image data represents a tradeoff between data volume, which can quickly become enormous and intractible over regional areas, even to large, very fast machines; and classification reliability, which is a function of the density and therefore the scale, of digital spatial data. Physical hardware constraints, even on "fast" machines necessitate the use of small scale digital image data with a relatively low density or resolution in the classification of large areas. This is because that in MSS data, volume increases as the square of the area times the dimensionality. Attempting regional classification over large areas utilizing large scale, high resolution, digital multispectral data would surpass the ability of the largest existing computers, and even the capabilities of theoretical machines to accomplish within a reasonable period of time.[4]

This research will investigate possible gains in large area classification accuracy utilizing low-density, hyper-altitude satellite MSS imagery by means of spatial preprocessing and channel synthesis utilizing such data. Spatial preprocessing will involve a nearest neighbor transform over the data set, which will be a measure of the information content or <u>entropy</u> of the multispectral image on a cell by cell

---

[4]Larry J. Stockmeyer and A. K. Chandra, "Intrinsically Difficult Problems," <u>Scientific American</u>, (May 1979), pp. 140-159.

basis. The transform is spatial due to the fact that every point on the image is transformed based on itself and all contiguous neighboring points. In such an information theoric transform, every point becomes a function not only of its neighboring points, but of the total occurrence over the complete image of each point class. Each point is thus defined according to local as well as global factors; an analogous, if not exact, approximation of the human retinal transform. This transform represents an additional channel of information which may be utilized in the classification process. It seems reasonable that the addition of such a channel to the image classification process would result in gains in accuracy, and it is the thesis of this research that the gains in accuracy thus obtained will offset the economic factor of added computational overhead. Such gains in accuracy, should they be significant, would be of immediate benefit in regional forest inventory. Since an accurate and updated inventory of forest resources in the hands of professional foresters is an essential element in a successful regional resource management and ecological monitoring system, this research is addressed to the problem of facilitating the accuracy and timeliness of such information with minimum computational overhead.

Research Methodology

The goal of software development for this study is creation of a classified image representing ground cover cluster categories over a study area which utilizes tranform channels in the image classification process. An identical area will be classified utilizing only the untransformed multispectral data. The methodology of the thesis will include the final integration of both images within an existing geobase information system for the Tallahala Wildlife Refuge in Jasper County, Mississippi. The classified imagery will be coordinate registered to the database, which contains ground truth information derived from medium altitude color infra-red imagery, and extensive field investigation, conducted by Mr. Jonathan Clark, a wildlife specialist at the M.S.U. Remote Sensing Center. Based upon ground truth information within the registered database, classification accuracies utilizing the transformed classification, as well as the untransformed classification, will be tabulated on a pixel by pixel basis over the entire study area. In this manner, gains in accuracy can be quantified according to an exact and unbiased procedure, in addition to visual inspection.

Chapter II

MATHEMATICAL FOUNDATIONS OF AN ENTROPY-BASED SPATIAL CLASSIFIER

Digital multiband remote sensor imagery is generally approached as a statistical assemblage in decision theoric clustering algorithms (in contrast to recent syntactic approaches).[5] In utilizing such a clustering approach, it is necessary that certain simplifying assumptions be made concerning the global characteristics of the data. This is due to the fact that attempts to deal with the generally large volume of such data in its full multi-dimensional, spatially intercorrelated complexity rapidly make numerical approaches to classification computationally intractible even on very fast hardware.

Important assumptions in existing classification software concern the independence of pixel probabilities across space, implying the assumption that the spatial distribution of pixels within a given spectral band describe a Gaussian or random distribution (i.e., are not spatially correlated). This assumption violates our intuitive concepts as well as our observational knowledge concerning the articulation of the earth's surface, yet they successfully describe the gross characteristics of the data set if approached as an "abstract" assemblage, free from the preconceptions associated with geographic phenomena.

---

[5] R. C. Gonzalez and M. G. Thomason, "Tree Grammars and Their Application to Syntactic Pattern Recognition," (Office of Naval Research: Arlington, Virginia, May 1974).

Information theory, as outlined by Shannon (1949), requires the assumption of discrete ergodicity of the data set. A gross definition of ergodicity is "statistical homogeneity." It refers to the characteristic of a data set that any assemblage or sequence from the data set is characteristic of the set as a whole, and that as the length of a sequence increases, our knowledge of the total set increases monotonically if we disregard pathological samples which may occur with low probability. Ergodicity of a data set implies that the set is homogeneous in the sense that subsets are not significantly clustered; not simply that the data distribution is perfectly uniform. It is necessary to assume for the purposes of this research, that in dealing with multiband, digital remote sensor imagery, points within the data set are not only spatially independent from a probabilistic point of view, but are in fact elements of an assemblage which meets the specifications of a discrete, ergodic source.[6] This is based on the observations that:

1) Each spectral channel is distinct and may be computationally treated as a separate source and thus statistically independent from other spectral channels - correlations are ignored. This permits the summation of the logarithms of independent probabilities across spectral channels, i.e., multi-

---

[6]Claude E. Shannon, The Mathematical Theory of Communication, (Chicago: University of Illinois Press, 1949), p. 45.

plication of probabilities across spectral channels is computationally permissable.

2) The possible state of any given pixel is discrete and discontinuous within a given channel; i.e., each channel may be treated as a discrete sequence of symbols from a finite set of symbols; the multi-spectral image being the superposition of such sets of sequences.[7]

3) Each spectral channel may be regarded as an independent ergodic source; as lengths of sequences of pixels from any given channel increase, they will monotonically tend to approach limits unique to that specific channel as a whole. Each channel may thus be regarded as statistically homogeneous and independent.

Given the above necessary assumptions which may be defended on the characteristics of the "abstract" data set and which fit the description of a discrete, ergodic source in information theory, we may feel justified in applying the tested and proven principles of information theory regarding discrete ergodic sources on digital, multiband remote sensor imagery.

---

[7] Abraham Moles, Information Theory and Esthetic Perception, (Urbana: University of Illinois Press, 1966), pp. 161-169.

Selected Concepts from Information Theory

This research is directed specifically to the quantification of the organization of the earth's surface at a given scale relative to a statistical framework defined by the total image under analysis. Each spectral channel in digital, multispectral imagery consists of recognizable image sub-areas or features, composed of a discrete range of spectral radiance values which are interpreted as shades of gray. Specific gray scale values representing instantaneous resolution elements, or pixels, occur with a frequency within selected areas which is characteristic of the spatial organization of such areas relative to the image as a whole. Spatial organization is related to the concept of texture because areas dominated by homogenous elements are characteristic of fine or "smooth" textures, while areas with many diverse or "nonhomogenous" elements are characteristic of "coarse" textures. Texture may therefore be regarded as a spatial-frequency function. Frequency of occurrence of a gray scale value over an image may be expressed as a probability function and the relative organization of sub-areas within an image may be quantified by utilizing this function across the range of possible gray scale values. Characteristic "textures" or categories of image spatial organization may therefore be regarded as a probability function of constituent pixels relative to the global image statistics. It is important to remember, however, that the concepts of texture and organization, although related, are not identical. Texture generally refers to a physical attribute of an area (rough, smooth, banded, etc.) while organization, as used in this study, refers only to a quantifiable relationship between the probability function of a sub-area relative to the global statistical ensemble defined by the image as a whole.

This study approaches the quantification of the spatial organization or "texture" of sub-areas of an image as a spatial frequency function utilizing global spectral probabilities derived from the total scene. In this technique, the occurrence of every grayscale value or pixel within each channel over the entire image is counted. This completed, each grayscale category is assigned a probability of occurrence, which is simply the ratio of the number of occurences within the category to the total number of pixels in the image. This is done for each grayscale category within each channel. In this manner each grayscale category is transformed into a probability function reflecting the uncertainty associated with the occurrence of any given pixel from that category within the image, since spatial correlation is defined to be zero.

This approach does not automatically give specific ranges of values to specific physical textures, but classifies existing organization within a sub-area, or cell, based on the spectral statistics of the entire scene. The important point is not the value assigned to specific textures, but that existing textural, or feature categories within an image are differentiated consistently. Such a result would be of significant utility in algorithms which classify spatial data based on the statistical "clustering" of similar categories across multiple spectral channels of information.

In attempting to differentiate spectral features based on "texture," we must, in effect, measure the degree of spatial organization of spectral information within multi-pixel cells of an image relative to the organization of the image as a whole which is expressed by the global statistics of the image. This organization can best be quantified in terms of the amount of information required to describe it completely.

To measure the amount of information we must first generalize a given situation in terms of event probabilities. In this case, an event is the occurrence of a specific gray-scale value or "pixel" within a given spectral channel. A pixel or instantaneous resolution element, is the discrete, digital representation of a specific segment within a continuous range of spectral reflectance intensity as viewed by the sensor platform in space. Since pixels within a given channel are mutually exclusive (only one value may occur at a given point in space/time) and the set of pixels within a given channel is exhaustive (the possible values which any pixel may acquire completely define the event space), then the sum of the probabilities of occurrence of all possible pixel values within a given channel must equal one, such that

$$\sum_{i=1}^{N} P_i = 1$$

where N is the range of GSV categories. Any measure of the information of an event is a function of the probability of the event, the event being the occurrence of a member of a particular grayscale category at a given point.

In 1928 in a study of the transmission characteristics of telegraph lines, R. V. Hartley demonstrated that the most natural choice for a measure of information is the logarithmic function. This measure was subsequently adopted by Claude E. Shannon in his classic work The Mathematical Theory of Communication. Shannon writes, "The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used, the resulting units may be called binary digits or bits...If the base 10 is used the units may be called decimal digits. Since

$$\log_2 M = \log_{10} M / \log_{10} 2 \doteq 3.32 \log_{10} M$$

a decimal digit is about 3 1/3 bits."[8]

In a related passage dealing specifically with digital image processing, and paraphrasing Shannon, E. L. Hall succinctly categories the reasoning behind the choice of the logarithmic function:

"...an information measuring function h should satisfy the following conditions:

1) $h(p)$ is continuous for $0<p<1$

2) $h(p) = \infty$ if $p = 0$

3) $h(p) = 0$ if $p = 1$

4) $h(p_2) > h(p_1)$ if $p_1 > p_2$

---

[8]Shannon, The Mathematical Theory of Communication, p. 32.

If another picture independent of the first is jointly considered with pixels $(n_1, n_2 \ldots n_N)$ and probabilities $(q_1, q_2 \ldots q_N)$, then an additional property is additivity.

5) $h(p_i q_j) = h(p_i) + h(q_j)$

It may be shown that these postulates are satisfied if, and only if,

$$h(p) = -\log_b(p)$$"[9]

The fifth condition is very important since it allows us to easily combine conditional but independent probabilities by summing their logarithms or, computationally, by multiplying them. Thus the information of a single pixel across several channels (1,2,...N) is given as

$$-(\log_b p_1 + \log_b p_2 + \ldots + \log_b p_N) = -\log_b(p_1 \times p_2 \times \ldots \times p_N).$$

In this study, calculated logarithms will be taken to the base 2. This is in consideration of the fact that the "bit" is the standard delimiter of information, since 2 raised to the power of the number of bits gives the range of the data space required to span a given set of possibilities. Choosing 2 as the base of the logarithm thus scales measurements to the standard unit measure for information; we have

$$\log_2 X = \ln X/\ln 2 \quad \text{or} \quad \log_2 X = \log_{10} X/\log_{10} 2.$$

---

[9] Ernest L. Hall, Computer Image Processing and Recognition, (New York: Academic Press, 1979), p. 313.

We now define the information measurements to be utilized in this study. Given pixels $(x_1, x_2 \ldots x_N)$ with associated probabilities $(p_1, p_2 \ldots p_N)$ then the <u>self information</u> of any particular pixel within a given spectral band is given as:

$$I_i = -\log_2 P_i$$

and the average information or ENTROPY for any sequence of pixels is defined as:

$$H = -K \sum_{i=1}^{N} P_i \log_2 P_i$$

where N equals the size of the sequence, and K is simply a positive constant used for calibration purposes. The negative sign converts the measure to a positive number since the logarithm of a probability is negative.

Entropy of Multiple Spectral Bands

Multiple spectral bands of information can be described as conditional but independent sets of events. Although pixel values for a given coordinate may be highly correlated across several spectral bands, any given value from a specific band is not regarded as being <u>determined</u> by any other spectral value from a different band, even should they prove to be identical. The common determining factor for each band is the state of the earth's surface and intervening atmosphere which effects each band independently, and not the subsequent correlation between bands. Therefore, although all bands are conditional on the same set of factors, i.e. the state of the earth's surface, they are considered mutually independent.

Given this approach, each multispectral pixel coordinate may be treated as a vector $V_i\{p_1,p_2...p_N\}$ of conditional but independent probabilities associated with each band of spectral values, and the probability (B_i) associated with the occurrence of any pixel vector may be computed as the logarithmic sum (or simple product) of its component probabilities such that B_i=ALOG($\log P_1 + \log P_2 ... + \log P_N$) or equivalently, $B_i=(P_1*P_2...*P_N)$.

With this approach, and utilizing multiple spectral channels of information, then given a vector V_i with associated probability B_i, the <u>self information</u> of a pixel vector becomes

$$I_i=-\log_2 B_i=-\sum_{b=1}^{N} \log_2 P_b$$

or equivalently,

$$I_i=-\log_2 B_i=-\log_2\left(\prod_{b=1}^{N} P_b\right)$$

and the average information or Entropy of a set of pixel vectors may be determined according to the following theorem (from Shannon):

<u>Theorem</u>: Let B_i be the probability of a sequence V_i of symbols from the source. Let

$$G_N=-1/N \sum B_i \log_2 B_i$$

where the sum is over all sequences B_i containing N symbols. Then G_N is a monotonic decreasing function of N, and $\lim_{N\to\infty} G_N=H$ [10]

[10] Claude E. Shannon, <u>The Mathematical Theory of Communication</u>, p. 55.

This theorem states that when the source is considered in terms of sequences of symbols instead of individual symbols, then as the lengths of the sequences increase, the entropy over each sequence will approach the entropy of the source. The constant N is, of course, the number of spectral bands in each pixel vector and we must alter 1/N to K/N to calibrate the function to the actual sampling procedure utilized in the study. Shannon's Theorem 5 will yield the entropy of an N-dimensional array of pixel vectors in a nearest neighbor sampling frame, and is proposed as the core of an entropy classification algorithm for the synthesis of a band of coordinate registered entropy values from multispectral digital remote sensor imagery. Spatial correlation among contiguous pixels within the sampling frame is assumed to be zero, as explained previously, thus permitting the summation of vector probabilities over a nine pixel "window" which scans the image.

Calibrating the Entropy Function

The expression

$$H= -K/N \sum B_i \log_2 B_i$$

must be calibrated to actual conditions in the sampling environment. This is accomplished through the use of the constant K where, under the circumstances of this study;

$$K=\left(\sum_1^9 B_i\right)^{-1}$$

The resulting expression

$$H=-\left(N\sum_{1}^{9}B_i\right)^{-1}\sum_{1}^{9}B_i\log_2 B_i$$

is therefore submitted as the "core" of an algorithm to synthesize a channel of data representing the degree of spectral organization of the earth's surface within a global statistical frame defined by an image at a given scale.

In utilizing the entropy function, probabilities of occurrence of grayscale categories in an image utilizing global statistics must be inverted at (1 / Range of categories) since this represents the value at which uncertainty is at a maximum for a particular category relative to the image as a whole. On a scale of 256 categories, or the scale utilized in this study, the greatest uncertainty possible for the occurrence of a particular grayscale value or pixel at any point on the image is 1/256, and categories whose occurrence probability within the image is less than this value must be inverted, since it is uncertainty, not probability, which we are measuring. Individual pixel values are thus treated as members of grayscale categories or sets whose probability of occurrence at any point is determined relative to the fixed range of possible values over the image as a whole. Each grayscale value or pixel is thus assigned the probability of the category or set of which it is a member. For an 8-bit image with 256 categories (an image with pixels coded within a range of 0-255) the maximum possible information association with a category is $-\log_2 1/256$, or 8 bits, which is associated with a perfectly random probability function of the category over the image set. Pixels within this category are considered as conveying maximum information about the image since their possible distribution over the image is attributable to pure chance; therefore, the entropy assigned such pixels is 8 bits or the maximum possible over the specified range of categories. The entropy function appears counter-intuitive, in this sense, assigning maximum information to maximum uncertainty. The paradox is resolved, however, when it is realized that entropy is a measure of the information required to completely specify an event of given uncertainty; or as negative entropy the amount of information conveyed by the known occurrence of an event of given uncertainty. When the maximum possible uncertainty is 1/256, or a perfect random probability function, the information required to totally remove the uncertainty associated with the event is 8 bits, and the information or negative entropy conveyed by a known occurrence of a random event is 8 bits.

Given an observer and an event space, entropy measures the amount of information conveyed by an event within the space relative to the observers existing knowledge of the organization of the event space. It is in this manner that entropy serves as a measure of organization. The more highly organized an event space is, the more predictable it is and less information is required to specify any given state; i.e., its associated entropy is low. As an event space becomes more disorganized, it becomes less predictable and increased information is required to specify a given state; i.e., its associated entropy increases. In the context of this study, the event space is the image, and the totality of our knowledge about the organization of the image is embodied in a table of global grayscale probabilities. The uncertainty associated with the occurrence of each grayscale category increases as its probability approaches 1/256. Any departure of a category probability from 1/256 yields information (negative entropy) about the global image state, reducing its associated uncertainty, and the associated entropy decreases monotonically as the probability departs, either positively or negatively, relative to this maximum uncertainty state. The entropy function will therefore classify an image based on the extent to which the various pixels, or grayscale components, of the image provide us with information regarding the global image state. Probabilities less than $(1/\text{scale}^2)$ are omitted as zero since the function gives negative values beyond this range. In the studies conducted, a typical image segment of 123,000 pixels yielded 2 to 5 pixels falling within this category, or .004 percent of the segment.

In scaling the entropy function result for output as a synthetic channel, the actual output values inserted onto tape with existing spectral information will be simply integer $(2.0^H)-1$. The resulting channel will thus be scaled compatible with existing spectral information scales in the information format utilized.

Other Useful Statistics

With the creation of a channel of spectral entropy of the data area on a cell-by-cell basis, little computational overhead is incurred in computing the RELATIVE ENTROPY, which is the ratio of the actual entropy to the maximum entropy, and therefore, the REDUNDANCY, which is one minus the relative entropy. Redundancy represents the extent to which the existing information may be regarded as being determined by organizational statistical processes within spectral categories which are manifest within the image at a given scale. While Entropy increases with diversity, uncertainty, or disorder, the redundancy is, in a sense, the purest measure of organization of a set of data since it represents the degree to which values depart from the random and are determined by the inherent organization of an image category relative to the image as a whole. The redundancy may be considered as a rough measure of the "conspicuousness" of an image category, since it increases monotonically as the occurrence of a category becomes less and less attributable to random events.

Both of these valuable statistical channels are available with minimum computation, and involve only two passes through the data set. By utilizing the entropy measure described, the algorithm will classify spatial cells composed of pixels whose frequency over the image approaches the random probability of 1/S (where S is the range of categories) as having the maximum "textural" coefficient or entropy. Cells composed of pixels whose frequency within the image is either very high or very low should have low textural coefficients. Each cell is thus classified in terms of its organization relative to the organization of the image as a whole. As the organization of a sub-area departs from the perfectly random, so will its textural coefficient, or entropy, decrease. The redundancy does roughly the opposite; areas composed of "random" pixel categories should have the lowest textural coefficient and as the organization of a sub-area departs from the random, relative to the global image, the redundancy should increase. Both channels promise to be of significant utility in a statistical clustering classification process.

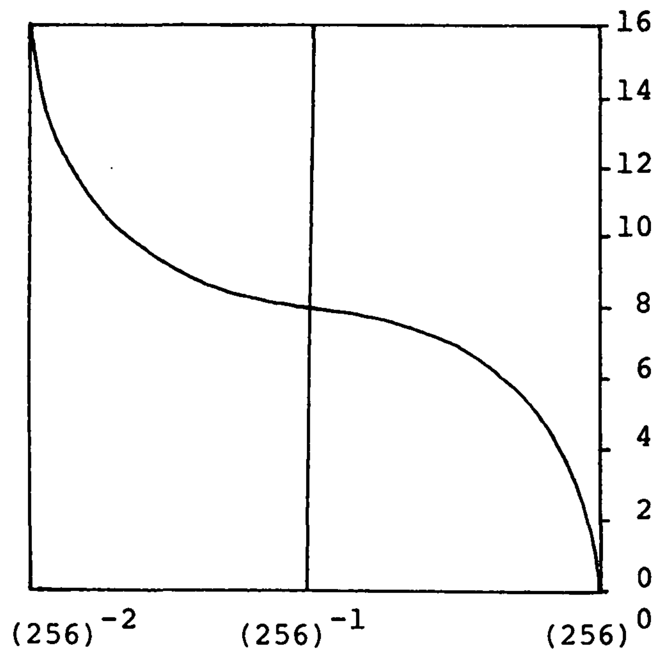
Figure 1 depicts actual algorithm behavior for multi-pixel cells containing homogeneous probabilities, as well as the probability inversion process. Figure 2 gives the program listing for implementing the algorithm on a Hewlett-Packard 34C hand held programmable calculator, if the algorithm is to be studied in greater detail.

Software written for this research (see Appendix) synthesizes channels of image entropy and redundancy from Landsat MSS data, using the algorithm defended in the previous section. In the study which follows, these channels are investigated for the purpose of proving or disproving their utility for increasing digital forest classification accuracy using the unsupervised LARSYS image classification system. The statistical and mathematical issues involved in defining specific image categories in terms of global image statistics are subtle, complex and unresolved. This study will attempt to shed light on these issues through the analysis of experimental results via unbiased procedures.

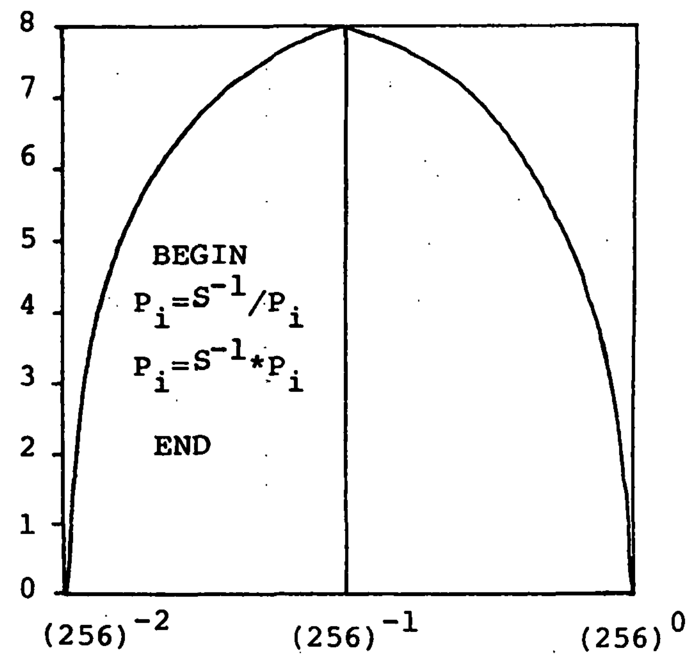
Figure 1

ALGORITHM BEHAVIOR
UTILIZING A HOMOGENEOUS
SAMPLE FRAME

SCALE=256



A.) Probabilities are unchanged; curve fits $\log_2 P_i$ but is not a measure of organization.



B.) Probabilities inverted at $(SCALE)^{-1}$; curve fits classical entropy function.

Figure 2

PROGRAM LISTING FOR IMPLEMENTATION OF
TRANSFORM ALGORITHM ON A
HEWLETT-PACKARD 34C HAND HELD CALCULATOR

```
C**  Probabilities (Bi) associated with pixel vectors are
C**  inserted into registers 0-8 prior to execution
C**  (i.e., product of multichannel probabilities for
C**  each pixel within the 9 pixel "cell").
C**  Minimum value for any given channel is 1/S^2

hLBL  A                                ÷
f(FIX)   7                             X
CF       0                             hLBL  3
2                                      STO  +9
f(LN)                                  ISG
STO     .0                             GTO   1
0                                      F?    0
R/S          ▶ ENTER # OF              GTO   4
STO     .1     CHANNELS                SF    0
O                                      RCL   9
R/S          ▶ ENTER RANGE             STO  .0
h(1/x)         OF GSV'S                GTO   0
RCL     .1                             hLBL  4
h(y^x)                                 RCL   9
STO     .2                             h(1/x)
hLBL     0                             RCL  .1
.00801                                 ÷
STO f  (I)                             RCL  .0
O                                      CHS
STO      9                             X
hLBL     1                             STO   9
RCL f  (i)                             RTN
RCL     .2
X◆Y                                    Display result is entropy
X>Y ?                                  of the contents of registers 0-8.
GTO      2                             Value is stored in register 9.
X = Y ?
GTO      2
÷
RCL     .2
X
hLBL     2
F?       0
GTO      3
ENTER
f (LN)
RCL     .0
```

Chapter III

IMPLEMENTATION AND EXPERIMENTAL RESULTS

This study utilized existing software at the Mississippi State University Remote Sensing Center for creation of a goebased information system (GIS) of the fixed grid type. Ground truth information concerning land cover and forest vegetation categories within the Talahala Wildlife Management Area in Jasper County, Mississippi were coded and input into the GIS. This data base served as the means by which image classification accuracy was accessed in a controlled fashion. In the techniques utilized in this study, unsupervised classification results utilizing Landsat MSS channels (3,4) over the study area for 5 May 1978, were coordinate-registered to the data base. These data were integrated as a variable class within the data base using software written for this study. Unsupervised classification results utilizing Landsat MSS channels (3,4), as well as synthetic transform channels Redundancy (MSS 3,4), and Entropy (MSS 3,4), were also coordinate-registered to the data base and integrated as a separate variable class. Classification accuracy for each classification technique was determined by tabulating the correspondence (i.e. the logical.AND.relation) of classification assignments to ground truth data on a pixel by pixel basis within the data base. Relative correspondence was quantified by means of the log-likelihood ratio test for goodness of fit, and the results of the study are presented at the end of this chapter.

The choice of MSS channels 3 and 4 for comparative classification assessment reflects constraints involving data accuracy and availability rather than empirical criteria related to the absolute suitability of these particular two channels for forest species type classification. MSS channels 1 and 2 had to be omitted because of severe scanner gain noise which prevented their implementation in the Entropy transform algorithm (See logistic problems, p. 47). Since the purpose of the study is to compare the relative accuracy of two classification techniques under identical controlled circumstances rather than to access the absolute or potential accuracy of any single technique, the arbitrary designation or number of channels utilized in the classification process is not considered to be a pivotal issue. The fundamental question addressed by this procedure is: Given a closed set of spectral information, can a transform applied to that set effect a significant and economically viable increase in the reliability of inferences based on that data set? Since the transform channels utilized are synthesized directly from the closed data set; i.e., no outside inferences or variables enter into the transform, no "new" information is added to the closed data set as a result of the transform. The transform process is one of "filtering" specific categories from the existing data set based on the spatial, as well as the spectral organization of the data set. The process is one of making available to the classification process information which would otherwise be lost, utilizing a transform which violates none of the mathematical foundations upon which the classification process is founded.

Image classification was accomplished through use of the LARSYS (Laboratory for Application of Remote Sensing) Image Processing System developed at Purdue University. In the technique utilized, spatially identical training areas from the raw multispectral data set and from the synthetic transform data set were first statistically "clustered" utilizing the LARSYS ISOCLS (Iterative Self-Organizing Clustering) processor, which is an unsupervised, maximum likelihood clustering algorithm. This accomplished, the entire study area for such data set was classified based on statistics derived from the ISOCLS processor over these identical sub-areas. When each classification was complete, the classified results were reformatted to LARSYS Universal data format for display and analysis, and were integrated as separate variable classes within the coordinate-registered data base. Since the scale ratio of the data base pixels to Landsat pixels was roughly compatible (2.5 acres to 1.1 acres), resampling of classification pixels for data base pixel determination was not deemed necessary or desirable in the coordinate registration process, and a one to one assignment was made.

Computer Procedures and Operational Techniques

Extensive software development was required for the synthesis of channels of spectral Entropy and Redundancy from Landsat MSS data. This development occurred over a ninety day period between late May and early August of 1980 through the generous support of the MSU Remote Sensing Center under the direction of Professor W. Frank Miller of the MSU Department of Forestry. Program descriptions and runstreams for executing channel synthesis on the Univac 1100/80 mainframe computer may be found in the Appendix.

The technique utilizes the concept of a "nearest neighbor" window, or (3 * 3 * N) matrix, where N equals the number of spectral channels represented within each pixel vector. The window, which "scans" the multispectral image in up to 4 data dimensions, represents a multipixel "cell" as discussed in the mathematical argument section. Pixels within the moving window serve as pointers to a lookup table of grayscale category probabilities for each spectral channel over the entire image. This table is constructed while making one full pass through the data set during the necessary reformatting of the data to pixel-interleaved format for the nearest-neighbor sampling scan.

In the transform process, the Entropy function is computed over all pixels within the moving cell using the previously described algorithm. The function value is then assigned to a location corresponding to the centroid pixel within the cell, the cell is shifted one pixel to the right, and the process repeats itself until the end of the scanline is reached. At the end of each scanline, the Redundancy is synthesized from the Entropy, and the values are then scaled and packed. The two channel synthetic scanline is then written onto tape along with the spectral channels used to synthesize it, at the users option. The next spectral scanline is then read, registered to the "window" array, or scan track, and the process repeats until the entire image has been transformed. While the array manipulations in the actual algorithm bear little resemblance to the above description due to the exploitation of computational economies in the quest for speed and efficiency, the result is directly analogous to that of a "scanning" window through the data. Available options in the software allow the user to requantize by equal probability the input data as well as output data to any desired number of categories less than or equal to 256. The program automatically adjusts key parameters within the transform algorithm to correspond to the specified input quantization levels chosen. The user may also omit the original MSS channels used to synthesize the transform channels on the output tape in order to conserve storage space, if such is desired.

Major computational economies are achieved through primitive spatial classification of image categories in terms of global image statistics computed during a single pass through the data set. Further discussion of this relatively simple technique is presented in the conclusion of this thesis and may be referenced there.

Investigative Results

Given the quality of the data utilized, and the necessity to omit two potentially useful channels after data smoothing because of residual scanner gain errors, with the result that only MSS channels 3 and 4 were utilized in the channel transform/synthesis process, the classification results were visually striking and beyond expectation. Classification results utilizing MSS channels 3 and 4 plus the Entropy (MSS 3,4) and the Redundancy (MSS 3,4) for 5 May 1978, yielded amazingly coherent imagery depicting bounded structural delineation of species classes within the forest community contained within the study area. In addition, the resultant classified imagery not only differentiated between species classes, such as hardwoods vs pine, but was apparently successful in breaking down each species class into coherent structural sub-categories based on ecological environment, stand history, and degree of mixture with other species classes.

Of particular interest upon visual inspection of the transform classification results is the coherence of the sub-categories within a species class in terms of their bounderies relative to the site's ecological environment, as well as to other sub-categories. For example, bottomland hardwoods were broken into no fewer than 6 sub-classes, each subclass being composed largely of discrete features nested within other hierarchically superordinate sub-classes. What is apparantly being revealed is the gradation of spectral variability within the forest species class in response to local variations within the habitat, such as soil fertility, height of water table and topography. Plate 1, which is a falsecolor image of the transform classification taken from the color viewer, illustrates this effect convincingly. In contrast, MSS classification results utilizing, only channels 3 and 4, while accurately delineating species classes, generally lacked such sub-class differentiation and, where it occured, it tended to form "blobs" with many apparently random arrangements of sub-category pixels (see Plate 2). MSS classification utilizing all four MSS channels from the data set was rendered incoherent by the poor quality of the multispectral data available (see Plate 3). Plate 4 depicts a close-up view of a central bottomland stream junction in the study area for comparison of hardwood sub-category definition between the two classification techniques. Plates 5 and 6 present the entire study area viewed through the entropy transform, and the redundancy transform channels; while Plate 7 is a medium altitude color infrared photo of the study area. Even a cursory visual comparison of transform classification results with classification results utilizing "classical" MSS spectral channels (3,4) indicates exciting prospects for new approaches in forest classification and inventory as the quality of available MSS data improves.

Before the comparative accuracy of MSS classification results and transform channel results could be accurately assessed, the actual correspondence of classification

PLATE 1

Color coded classification results utilizing MSS channels (3,4), redundancy (3,4) and entropy (3,4). Area represents central region of the Tallahala Wildlife Management Area in Jasper Co., Mississippi utilizing Landsat imagery taken 5 May 1978.

PLATE 2

Color coded classification results utilizing MSS channels (3,4). Area is identical to Plate 1: Tallahala Wildlife Management Area, Jasper Co., Mississippi, 5 May 1978.

PLATE 3

Color coded classification results utilizing MSS channels (1,2,3,4). Area is identical to Plate 1; incoherent due to scanner noise: Tallahala Wildlife Management Area, Jasper Co., Mississippi, 5 May 1978.

PLATE 4

Transform Classify MSS (3,4) Classify

Comparison of sub-category definition at central bottomland junction: Tallahala Wildlife Management Area, Jasper County, Mississippi, 5 May 1978.

PLATE 5

Entropy of MSS channels (3,4): Tallahala Wildlife Management Area, Jasper County, Mississippi, 5 May 1978. (64 Level Grayscale).

PLATE 6

Redundancy of MSS channels (3,4): Tallahala Wildlife Management Area, Jasper County, Mississippi, 5 May 1978. (64 Level Grayscale)

PLATE 7

Medium altitute color infrared imagery of Tallahala Game Management Area, Jasper County, Mississippi, June 1974. Scale 1:120,000.

"clusters" to data base ground cover categories had to be visually established. Unfortunately, there seems to be no escape from the introduction of bias in assignment of statistical clusters to arbitrary ground cover categories on a map. Arbitrarily defined ground cover categories, when mapped, cannot accomodate the intense structural intermixing of categories due to topographic and hydrological factors in many areas, nor do their boundaries necessarily conform to existing "statistical" boundaries. Thus, while parts of the image were reflecting coherent, structurally intermixed categories, the data base indicated only the dominant category for the given area. The procedure of tabulating data base to classification correspondence therefore does not determine the absolute accuracy of a classification procedure, but only attempts to assess the accuracy of one procedure relative to another. Final determination of image classification accuracy before inclusion into a geobase variable class can only be entrusted to persons intimately familiar with the classified area and with training in ecological science.

The tabulated results of the study begin with Table 1. While the figures apparently lend some support to the hypothesis that channel synthesis utilizing MSS data will increase classification reliability relative to raw MSS imagery, the tabulated results are far from conclusive. To this researcher they suggest that at the gross levels of categorization and resolution used as a control in the data base, a subtle but statistically significant increase in accuracy is evident, even after cluster re-assignment, when utilizing transform channels in the classification process.

Table 2 represents the results of comparing the two classification procedures with ground truth data by using the log likelihood test for goodness of fit. The smaller the absolute value of the test, the better the fit between the two sets of data. The dismal scores for both classified data sets tested against the ground truth data set result from the incompatibility of the two sampling frames compared; manually mapped and digitized boundaries in the ground truth set were tested against statistically clustered, unsmoothed multispectral data. This effect is discussed in the conclusion. The magnitude of the unavoidable sampling error prevents hard conclusions regarding either data set based upon purely quantitative criteria.

Table 1

CLASSIFICATION TO GROUND TRUTH: COMPARATIVE ACCURACY UTILIZING TWO TECHNIQUES

Ground Cover Category	Database Control	MSS(3,4) Incidence	ENT/RDN(3,4) Incidence
1. Mature Bot. Hardwd	3380	2635	2400
2. Other Mat. Hardwd	2024	438	470
3. Hardwood/Pine	5512	1475	3224
4. Void	0	0	0
5. Pine (dense)	4672	1404	2122
6. Open Pine	345	0	0
7. New Pine (dense)	12	0	0
8. Scrubland	637	0	0
9. Open Land	1899	553	26
10. Water	32	0	0
TOTAL	18,513	6505	8242
PERCENT	100	35	44

Percent increase in relative classification accuracy utilizing transform channels is 9 percent.

Table 2

LOG-LIKELIHOOD TEST FOR GOODNESS OF FIT;
COMPARISON OF TWO TECHNIQUES

Log-Likelihood Analysis of Results

$$G=2\sum_{i=1}^{n} f_i \ln \frac{f_i}{F_i}$$

equivalently

$$G=4.60517\sum_{1}^{n} f_i \log_{10} \frac{f_i}{F_i}$$

f_i = Observed Frequency
F_i = Control Frequency
N = Number of Classes
DF = 10

Transform Log Likelihood

-10,046.754 H, Rejected

MSS(3,4) Log Likelihood

-11,282.352 H, Rejected

Results: Magnitude of Error Precludes Meaningful Comparison. Inconclusive.

Chapter IV

CONCLUSION

Unfortunately, the quantitative results do not conclusively determine the efficacy of one image classification approach to another at the level of category definition existing in the ground truth control set. Upon reflection, it becomes apparent that the assignment of classified image clusters to arbitrary, pre-defined ground cover categories is a particulary weak point in the technique, since the clusters sometimes span different, though related, data base categories as well as differentiating sub-categories not included in the ground truth control set. Choosing to assign a cluster to one category instead of another under these circumstances often involves the same type of decision as drawing a boundary to a category on a map or aerial photograph; the best attempts serve merely to minimize error but cannot eliminate it, especially as the number of clusters becomes large. The primary manifestion of this problem appears in the discordance between the boundaries of categories in the ground truth data set and the boundaries of classification clusters. This boundary problem eliminated many correct assignments of correspondence between ground truth and classification results, since classified image clusters generally occupied the central regions of designated data base ground-truth categories, but the existing extension of ground truth category boundaries within the data base beyond actual cluster boundaries obliterated correct classifcation correspondence with large areas of neighboring ground truth categories of a different class. This effect is especially apparant when viewing line printer maps of the logical .AND. relation between ground truth categories and cluster assignments from classified imagery. This is, in itself, a valuable finding, since it suggests a modification of technique and assumptions in constructing data bases utilizing classified imagery as a key variable.

In the construction of geobase data systems which depend upon classified imagery as key variable classes in the total ecological data network, the most promising approach to inventory of ground cover categories within such a system would be to define data base categories in terms of the characteristics of dominant clusters within reliably classified imagery. This procedure would increase the correspondence between boundarys as necessary divisions of land cover types, and as delimiters of actually occuring clusters in classified image data. Total acreage estimates within each "cluster/category" would therefore be greatly enhanced in accuracy.

In the opinion of the writer, the finding of central importance to this study appeared in the prominent subcategory definition within existing "global" cover type categories in the data base, as well as the coherence and hierarchical structure of such subcategories. This effect is well below the threshold of definition of the data base ground truth variable class, and therefore did not appear as a factor in the findings since all sub-categories were simply assigned to the closest hierarchically superordinate data base landcover category. The quantified results therefore do not address this finding conclusively.

In conclusion, while unmodified MSS imagery, and MSS imagery with the addition of transform channels of Entropy and Redundancy are roughly equivalent in the reliability of their classifications of gross forest species classes, transform imagery apparently yields markedly superior sub-category definition within such ground cover units, and this sub-definition appears to be in consistent agreement with known detailed topographic and ecological information regarding the forest system. It is therefore tentatively contented that the transform imagery does indeed yield a marked increase in accuracy relating to the state of the forest ecosystem; however, this increase occurs as the hierarchical sub-division within gross land cover categories, and was therefore below the threshold of measurement of the data base ground truth variable class. The reader is referred to a comparative inspection of the color plates.

Futher research into this area is planned.

Logistic Problems of Interest to Future Investigations of this Nature

A major difficulty encountered with the utilizaton of global image statistics in primitive image feature classification is the extreme sensitivity of the algorithm to electronic gain malfunctions in the Landsat scanner mechanism, which badly skew image statistics. In currently available raw Landsat MSS data, this problem can only be described as massive, and even images of "good" quality tend to be noticably striped in one or more spectral channels. The result is badly skewed global statistical distributions over the study area for the effected channels. Transform channels computed from such data result in a pronounced, highly selective, striped pattern within specific image categories, even when the striping is visually subtle in the raw data set.

The extreme sensitivity of the algorithm to global errors requires that the data be of consistent quality over large areas. Since the acquisition and dispensation to users of data of acceptable quality is an area in which massive improvement is possible, this weakness of the global statistical approach is not considered to be insurmountable. To cope with the difficulty of gain-striped data within a limited time frame, an algorithm was quickly developed which successfully smoothed moderately striped data by utilizing a "moving average" technique, which compared the mean of each individual scanline to the global mean for each channel, as well as the computed mean for the preceeding scanline in that channel. Each pixel in the scanline for each channel was then adjusted by the resulting channel specific constant. Severely striped data proved unsalvagable, since gain errors within the scanline were non-linear, or the data simply did not exist. Noise other than a simple gain constant caused a propogating "ripple-effect" in the smoothing process resulting in residual striping and rendering the data unusable for global analysis. The algorithm does work for moderately striped data, however, and MSS channels 3 and 4 for 5 May 1978 over the study area were in this manner rendered of marginal but acceptable global statistical quality for channel synthesis.

To obtain MSS data of usable quality for future investigations of this nature, extensive preprocessing of currently available Landsat MSS data utilizing gain calibration information, as well as "in-line" spatial statistics is necessary if the described techniques, as well as statistical clustering classification techniques in general, are to be utilized on a regular basis in which reliability is an issue. Placing this significant software development and computational burden on the user does little, however, to encourage the expansion of remote sensing technology into the private sector on a broad scale. If the quality of accessable data is an indication of existing agencies' future committment towards this end, then there is little probability of attaining this goal in the forseeable

APPENDIX

PROGRAM DESCRIPTIONS AND COMPUTER RUNSTREAMS

The Destriping Runstream

```
* RUNSTREAM FOR DESTRIPING LARSYS FORMAT
* LANDSAT MSS DATA BY USE OF "MOVING AVERAGE"
* TECHNIQUE.  HISTOGRAMS AND ALL CHANNEL STATISTICS, BOTH
* PRIOR TO DESTRIPING AND AFTER DESTRIPING, ARE
* DISPLAYED.  OPTION PERMITS OMITTING DESTRIPING
* OPERATION, AND SIMPLY DUMPS HISTOGRAM AND CHANNEL STATISTICS
* OVER STUDY AREA FOR ANALYSIS.
* VARIABLES:

     IYST - Start scanline on LARSYS tape
     IXST - Start pixel on LARSYS tape
     NUMSLY - Number of scanlines to process
     NPIX - Number of pixels to process
     NBANDS - Number of spectral bands (Max = 4)
     IBANDS - Designated Bands (no limit)
     IFILE - Input file on LARSYS tape (1-N)
     IOFILE - Output file of destriped data (1-N)
     IWCTL - If (0) dump stats only; no destriping
             If (1) dump stats and destrip data

*RUNSTREAM (UNIVAC 1100/80)
@RUN USERID, ACCOUNT #, CARTER, 7, 50
@ASG,A LANDSAT.
@ASG,TJH CCT, 16N, "LARSYS INPUT TAPE"
@ASG,TJH PILTAP, 16N, "SCRATCH TAPE"
@MSG,W PLEASE LOAD "SCRATCH TAPE" RING IN
@ASG,TJH FUNTAP, 16N,"DESTRIPED OUTPUT TAPE"
@MSG,W PLEASE LOAD "OUTPUT TAPE" RING IN
@FTN,S LANDSAT. DESTRIPE, TPF$.DST
@MAP,I
IN TPF$.
IN LANDSAT.FTNTAPES
LIB SYS$*FTNLIB$.
@XQT
IYST,IXST,NUMSLY,NPIX,NBANDS,(IBANDS(I),I=1,NBANDS),
IFILE,IOFILE,IWCTL
@FIN
*MAXIMUM WIDTH OF AREA IS 3499 PIXELS.
*IF IWCTL = 0, NO NEED FOR SCRATCH TAPE OR
*OUTPUT TAPE; AREA STATISTICS AND HISTOGRAMS
*ONLY ARE DISPLAYED
```

Entropy//Redundancy Transform Runstream

```
* RUNSTREAM FOR GENERATING SYNTHETIC
* CHANNELS OF ENTROPY AND REDUNDANCY
* UTILIZING LARSYS FORMAT LANDSAT MSS
* DATA. HISTOGRAMS AND STATISTICS OF INPUT
* CHANNELS AS WELL AS SYNTHESIZED
* CHANNELS ARE DISPLAYED.  MAIN PROGRAM
* DRIVES SAMPLING AND I/O ROUTINES.
* SUBROUTINE FUNCT WHICH PERFORMS THE
* TRANSFORM ON THE DATA "WINDOW" MAY
* BE MODIFIED AND EXTERNALLY COMPILED
* FOR FUTURE CHANNEL SYNTHESIS STUDIES.
* MODIFICATION OF MAIN PROGRAM IS NOT
* RECOMMENDED.
* VARIABLES:

     IYST - Start scanline on LARSYS tape
     IXST - Start pixel on LARSYS tape
     NUMSLY - Number of scanlines to process
     NPIX - Number of pixels to process
     NBANDS - Number of spectral bands (MAX = 4)
     IBANDS - Designated bands (no limit)
     IFILE - Input file on LARSYS tape (1-N)
     IOFILE - Output file on output tape (1-N)
     IWCTL - If (0) write synthetic channels only
             If (1) write MSS and synthetic channels
     NSYNB - Number of synthetic channels (2)
     NIQ - Number of input requantization levels
     NOQ - Number of output requantization levels
           If either NIQ or NOQ equal 256 NO
           corresponding requantization occurs

* RUNSTREAM (UNIVAC 1100/80)
@RUN USERID, ACCOUNT, CARTER, 7, 50
@ASG,A LANDSAT.
@ASG,TJH CCT, 16N, "LARSYS INPUT TAPE"
@ASG,TJH PILTAP, 16N, "SCRATCH TAPE"
@MSG,W PLEASE LOAD, "SCRATCH TAPE" RING IN
@ASG,TJH FUNTAP,16N, "OUTPUT TRANSFORM TAPE"
@MSG,W PLEASE LOAD "OUTPUT TAPE" RING IN
@FTN,S LANDSAT. IMAGENTROPY, TPF$.IGE
@MAP,I
IN TPF$.
IN LANDSAT.FTNTAPES
LIB SYS$*FTNLIB$.
@XQT
IYST,IXST,NUMSLY,NPIX,NBANDS,(IBANDS(I),I=1,NBANDS),
IFILE,IOFILE,IWCTL,NSYNB,NIQ,NOQ

* MAXIMUM WIDTH OF AREA IS 3499 PIXELS
```

Entropy and Redundancy Computation Subprogram (FUNCT) Program Listing

```
      SUBROUTINE FUNCT(NOQ,NIQ,NBANDS,NPIX,PTABLE,
     +                  WINDOW, IFLINE, NFPIX, IRD)
      DIMENSION  PTABLE(265,7), SCANLN(1000), CONST(1000)
      INTEGER  WINDOW(1000,3), IFLINE(1000), IRLINE(1000)
      DATA   IROLL/32767/, ILN/0/, IFLAG/0/
      IF(IFLAG.NE.0)GO TO 5
      NFPIX = NPIX - 2
      NPX = NPIX + 1
      DO 10 IB = 1,NBANDS
        HLIM = 1.0 / PTABLE(262,IB)
        IF(NIQ.LT.255)HLIM = 1.0 / NIQ
        HLX = HLIM**2.0
        IST = IFIX(PTABLE(260,IB))
        IFN = IFIX(PTABLE(261,IB))
        DO 10 IX = IST,IFN
          IF(PTABLE(IX,IB).LT.HLX)PTABLE(IX,IB) = 0.0
          IF(PTABLE(IX,IB).EQ.0.0)GO TO 10
          IF(PTABLE(IX,IB).GE.HLIM)GO TO 10
            TEMP = HLIM / PTABLE(IX,IB)
            PTABLE(IX,IB) = HLIM * TEMP
10    CONTINUE
      ZLOG2 = LOG(2.0)
      FACTOR = LOG(NIQ)  / ZLOG2
      FACTH = FLOAT(NIQ)
      IF(NIQ.LT.255)GO TO 12
         FACTH = 128.0
         FACTOR = LOG(FACTH)  / ZLOG2
12    IF1 = NBANDS + 1
      IF2 = NBANDS + 2
      IWR = IRD
      IFLAG = 7
5     CONTINUE
      DO 53 IX = 2,NPX
        SCANLN(IX)  = 0.0
53      CONST(IX)  = 0.0
      BSUBI  = 1.0
      DO 70 IX = 2,NPX
        DO 65 IW = 1,3
          DO 60 IB = 1,NBANDS
            IBIT = ((4 - IB)  * 8)  + 5
            IGSV  = BITS(WINDOW(IX,IW),IBIT,8)
            IF(PTABLE(IGSV,IB).EQ.0.0)GO TO 60
            BSUBI  = BSUBI * PTABLE(IGSV,IB)
60        CONTINUE
          CONST(IX)  = CONST(IX)  + BSUBI
          TEMP = BSUBI
          BSUBI  = LOG(BSUBI)
          BSUBI  = BSUBI  / ZLOG2
          BSUBI  = BSUBI  * TEMP
          SCANLN(IX)  = SCANLN(IX)  + BSUBI
          BSUBI  = 1.0
```

```
65      CONTINUE
70    CONTINUE
      TSP = SCANLN(2)
      BSP = CONST(2)
      DO 75 IX = 3,NPIX
        IX2 = IX + 1
        ESI = ESI + TSP
          ESI = ESI + SCANLN(IX)
            ESI = ESI + SCANLN(IX2)
        BSI = BSI + BSP
          BSI = BSI + CONST(IX)
            BSI = BSI + CONST(IX2)
        BSI = 1.0 / BSI
        BSI = BSI / NBANDS
        BSI = -1.0 * BSI
        TSP = SCANLN(IX)
        BSP = CONST(IX)
        SCANLN(IX) = ESI * BSI
        REDN = SCANLN(IX) / FACTOR
        REDN = 1.0 - REDN
        IRLINE(IX) = IFIX(FACTH * REDN) + 1
        IF(IRLINE(IX).LT.1)IRLINE(IX) = 1
        IF(IRLINE(IX).GT.256)IRLINE(IX) = 256
        IFLINE(IX) = IFIX(2.0**SCANLN(IX))
        IF(IFLINE(IX).LT.1)IFLINE(IX) = 1
        IF(IFLINE(IX).GT.256)IFLINE(IX) = 256
        ESI = 0.0
        BSI = 0.0
75    CONTINUE
      IF(NOQ.GT.255)GO TO 77
      CALL KEQUAN(IFLINE(3),256,NOQ,NFPIX)
      CALL KEQUAN(IRLINE(3),256,NOQ,NFPIX)
77     CONTINUE
      DO 80 IX = 3,NPIX
        PTABLE(IRLINE(IX),IF1)=PTABLE(IRLINE(IX),IF1)+1.0
        IRLINE(IX) = IRLINE(IX) - 1
        PTABLE(IFLINE(IX),IF2)=PTABLE(IFLINE(IX),IF2)+1.0
        IFLINE(IX) = IFLINE(IX) - 1
80    CONTINUE
      CALL MRGARY(NPIX,IRLINE(2),IFLINE(2))
      IFLINE(1) = IROLL
      ILN = ILN + 1
      BITS(IFLINE(1),1,20) = ILN
      IF(ILN.LT.25)WRITE(6,100)(IFLINE(IB),IB=1,15)
      RETURN
      ENTRY NLINES
      WRITE(6,101)ILN
      RETURN
100   FORMAT('0**FUNCT ',O12,14(1X,O6))
101   FORMAT('0****LINES WRITTEN ON OUTPUT TAPE = ',I4)
      END
```

<u>Cluster Analysis (ISOCLS) Runstream</u>

```
* RUNSTREAM TO ISOCLS OUTPUT TRANSFORM
* TAPE, UTILIZING LARSYS ISOCLS PROCESSOR.
* THIS PROCESSOR STATISTICALLY CLUSTERS BY
* MAXIMUM LIKELIHOOD THE INPUT TAPE BASED
* ON MULTISPECTRAL CORRESPONDENCES WITHIN
* EACH PIXEL VECTOR

* RUNSTREAM (UNIVAC 11/80)
@RUN USERID, ACCOUNT, CARTER, 7, 150
@ASG,A JSC.
@ASG,TJH 3, 16N, "LARSYS INPUT TAPE"
@ASG,TJH 4, 16N, "STATISTICS TAPE"
@MSG,W PLEASE LOAD "STATISTICS TAPE" RING IN
@ELT,LI INPUT
$ISOCLS
DATAFILE  FILE=1
CLASSES   1
CHANNELS  DATA=1,2,3,4, STAT= 1,2,3,4
OPTION    STATS
STATFILE  OUTPUT/UNIT=4, FILE=1
STDMAX    4.5     SEE LARSYS MANUAL
DLMIN     3.5     SEE LARSYS MANUAL
*END*
CLASSNAME ENTROPY
AREA1     (1,1)(NWX,NWY),(NEX,NEY)(SEX,SEY),(SWX,SWY)
$END*
$EXIT
@XQT JSC.LARSAA
@ADD INPUT
```

Classification Analysis ($CLASSIFY) Runstream

```
* RUNSTREAM TO CLASSIFY STUDY AREA
* UTILIZING SPECTRAL STATISTICS GENERATED
* BY $ISOCLS PROCESSOR AND CONTAINED
* IN STATISTICS TAPE

* RUNSTREAM
@RUN USERID, ACCOUNT, CARTER, 10, 80
@ASG,A JSC.
@ASG,TJH    3,16N, "LARSYS INPUT TAPE"
@ASG,TJH    4,16N, "STATISTICS TAPE"
@ASG,TJH    2,16N, "CLASSIFY OUTPUT TAPE"
@MSG,W PLEASE LOAD CLASSIFY TAPE RING IN
@ELT,LI     INPUT
$CLASSIFY
OPTION      STATS
STATFILE    UNIT=4, FILE=1
CHANNELS    STAT=1,2,3,4,DATA=1,2,3,4
DATAFILE    FILE=1
HED1        HEADER COMMENT
HED2        HEADER COMMENT
*END*
FIELD       (1,1)(NWX,NWY),(NEX,NEY),(SEX,SEY),(SWX,SWY)
$END*
$EXIT
@XQT JSC.LARSAA
@ADD INPUT
```

$CLASSIFY Output Reformat to LARSYS Runstream

```
* RUNSTREAM TO REFORMAT $CLASSIFY
* OUTPUT TAPE TO LARSYS FORMAT FOR
* STORAGE, DISPLAY AND ANALYSIS

* RUNSTREAM (UNIVAC 1100/80)
@RUN USERID, ACCOUNT, CARTER, 5
@ASG,A LANDSAT.
@ASG,TJH 9,16N, "$CLASSIFY TAPE"
@ASG,TJH OUTTAPE, 16N, LARSYS OUTPUT TAPE
@MSG,W PLEASE LOAD OUTPUT TAPE RING IN
@FOR,S LANDSAT.REFCLASS, TPF$.RFC
@MAP,I
IN TPF$.
IN LANDSAT.FORTAPES
@XQT
OUTPUT FILE (1-N)

* OUTPUT TAPE IS IN STANDARD LARSYS
* FORMAT AND MAY BE DISPLAYED ON
* MINICOMPUTER SYSTEM AND USED AS
* INPUT INTO CALUP DATABASE (NEXT SECTION)
```

UDCLASSIF1 - Database Update With Classified Imagery

```
* RUNSTREAM TO CREATE FILE OF LARSYS
* CCT COORDINATES WHICH CORRESPOND TO
* CALUP DATABASE PIXELS.
* PROGRAMS CONVERT BETWEEN CCT AND DATABASE CONTROL POINT
* COORDINATE SCHEMES.
* VARIABLES:

     NCP - Number of control points utilized
     NSLY - Number of scanlines in classified area
     IXWIDT - Width of study area in tape pixels

* RUNSTREAM (UNIVAC 1100/80)
@RUN USERID, ACCOUNT, CARTER 5
@ASG,A LANDSAT.
@PREP LANDSAT.
@ASG,A DATABASE.
@USE INPUTDB, DATABASE
@ASG,T 9.,f///250
@FOR,S LANDSAT. UDCLASSIF1, TPF$.UD1
@MAP,I
IN TPF$.
LIB LANDSAT.
@XQT
NCP,NSLY,IXWIDT
     CCT CONTROL POINT COORDINATES (1-NCP)
     CALUP DATABASE COORDINATES (1-NCP)
@ASG,C SORTC1.,F///250
@SORT,SM
FILEIN=9.
FILEOUT=SORTC1.
RSZ=20
KEY=6,5,A,D:1,5,A,A
RECORD=100
@EOF
@FIN

* THE FILE SORTC1 NOW CONTAINS SORTED POINTERS
* TO PIXELS ON CLASSIFIED TAPE WHICH
* CORRESPOND TO CALUP DATABASE PIXELS.
* SORTC1 IS USED AS INPUT, ALONG WITH
* CLASSIFIED TAPE, TO UDCLASSIF2.
```

UDCLASSIF2 - Database Update With Classified Imagery

```
* RUNSTREAM TO CREATE A SORTED FILE OF
* CALUP DATABASE COORDINATES WITH ASSIGNED
* SUB-VARIABLE VALUES WHICH ARE DETERMINED
* BY USER RELATIVE TO GROUND TRUTH CORRES-
* PONDENCES TO CLUSTER VALUES.
* VARIABLES:

     NUM2 - Variable designation of classified tape in
            database scheme
     IFILE - File on tape containing classified data (1-N)
     NSLY - Number of scanlines in classified data
            (same as UDCLASSIF1)
     IRANG - Length of re-sampling tansect centered on
             designated coordinate for determination of
             cluster value.  If (IRANG = 0) then no
             resampling

* RUNSTREAM (UNIVAC 11/80)
@RUN USERID, ACCOUNT, CARTER, 5
@ASG,A LANDSAT.
@ASG,A SORTC1.
@USE 7, SORTC1.
@ASG,T 9., F///250
@ASG,TJH CCT,16N, "LARSYS FORMAT CLASSIFIED TAPE"
@FOR,S LANDSAT.UDCLASSIF2, TPF$.UD2
@FOR,SI
     SUBROUTING CLSCOV(IC)
C** SUBROUTINE LOGICALLY ASSIGNS CALUP SUB-
C** VARIABLE VALUE BASED ON CLUSTER VALUE (IC)
C** ON TAPE AND PLACES RESULT IN (IC)
     END
@MAP,I
IN TPF$.
IN LANDSAT. FORTAPES
@XQT
NUM2, IFILE, NSLY, IRANG
@ASG,C DBCLUST.,F///250
@SORT,SM
FILEIN=9.
FILEOUT=DBCLUST.
RSZ=26
KEY=7,5,A,D:2,5,A,A
RECORD=200
@EOF
@FIN

* FILE DBCLUST CONTAINS SORTED UPDATE FOR
* CALUP DATABASE AND IS TREATED AS A CORR-
* ECTIONS FILE (ITYPE=1) USING THE CALUP
* UPDATE PROGRAM. DELETE FILES WHEN
* SUCCESSFUL UPDATE IS VERIFIED.
```

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

Ashby, W. Ross. An Introduction to Cybernetics. New York: John Wiley and Sons Incorporated, 1956.

Cornsweet, Tom N. Visual Perception. New York: Academic Press, 1970.

Hall, Ernest L. Computer Image Processing and Recognition. New York: Academic Press, 1979.

Julez, Bela. Foundations of Cyclopean Perception. Chicago: University of Chicago Press, 1971.

Khinchin, A. I. Mathematical Foundations of Information Theory. New York: New York University, 1957.

Moles, Abraham. Information Theory and Esthetic Perception. Chicago: University of Illinois Press, 1966.

Rosenfeld, A. and Kak, A. Digital Picture Processing. New York: Academic Press, 1976.

Shannon, Claude E. and Weaver, Warren. The Mathematical Theory of Communication. Chicago: University of Illinois Press, 1949.

Uttal, William R. An Autocorrelation Theory of Form Detection. New York: John Wiley and Sons Incorporated, 1975.

Zar, Jerrold H. Biostatistical Analysis. Englewood Cliffs: New Jersey: Prentice-Hall Incorporated, 1974.

Articles

Batty, M. "Entropy and Spatial Geometry." Area 4 (1972): 230-236.

Bryant, Jack. "On the Clustering of Multidimensional Pictoral Data." Pattern Recognition 2 (1979): 115-125.

Berthold, K. P. H. and Bachman, B. L. "Using Synthetic Images with Surface Models." Communications of the ACM 21 (November 1978) 914-924.

Carlton, S. G. and Mitchel, O. R. "Image Segmentation Using Texture and Graylevel." In Proceedings, IEEE Computer Society Conference on Pattern Recognition and Image Processing. p. 387-391. Long Beach, California, IEEE Computer Society, 1977.

Chapman, G. P. "Application of Information Theory to the Analysis of Population Distributions in Space." Economic Geography 46 (1970): 317-331.

Frieden, Roy. "Restoring With Maximum Likelihood and Maximum Entropy." Journal of the Optical Society of America 62 (April 1972): 511-518.

Gonzalez, R. C. and Thomason, M. G. "Tree Grammars and Their Application to Syntactic Pattern Recognition." Arlington, Virginia: Office of Naval Research, 1974.

Haken, H. "Pattern Formation and Pattern Recognition - An Attempt At A Synthesis." in Dynamic Systems and Pattern Recognition. p. 2-15. New York: Springer-Verlag, 1979.

Hubel, David H. and Wiesel, Torsen N. "Brain Mechanisms of Vision." Scientific American (August 1979): 150-163.

Jaynes, E. T. "Information Theory and Statistical Mechanics." Physical Review 106 (1957): 620-630.

Jaynes, E. T. "Gibbs vs. Boltzmann Entropies." American Journal of Physics 33 (1965): 391-398.

Julez, Bela. "Texture and Visual Preception." Scientific American (February 1965): 38-48.

Leopold, Luna B. and Langbein, Walter B. "The Concept of Entropy in Landscape Evolution." General Systems 9 (1964): 24-43.

Margalef, Ramon D. "Information Theory in Ecology." General Systems 3 (1958): 36-71.

Peterson, David L. "Issues Arising From the Demonstration of Landsat Based Technologies to Inventories and Mapping of Forest Resources of the Pacific Northwest." NASA: Ames Research Center, Moffett Field, California, 1978.

Rachevsky, N. "Life, Information Theory and Topology." Bulletin of Mathematical Biophysics 17 (1955): 229-235.

Sheppard, E. S. "Entropy, Theory Construction and Spatial Analysis." Papers of the Regional Science Association (1976): 741-752.

Slagle, James. "Experiments With Some Algorithms That Find Central Solutions for Pattern Classificaton." Communications of the ACM 22 (March 1979): 178-183.

Stockmeyer, L. J. and Chandra, A. K. "Intrisically Difficult Problems." Scientific American 240 (May 1979): 140-159.