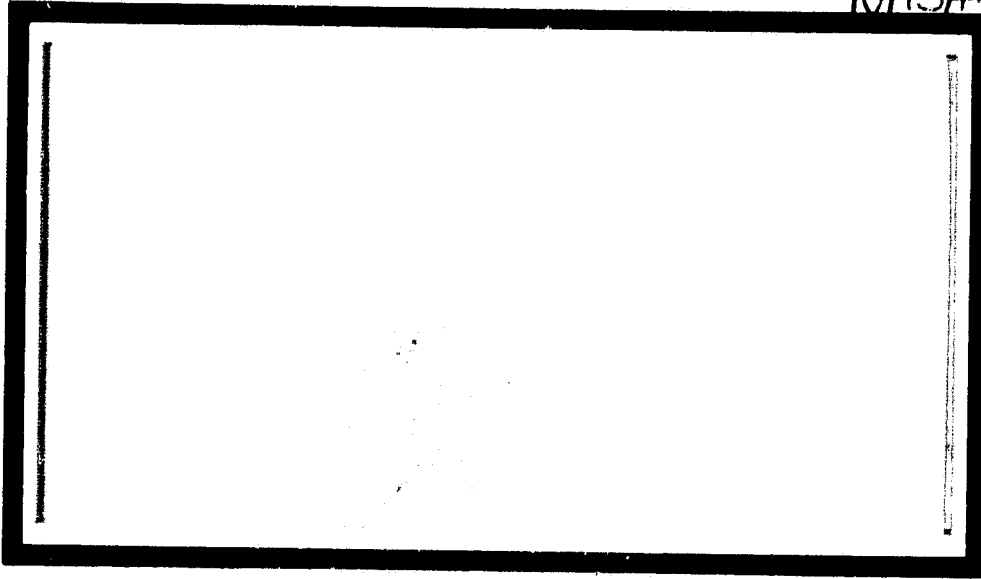


N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM
MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT
CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED
IN THE INTEREST OF MAKING AVAILABLE AS MUCH
INFORMATION AS POSSIBLE

NASA-CR-160946



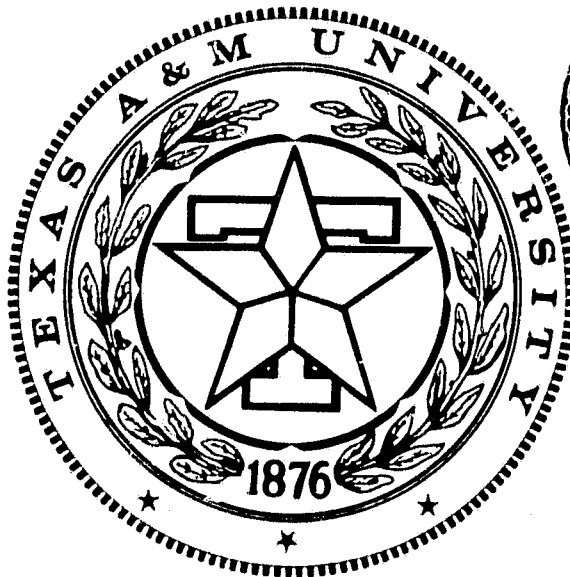
(NASA-CR-160946) BASIC RESEARCH PLANNING IN
MATHEMATICAL PATTERN RECOGNITION AND IMAGE
ANALYSIS Final Report (Texas A&M Univ.)
77 p HC A05/MF A01

N81-23561

CSSL 05B

Unclas
42287

G3/43



DEPARTMENT OF MATHEMATICS

TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS

BASIC RESEARCH PLANNING IN
MATHEMATICAL PATTERN RECOGNITION
AND
IMAGE ANALYSIS

Contract NAS 9-15964

January 1981

FINAL REPORT

Prepared for

Space and Life Sciences Directorate
NASA/Johnson Space Center
Houston, Texas 77058

by

Jack Bryant
L. F. Guseman, Jr.
Co-Principal Investigators
Department of Mathematics
Texas A&M University
College Station, Texas 77843

1.0 Introduction

2.0 Mathematical Pattern Recognition and Image Analysis

2.1 Preprocessing

2.1.1 Geometric

2.1.2 Radiometric

2.2 Digital Image Representation

2.2.1 Spatial

2.2.2 Spectral

2.2.3 Temporal

2.2.4 Syntactic

2.2.5 Ancillary

2.3 Object Scene Inference

2.3.1 Image Partitioning

2.3.2 Proportion Estimation

2.3.3 Error Models

2.4 Computational Structures

2.4.1 Parallel Processing

2.4.2 Image Data Structures

2.5 Continuing Studies

2.5.1 Polarization

2.5.2 Computer Architectures and Parallel Processing

2.5.3 Applicability of "Expert" Systems to Interactive Analysis

References

Appendix

BASIC RESEARCH PLANNING IN MATHEMATICAL PATTERN RECOGNITION
AND IMAGE ANALYSIS

1.0 INTRODUCTION

In fiscal year 1980, the National Aeronautics and Space Administration initiated a planning study to develop a program for basic research that could be initiated in fiscal year 1981 and continued for a five- to ten-year period. The planning study was sponsored by the Renewable Resources Branch of the Resources Observation Division of the Office of Space and Terrestrial Applications (OSTA) and coordinated by the Space and Life Sciences Directorate of the Johnson Space Center.

The purpose of the study was to define the basis for a research program which would significantly broaden and strengthen the foundation for continued technological development and support future NASA projects using aerospace remote sensing for mapping and monitoring the Earth's renewable resources.

The basic research problems related to using remote sensing can be generally grouped into the following research categories:

1. Scene Radiation and Atmospheric Effects Characterization
2. Mathematical Pattern Recognition and Image Analysis
3. Electromagnetic Measurements and Data Handling
4. Information Utilization and Evaluation.

The acquisition of information concerning the existence, state, and/or condition, and location of the Earth's renewable resources utilizing aerospace remote sensing is based on these four interrelated categories.

Here, remote sensing is the observation (measurement) of a portion of the Earth's surface (object scene) through the intervening medium(ia) of the surrounding atmosphere. A typical object scene is composed of physical material (scene radiators) reflecting sunlight or emitting electromagnetic radiation characteristic of and dependent upon material type, condition, and configuration. Scene radiation is usually significantly altered by the atmosphere through which it must pass to reach a sensor located in space some distance away. Notwithstanding atmospheric and other effects, a sensor collects some portion of the energy radiated by the scene and converts it to electrical signals representative of that incoming energy. Whether the sensor is a human eye or a manmade electro-optical measurement instrument, it has particular response characteristics that help determine the inherent information content of the measured electromagnetic energy. The measurements of the object scene radiation determined by the sensor give rise to an associated digital image which can be analyzed in terms of relationships which exist between the radiated energy (from the object scene) and the characteristics of the digital image. Automated approaches for analyzing digital images make use of pattern recognition techniques based on mathematical models which incorporate the spatial, temporal, spectral and polarization characteristics of the digital image. The digital image contains "noise" due to atmospheric, sensor, communications and recording effects and is produced by an imperfect sensor whose location and look direction are usually imperfectly known as a function of time.

This document presents the results of the definition study carried out in the research category of Mathematical Pattern Recognition and Image Analysis.

A Working Group composed of the following individuals was formed to carry out this study leading to the definition of fundamental research issues in this category.

MATHEMATICAL PATTERN RECOGNITION AND IMAGE ANALYSIS

WORKING GROUP

Chairman:

Dr. L. F. Guseman, Jr., Associate Professor
Department of Mathematics
Texas A&M University
College Station, Texas 77843
(713) 845-6931

Members:

Dr. Jack Bryant, Associate Professor
Department of Mathematics
Texas A&M University
College Station, Texas 77843
(713) 845-6334

Dr. William A. Coberly,
Associate Professor & Chairman
Division of Mathematical Sciences
University of Tulsa
Tulsa, Oklahoma 74104
(918) 692-6000, Ext. 228

Dr. Henry P. Decell, Jr., Professor
Department of Mathematics
University of Houston
Houston, Texas 77004
(713) 749-2126

Dr. Richard P. Heydorn
Earth Observation Division
NASA/Johnson Space Center (SF3)
Houston, Texas 77058
(713) 483-4763

Mr. R. B. MacDonald, Chief Scientist
for Earth Resources
NASA/Johnson Space Center (SA)
Houston, Texas 77058
(713) 483-5305

Dr. Edward M. Mikhail, Professor
School of Civil Engineering
Purdue University
West Lafayette, Indiana 47907
(317) 494-1475

Dr. George Nagy, Professor & Chairman
Department of Computer Science
University of Nebraska
Lincoln, Nebraska 68588
(402) 472-3200

Dr. W. B. Smith, Professor & Director
Institute of Statistics
Texas A&M University
College Station, Texas 77843
(713) 845-3141

Dr. Alan Strahler, Assistant Professor
Department of Geography
University of California
Santa Barbara, California 93106
(805) 961-3772

Dr. Philip H. Swain, Associate Professor
LARS/Purdue University
1220 Potter Drive
West Lafayette, Indiana 47906

For identifying research issues, the Working Group planned and conducted workshops in "Registration and Rectification of Remote Sensing Data," "Digital Image Modeling," and "Digital Image Pattern Recognition." Scientists from other universities, from research institutions, and from industrial and governmental organizations were invited to attend these workshops and to assist in identifying critical research topics in their areas of expertise. An agenda, a list of presentations, and a list of invited attendees for each workshop appear in the Appendix. A list of planning meetings, briefings, and documentation sessions held during the course of this study also appears in the Appendix.

The research issues identified during this study are presented in Section 2.0.

2.0 MATHEMATICAL PATTERN RECOGNITION AND IMAGE ANALYSIS

In this section we address fundamental research issues that arise in developing automated approaches to extracting information by remote sensing. The goal is to make effective and efficient use of the sensor output (digital image or images) in conjunction with ancillary data to determine the required attributes of the specific taxonomy comprising the object scene. For our purposes, taxonomy refers to the collection of classes of interest as defined by the particular application (e.g., vegetation types, rock types, etc.).

On a single calendar date, the electromagnetic radiation being reflected and/or emitted from the object scene can be sampled by an electro-optical-mechanical sensor mounted on an earth-orbiting satellite. The sensor records energy measurements at one or more polarization angles, in one or more spectral bands, over each resolution element (area of fixed size determined by the sensor) in the object scene. The sensor thus associates with each resolution element in the object scene a measurement vector (pixel) each of whose components is an energy measurement in one spectral band at one polarization angle. The vectors are arranged in an array which maintains the spatial relationships of resolution elements in the object scene; that is, adjacent vectors in the array represent adjacent resolution elements. The sensor may, of course, introduce some spatial distortion.

The resulting array of measurement vectors is a multi-dimensional digital image of the object scene on a particular date. The measurements

take into account aberrations due to atmospheric effects. A single sensor may produce many digital images of the same object scene on different calendar dates (that is, temporally), and a combination of sensors with differing resolutions can produce many digital images of the same object scene on the same or on different calendar dates. Hence, each object scene can be represented by a digital multi-image, that is, by a set of digital images produced by a combination of sensors on different calendar dates.

Ancillary spatial data, possibly in the form of digital images, such as maps, aerial photographs, and weather data at various locations, may also be available for integration into the digital multi-image. Ancillary calibration data, which are not necessarily related spatially nor applicable to the whole digital multi-image, may also be available. Training sets or reference signatures are examples; a set of meteorological data provided for calibrating the entire area (as in yield prediction) is another example.

The problem is how to make efficient use of the remotely sensed digital multi-images and ancillary data to infer the identity of the classes comprising the taxonomy of an object scene, or the proportion and/or location of those classes in the object scene and the attributes of each class.

The ability to make inferences about taxonomic classes in an object scene requires that one understand the intrinsic properties of the digital multi-image and subsequently can apply this understanding to establish relationships between the object scene classes and attributes and their

distributions of measurement vectors. These vectors may be recorded at different times and under differing conditions in the atmosphere and in the source of illumination by sensor(s) with different recording properties (signal-to-noise ratio, gain, bias, etc.).

The first steps to understanding the complicated phenomena of the remote sensing process involve the acquisition, presentation, and statistical treatment of the data (digital images). This statistical treatment may involve only simple computations such as histogramming spectral values from a single component of the measurement vectors in the digital image. However, it may also involve sophisticated mathematical ideas and complicated experimental designs. Once a sufficient number of digital images has been analyzed, mathematical systems are constructed which attempt to account for the results of the analyses. These systems usually are called mathematical models. Given an observed set of measurement vectors and a model of the relationship of the measurements to object scene classes and attributes, inferences can be made about the composition of the object scene.

Various fundamental problems are encountered while attempting to develop automated techniques for applications of remote sensing. Many of these problems fall into the category "Mathematical Pattern Recognition and Image Analysis."

RESEARCH SUBCATEGORIES

From the workshop presentations and discussions, and from subsequent meetings of the Working Group, a number of research issues were identified and, for presentation, grouped into the following subcategories:

- 2.1 Preprocessing
- 2.2 Digital Image Representation
- 2.3 Object Scene Inference
- 2.4 Computational Structures
- 2.5 Continuing Studies

Related issues within each subcategory were again grouped in to areas and, in some cases, subareas. A priority of I, II, or III was assigned to each research issue to indicate either the issue's importance or its dependence on prior investigations.

The ordering of the first three subcategories is intentional; it represents the steps usually performed in carrying out an approach to a given problem--readying the data, developing the model, and implementing the model. Each of these subcategories both influences and is influenced by the others. For example, modeling of a digital image for a given application is clearly affected by the approach used to register the data (preprocessing). Also, the model clearly influences the choice and implementation of approaches formulated for making inferences about the object scene. The model developed also dictates which digital images are registered. Implementing the techniques is dependent on both methods of data storage and limitations imposed by existing computer architectures. These issues are discussed in 2.4.

Additional topics, which were deemed important by the Working Group, but not adequately addressed in the study definition, are discussed in 2.5.

2.1 Preprocessing

By preprocessing we mean those operations or transformations applied to the original digital image(s) which involve correcting, compressing, or combining images to reduce the magnitude of unwanted effects or otherwise to improve the quality of the digital image data for subsequent processing and modeling. This preprocessing can be geometric, where the spatial structure in the image is required for processing. Registration is an example of this type of processing. By radiometric, we mean that processing in which the radiance measurements from the ground are corrected. Here the spatial location of the pixels in the image may or may not be important. A sun angle correction in which the points in the image are adjusted to values representing radiances at a given sun angle is an example of this type of processing.

Registration is the operation by which a digital image is mapped onto another equivalent digital image using transformations of a specified form. In the work addressed here, two digital images are equivalent when they represent the same segment of the earth's surface. While the main task is to determine specific parameters of the transformation desired, it is equally important to evaluate the accuracy with which registration has been accomplished. This evaluation includes first establishing criteria for accuracy, then implementing and testing procedures based on these criteria.

While registration merges two or more digital images to form one reference array, which may not have geographic significance, rectification is the procedure that provides such significance. It is the operation which establishes the appropriate correspondence between a digital image and

the segment of the earth's surface characterized by the image. It is important to recognize that while the digital image is two-dimensional, the corresponding portion of the earth's surface is essentially three-dimensional. Hence, in rectifying digital images, careful consideration must be given to the role played by elevations (the third dimension) and to map projection. As in registration, there is a need to establish criteria on which to base accuracy measures (measures of performance) for rectification procedures. In addition, methods need to be formulated for applying the accuracy measures.

Problems involving registration and rectification should be addressed for remote sensing data acquired from both aircraft and spacecraft. It is easy to recognize the potentially significant differences in the characteristics of these two types of data. One such difference is related to modeling the sensor/platform system. Consideration should be given to treating two digital images acquired either from the same platform or from two different platforms.

Research Issues By Areas--Preprocessing

2.1.1 Geometric Preprocessing

Many of the issues related to the geometry of the remote sensing data are enumerated in the following sections.

Reference Coordinate System. Points on the earth's surface may be located on several types of coordinate systems. At present, the systems used include: geographic systems which designate a point location by latitude ϕ , longitude λ , and height, h ; Universal Transverse Mercator (UTM), which designates points by Easting E , Northing N , and height h ; a local space cartesian system, which identifies coordinates as X , Y , and Z ; and a variety of other map projection coordinate systems (McEwen, 1979). Most of these systems have been developed for representing the earth's surface without particular applications to aerospace remote sensing. Some systems (Colvocaresses, 1974; Synder, 1978) have been proposed for this purpose, but there has been no evaluation of their universality. A basic question is whether or not a universal coordinate system (UCS) should be established for remote sensing data. This UCS, which could serve as a common reference system for data collected and processed from a number of different sensors and platforms, should be suitable for the largest number of applications. It would be necessary also to establish a dense set of points over the surface of the earth so that accurate registration would result in overlapping zones of image data. Other considerations include the relation between efficient computer data structures (rectangular arrays) and other systems (e.g., geographic), and transformation of large sets of data between

different systems and the resulting errors. It is estimated that a minimum of two years is needed for this research. (Priority II).

Control and Correspondence. In both registration and rectification, reference points, features, or areas are required to control the operation. Currently, reference points, or points of known coordinates in both image spaces in registration or in the image space and object or ground space in rectification, are used most commonly (Mikhail, 1979). Only very recently (Leberl, 1978) has an attempt been made to use more than single points. Therefore, the whole question of control still requires considerable investigation. The following issues should be considered.

1. A systematic classification of different types of control (points or nets, for example) and an indication of when to use which type should be investigated. Also, the problem of when to use relative control and when to use absolute control needs to be addressed. (Priority I).

2. The U. S. Geological Survey produces DEM (Digital Elevation Model) data and DLG (Digital Line Graph or digital planimetric) data. How can DEM best be used as a "multispectral" parameter in classification (e.g., to correct for sun illumination), and as a means for achieving accurate rectification? How can DLG be used either to create "control" for registration and rectification or to serve as ancillary data for the remotely sensed digital image. (Priority II).

3. What are the characteristics of optimum spatial control for merging sensor data sets of different geometries (e.g., MSS and radar)? (Priority III).

4. Investigate techniques of pattern recognition for determining control points and/or features which are easily identified and located without having to correlate the scene content with a reference image chip. There is also a need to define algorithms for these techniques, and to quantify features as a function of scene content, spectral band, resolution, spatial and frequency characteristics, etc. (Priority I).

5. An important research problem, particularly for agricultural applications, is to determine a good strategy for obtaining absolute and relative ground control for Landsat D-type imagery. This research would consider the problem of recognizing control features whose appearance may change significantly or even radically with the crop calendar, season, local meteorological conditions, etc. (Priority II).

6. Other important research should examine how to determine theoretically the maximum amount of control needed, beyond which only a negligible or insignificant improvement in accuracy would result. This naturally depends, at least to some extent, on the error modeling algorithms used. (Priority II).

7. Registration and rectification, although somewhat similar, can be quite different operations. Should the control, then, be the same for both operations? (Priority III).

The research problems concerning control are closely associated with correspondence, which can be between similar images obtained at the same time, dissimilar images obtained at the same time, images (both similar and dissimilar) obtained at different times, or images of various types and ancillary data of various kinds. It is important to investigate

different methods of selecting features for establishing such correspondences. Some of the research tasks and questions are discussed briefly below.

1. There is a need to investigate various existing techniques for establishing correspondences (spatial domain/frequency domain cross-correlation, etc.) and to determine circumstances under which particular techniques are optimal. Also, the effects of scale, sampling resolution, orientation, and other sensor variations on the accuracy of these techniques should be studied. (Priority I).

2. How can correspondence accurately be established between a digital image and digital terrain data such as DEM/DTM? (It is possible that an interactive approach may be the best, because it would allow cross-identification.) (Priority I).

3. The best means for establishing a precise correspondence between multitemporal data needs to be determined. (Priority I).

4. The "mosaic seam problem" still needs solving. This problem arises due to errors in registration between adjacent images, radiometric imbalances within images, and radiometric imbalances and differences in image content. Temporal differences compound the problem, and it is expected that only when there is a proper photogrammetric model will the problem be solved. (Priority II).

5. The effect of significant differences in spatial and spectral resolution on establishing a correspondence should be investigated. (Priority III). Since the research on control and correspondence is varied and is influenced by other research areas, it may require three to four years to complete. Of course, some results will become available earlier.

Resampling. As a result of instabilities in the air or in the spaceborne sensor platform, or as a result of geometric and radiometric distortions introduced by the sensor optics and electronics, there is no simple method for determining a transformation which approximates the one-to-one correspondence between the digital image scene radiance map and the object scene radiance map. Ancillary information such as ground control points, a priori knowledge, or sensor properties, can be used to estimate the geometric transformation linking measurement vectors in the digital image to locations in the object scene. A radiometric transformation, which will specify the radiometric values at the new locations identified by the geometric transformation, needs to be defined.

There are two alternative methods for removing geometric distortion. An investigator may perform all operations relating to object scene inference in the untransformed digital image and then remove geometric distortions, or he may perform geometric and radiometric transformations on the digital image and then make inferences about the object scene in the transformed digital image. Both alternatives involve resampling-- interpretation of radiometric values in the transformed scene, as in the second alternative, or interpolation of object scene attributes and classifications between pixel locations in the transformed scene, as described in the first alternative.

Several questions, discussed below, arise regarding the best approach to resampling using either the first or the second alternative and taking into consideration the effects of resampling.

1. Under what conditions and for what applications is the first alternative the most appropriate approach to resampling? The second alternative? (Priority II).

2. To what extent does resampling the radiometric data using the second method affect accuracy in classification and in other techniques used for making inferences about the object scene? (Priority I).

3. To what extent does resampling the processed image scene using the first alternative affect the accuracy of techniques for making inferences about the object scene? (Priority I).

4. What are the considerations for designing future sensor/platform combinations that reduce the errors introduced by geometric and radiometric distortions? (Priority II).

Registration-Rectification Sequence. Several questions arise when we consider two digital images of the same segment of the earth's surface. If each image is rectified to ground control, the two images should be registered. If it is possible to register one image to the other, then they should be rectified. There has been no clear-cut way to determine which sequence is optimal, mainly because "optimal" has not been defined clearly; this whole question therefore remains a basic research problem. Both theoretical and experimental work may be involved in investigating the sequential relationships between registration and rectification. In such a research effort it is important to recognize that one sequence may be more suitable for similar images while the reverse sequence may be more appropriate for dissimilar images. A fundamental approach considers extracting whatever information is needed first, then rectifying such information; in other words, we should attempt to look beyond the gray scale or radiometric domain and into a symbolic domain. In this case registration and rectification of symbolic features should be considered.

This task is sufficiently well-defined so that some results can be published after one year of research. However, a complete solution to the problem will require longer. (Priority I).

Errors, Tolerances, and Accuracy Measures. In geometric and other aspects of preprocessing (e.g., registration and rectification) accuracy is paramount. So far, the practice with respect to rectification (and to some extent registration) has been to calculate mean square errors at check points (Konecny, 1976; Mikhail, 1977). Other measures, such as maximum deviation, have been used, particularly in registration. Such procedures, it is well-known, are not necessarily the best ones. Important research problems in this area, described below, should be addressed.

1. Careful investigation is needed to precisely define distortions, errors, tolerances, and, in particular, defining accuracy measures for image data, reference data, and various operations. These definitions should address both geometric and radiometric processes. (Priority I).

2. The distinction between radiometric and geometric errors, if it exists, should be made. (Priority I).

3. Once precise definitions are established, procedures for evaluating errors in rectification should be determined and accuracy measures for registration and rectification must be established. (Priority I).

4. For various images, particularly those obtained by radar, improved procedures for correcting for both radiometric and geometric "errors," caused by relief, are required (such as those due to antenna pointing and to radar ranging geometry). (Priority II).

5. Processing quality should be more well-defined, perhaps theoretically. The interaction between processing and the quality of the data being processed, which is related to the need for standard synthetic and real images, is an important matter for investigation. (Priority III).

6. Accuracy can be regarded as either absolute or relative. One method for assessing absolute accuracy is by extensive control, which can be difficult. What are other alternatives? (In absolute accuracy, the accuracy of the control must also be considered). (Priority II).

7. Measures of relative accuracy that might support limited absolute accuracy can be devised. A lot of research could be done in devising reliability factors (as used by CDC, Panton, 1978), or ground control point correlation factors (as used by IBM), or correlation factors not involving ground control points (as at TRW), or figures of merit. (Priority II).

This research effort (issues 1-7 above) is expected to require two years, although some results can be published after one year. More extensive research, however, could continue for a total of three to four years.

Sensor/Platform Modeling. Although registration, and to a lesser extent rectification, can be accomplished using global warping functions (Anuta, 1973), these are not substitutes for accurate sensor/platform modeling. In rectification there has been some attempt at such modeling (Mikhail, 1977 and 1979; Panton, 1978; TRW); for registration some work has been done by TRW. It is important to note that far better results will be obtained from both rectification and registration once we know the behavior of both the sensor and the platform. Thus, intensive research is needed for the development of generally available models for different sensors and platforms. Other research tasks are outlined below.

1. An analysis should be made to determine the feasibility of on-board determination of sensor location and orientation relative to earth.

This requires a data base on board. (The NASA NEEDS system is relevant to this task). (Priority II).

2. Determine the feasibility of performing nearly real-time sensor modeling. (Priority III).

3. There is a need to investigate the use of star sensors for sensor modeling in order to achieve sub-pixel accuracy. (Priority III).

4. An accurate parametric sensor model (for both internal and external geometry) for Landsat type images as an alternative to existing global rectification models needs to be developed. (Priority I).

5. Research is needed to determine the possibility and the advantage of performing nearly real-time rectification on board the platform as opposed to on the ground. (Priority III).

6. How well can recursive techniques, such as Kalman filtering, be used for orbital modeling of errors in attitude? (Priority I).

7. The need for introducing accurate reference marks (such as time spikes or angle marks) in the image to help in effective sensor calibration and modeling should be analyzed and documented. (Priority I).

8. As the resolution improves, sensor modeling should, in principle, work better; this needs to be ascertained through controlled investigation. (Priority II).

9. Whether or not a type of sensor should be matched to a particular application, and which sensors are amenable to which registration/rectification techniques, needs to be investigated. (Priority II).

Because this research (issues 1-9) is extensive and the variety of sensing systems large, it can require five years. As some sensors are modeled, results can be published.

Topographic Problems. The extensive work in photogrammetry dealing with stereo data naturally raises a variety of questions when one considers applications to remote sensing. Therefore, whenever overlapping remote sensing data with significant Base/Height ratios exist, rigorous photogrammetric techniques should be applied. (Base/Height ratio measures the distance between the location of the platform for two images and the height of the platform above the terrain.)

Only very recently (Mikhail, 1979) has there been an attempt to reduce photogrammetrically sidelapping MSS data (from aircraft). More research is needed to investigate the various problems arising from adapting current photogrammetric techniques, and perhaps to develop new ones, for use with overlapping remote sensing data. Some areas for such research are briefly discussed below.

1. In regard to stereo imagery, research is needed to evaluate the accuracy of recovering point evaluations and the impact on registration and rectification of using such elevations. (Priority I).

2. The concept and use of "orthophotos," or a set of images equivalent to a map where effects of relief and tilt have both been eliminated, needs to be critically examined. Different sensors produce characteristically different images which may or may not be suitable for producing orthophotos. In fact, there seems to be a slow shift away from orthophoto production. What alternative products from remote sensing data are useful and suitable? (Priority I).

This research on topographic problems could yield results after one year, but may extend to two, three, or even four years, depending on the range of sensors considered.

2.1.2 Radiometric Preprocessing

The term preprocessing refers to computations made to remove unwanted (distortions" or "noise") elements from a set of measurements, and is a step generally taken prior to other steps in processing, such as classification. Radiometric, as opposed to geometric, preprocessing refers to processing on a given pixel which ignores the fact that the pixel has certain properties relative to its immediate neighbors. The so-called sun angle correction made on Landsat imagery is an example of radiometric preprocessing. Here the radiance in all channels for each pixel is adjusted to correspond to what would have been the radiance if the angle of the sun from the zenith corresponded to some given angle.

A limited amount of research has been done on models, independent of specific applications, which describes or predicts physical characteristics or distortions of scenes. Lambeck and Potter (1978) consider procedures for the correction of spectral signatures for the effects of atmospheric haze. Here a physical model describes how radiance values are affected by atmospheric scattering and absorption, and this model is used to correct radiance values in the digital image for atmospheric effects. This type of general correction is universally applicable. Other research might consider models which correct for differential illumination due to topographic effects, and models which remove sensor effects.

The development of models for correcting haze, illumination, and the sensor itself is an objective of research being examined by the Scene Radiation and Atmospheric Effects Characterization and Electromagnetic Measurements and Data Handling working groups. Their emphasis is on the

physical modeling of the interaction between fundamental electromagnetic phenomena and surface matter. As they relate to image analysis, these models would be used to develop transformations for correcting the digital image. While this approach is the one which is likely to produce the desired solutions, there may be other solutions.

Some solutions have produced simple, empirically calibrated models. Lambeck and Potter (1978) have developed a procedure which corrects for haze in the greenness-brightness plane. This procedure appears useful for multi-temporal analysis. Correction of differential illumination using terrain models registered to spectral multi-images has been explored by several researchers. Woodham's approach at the University of British Columbia has assumed Lambertian scattering; Sadowski and Malila at ERIM have used the Suits bidirectional reflectance model for correcting the differential illumination of a forest image.

Research Issues--Radiometric Preprocessing

Some research issues, including an estimate of the length of time required to make significant progress, are discussed below. An attempt is made to rank the issue in terms of its importance.

1. Certain distortions occur in satellite observations because they are made through the earth's atmosphere. With Landsat, this distortion is approximately affine. As haze levels increase, contrasts decrease. This generally means that the discrimination of object types is not absolute unless a haze correction is made. That is, if a classifier is trained on haze-free data and haze is added, the spectral distortion will cause errors in classification. Therefore, methods to correct digital images for

distortions resulting from haze need to be developed. As mentioned earlier, some work has been done on correcting haze which makes use of targets of known reflectance and/or shifts in the data along certain transformed coordinates; these approaches make use of data only from the primary data sources (Lambeck and Potter, 1978). More research along similar lines is needed to determine the extent to which corrections of this type are possible. Using Landsat data, significant progress could conceivably be made in 2-3 years. (Priority I).

2. Discrimination among vegetation types, particularly in agricultural crops, depends largely upon the rate at which the vegetation covers the soil. Thus, it is often desirable to adjust the data for differences in soil color in visible bands and for differences in temperature in thermal bands. Research is needed to develop methods for making these corrections.

Developing these methods may require data other than that which can be obtained from the primary data source. For example, soil color can change depending upon the soil moisture, so observations about current precipitation may be needed to predict soil color. Therefore, to predict soil color and temperature backgrounds, models driven by satellite-derivable point measurements may be needed. Satisfactory solutions may require 5-8 years of research.

Methods which depend essentially upon contrasts (ERIM's greenness coordinate, Kauth et al.) or some other such methods of transformation may suffice as more sophisticated pattern recognition or sensor systems are developed. Because this form of preprocessing may depend upon progress

in other areas, this task is assigned a ranking of Priority III.

3. Terrain relief can introduce noise into the spectral signatures of vegetation targets. For example, shadowing effects in high slope areas can distort signatures. Methods are needed to correct or to minimize such effects.

Significant progress in this area may be possible using digital elevation models which have already been developed. Some results, therefore, may appear in 2-3 years. Further progress will probably depend upon the development of good canopy reflectance models. This research may take 6-8 years to develop. While this is an important area for research, as with the preprocessing for soil background effect, it may not produce significant change once more sophisticated sensors or methods of pattern recognition (discrimination or proportion estimation) are developed. This effort is consequently given a ranking of Priority III.

4. Changes in view angle can cause changes in the spectral appearance of a target. Field experiments performed at Purdue (Vanderbilt, 1980) show that a canopy reflective response is a pronounced function of illumination angle, scanner view angle, and wavelength. Since oblique viewing sensors such as the Multispectral Resource Sampler (MRS) have been proposed, this may be a significant source of signal variation in the future. In fact, even with Landsat view angle effects are noticeable in data acquired in the overlapping portions of the ground tracks. Methods should be developed to correct data in order to remove or minimize differences in the viewing angle. Since a minimal amount of research has been devoted to this problem, it is estimated that 4-7 years would be required before good

solutions would emerge. Moreover, "different" sensors would present different problems. Consequently, this problem may be one that would follow the sequence of sensor development:

The problem is considered serious with future sensors much more than with Landsat, and it is thus given a Priority III ranking.

2.2 Digital Image Representation

Digital image representation is the determination and modeling of basic characteristics or features of the digital image which can be incorporated into the process of identifying classes and attributes in object scenes. Implicit in scene representation is determining the extent to which the information content of the digital image can be used to identify those basic characteristics which are useful for various applications. This is especially important in identifying those characteristics of real classes and attributes in the object scene which are also represented in the digital image.

The term "digital image modeling" should be distinguished from the terms "scene modeling" and "sensor modeling." These modeling efforts are not a part of the issues described in this section. However, the efforts at digital image representation will require some understanding of and some input from the efforts in Scene Radiation and Atmospheric Effects Characterization and Electromagnetic Measurements and Data Handling.

2.2.1 Spatial Representation

The intrinsic geometric relationships of pixels within an image require that a pixel must be interpreted in the context of its spatial

neighbors. In this section three areas of research for aiding the spatial understanding of digital images are identified.

Texture. Image texture is a concept which has given rise to numerous descriptions. There has been no general agreement on a formal definition of texture, either from a psychological or a mathematical point of view. There is, however, general agreement that texture is important in understanding digital images (Haralick, 1979).

Research Issues--Texture

1. Most of the work on texture has been done using images of much finer spatial resolutions, relative to the size of the objects, than those of the sensors under consideration in this effort. Moreover, little work has been done on texture in multiple images such as those generated by multispectral sensors. Therefore, basic research to define texture for such images and applications is required. (Priority I).

2. The atmosphere and the sensor system introduce spatial correlation into the digital image array. Transfer functions need to be determined for new and existing sensor systems, and a study is needed to incorporate the spatial correlation into the digital image model (Tubbs and Coberly, 1978). This issue requires input from the research in scene radiation and sensor modeling. (Priority II).

Spatial Scene Segmentation. In most applications, multi-pixel spatial structures (fields in agricultural applications, for example) are important components of the digital image model. Automatic spatial segmentation (delineation of the spatial structures) of multi-images is an important step for future development of procedures for image classification and analysis. (A recent survey article on this topic is

Rosenfeld and Davis, 1979.) Automatic delineation of agricultural fields has been addressed by Bryant (1979) and ERIM, and general approaches have been explored by Haralick (1980) and Haralick and Watson (1980).

Research Issues--Spatial Scene Segmentation

Most of the research on segmentation has been for single images. There is a need to determine the important spatial structures, especially for agricultural fields, and to develop the capability for automatic multi-image spatial segmentation. (Priority I).

Mixed Pixel Models. The basic element of the digital image is a pixel. We assume that each pixel is associated, possibly in a complex way, with a point or area in the object scene. We define two types of pixels: pure and mixed. If the related area in the scene consists of material from only one taxonomic class, then the pixel is said to be pure. If more than one class is present, then the pixel is said to be mixed. Of course, one pixel from a digital image might be pure under the taxonomy dictated by one application and mixed under another.

Research Issues--Mixed Pixel Models

The coarse resolution of multispectral scanners makes the mixed pixel an obstacle in accurately classifying and estimating acreage in most applications. Automatic recognition of mixed pixels and their treatment in procedures for classification and aggregation will require a better understanding of this digital image phenomenon. (Priority I).

2.2.2 Spectral Representation

Spectral representation refers to quantization of spectral image data using mathematical models. For example, one might model certain spectral image data as a statistical mixture of spectral classes corresponding to crop types. Kanal (1974) and the Proceedings of the LACIE Symposium both survey current approaches.

There are basically two general approaches, parametric and non-parametric. They differ and are appropriate in proportion to what one can assume to be true about the spectral image data. The purpose of either approach is to provide a framework (model) for subsequent use in classification, proportion estimation, etc.

Parametric statistical models attempt to describe spectral images in terms of a finite set of parameters. Initially, one selects an appropriate family of (parameterized) density functions and accepts the postulate that an appropriate choice of the parameters will provide a suitable statistical representation of the digital image. Subsequently, one actually estimates the parameters involved.

Non-parametric statistical models are those that do not depend upon preselecting a parameter-dependent family of density functions (density estimation, stochastic approximations, clustering, etc.).

Closely related to these approaches is dimension reduction, the process of transforming digital images (single or multiple) having spectral measurement vectors of dimension n , to a transformed digital image whose spectral measurement vectors are of dimension k , where $k < n$. It includes feature extraction, feature subset selection, and

linear and non-linear combinations. It is an essential part of multivariate parametric and non-parametric spectral representation from the point of view of data structure simplification, computational economy, and data displays. Summaries of current approaches can be found in Decell and Guseman (1979).

Research Issues--Spectral Representation

Parametric and non-parametric statistical models for typical multivariate spectral image data need to be developed. (Priority I).

The criteria for selecting the most appropriate transformation from among a prescribed class of transformations are usually based upon preserving the statistical information in the original digital image, preserving information which is other than statistical (spatial or textural information, for example), or both. While it is not always possible to determine a dimension-reducing transformation which preserves all of the information contained in the original digital image, it is important to detect when such transformations do exist. It is equally important to be able to determine (measure) the loss of information when the transformation does not exist. Only by examining our description of information content (class separability) can we determine the acceptability of a dimension-reducing transformation which does not preserve the original digital image, but which may nearly do so.

1. Techniques for reducing dimension (linear and non-linear) for parametric and non-parametric statistical models of typical multivariate spectral image data based upon the preservation of all (or nearly all) statistical information should be developed. (Priority I).

2. Methods for reducing dimension (linear and non-linear) for typical multivariate spectral image data based upon the preservation of all (or nearly all) data structure information which is other than statistical need to be investigated. (Priority II).

2.2.3 Temporal Variation

Acquiring digital images of an object scene on different calendar dates makes it possible to study the object scene classes in terms of their temporal variation.

The LACIE pointed out that the use of multi-temporal digital images is critical in discriminating between classes which are separable only at certain times during the growing season. Most of the work involving the use of temporal variation in remote sensing applications can be found in various reports presented at the LACIE Symposium (1978).

Research Issues--Temporal Variation

The general problem is to develop adequate models of temporal variation which are best suited for different remote sensing applications. Introducing several time-dependent digital images usually requires the application of procedures for registration/rectification, and subsequent temporal models could be quite complicated. The following specific issues should be considered.

1. Temporal variation is not restricted only to taxonomic classes; models for temporal variation in digital images need to be developed, as well. The model should attempt to distinguish useful temporal variation (crop phenology, etc.) from irrelevant temporal variation (sun angle, haze, moisture, etc.). (Priority I).

2. Approaches need to be formulated for detecting and quantifying changes in the content and configurations of object scenes. In particular, models for change need to be developed. (Priority II).

3. The sensitivity of temporal models to errors in registration/rectification needs to be investigated. (Priority II).

4. In the past, temporal models have required accurate image-to-image registration. There is a need to explore the possibility of developing temporal models which require less precise registration or which bypass the registration process. (Priority II).

2.2.4 Syntactic Modeling

By syntactic modeling, we mean constructing models that specify the spatial, spectral, or temporal constraints or characteristics of the objects in the object scene and using such models in pattern recognition information extraction. (The term syntax is borrowed from linguistic analysis, where words only have meaning if interpreted in context.) As a simple example, consider a spectral classifier which uses ancillary slope data to identify water only from flat locations. Spatial relationships may be exploited, as in the example of shape recognition (airplanes, tanks) or adjacency models (a beach is adjacent to both land and water). Temporal relations may be important as well, as in recognizing the sequence of land use change: forest→bare ground→asphalt and not recognizing asphalt→forest or asphalt→bare ground. Identifying crops using multi-temporal models of changes of signatures through time is another example.

The application of syntactic models has not progressed much beyond the simple examples cited above. However, theoretical work within the discipline of pattern recognition is an active area, especially in the work of Pavlidis (1972, 1977), Haralick (1977), and Fu (1974), Moayer and Fu (1976). Pavlidis's research emphasizes the construction of connected graphs describing image structure and their relationships with semantic graphs depicting the syntactic restraints, whereas Haralick has approached the problem by increasing the probability of correct object identification, given a finite set of possible object structures or relationships.

Research Issues--Syntactic Modeling

The primary research problem is to select and develop approaches to syntactic modeling which are best suited to remote sensing. Perhaps Pavlidis's approach using graph theory will be most fruitful since it is closely related to research on segmenting images and on data structures and storage for remotely sensed data. For applications to renewable resources, however, this research should emphasize appropriate types of syntactic models. Since many remotely sensed data are multidimensional, the specification of multidimensional syntactic structures is important in advanced research. Because research in pattern recognition is supporting developments in this area, NASA should support limited studies at a Priority II level.

2.2.5 Ancillary Data

Remotely sensed data consist of sets of measurements of electromagnetic radiation at points on the earth's surface. However, in many cases the objects in the digital image are complex and may not always be separable on the basis of electromagnetic radiation alone. In such cases, the use of ancillary spatial data, shown by additional registered layers in a multi-image, may be incorporated into the algorithm for extracting information to improve object recognition. As the use of remotely sensed imagery, especially from satellite platforms, becomes more widespread, more and more digital images will be incorporated into geo-based information systems. These systems, then, will combine not only temporal files of spectral data, but also image layers of ancillary spatial data. Thus, the demand for algorithms which use ancillary as well as spectral data to extract information will increase.

Spectral data are usually recorded as continuous measurements, or at least as continuous measurements which have been quantized into a reasonably large number of integral values; however, ancillary data may be continuous, stepwise or discrete, or categorical in nature. Thus, to exploit ancillary information one must combine disparate data types in a common framework for extracting information. Some progress has already been made in this area. In the past few years, Strahler (1980; Strahler et al., 1980) has demonstrated several mechanisms for combining spectral and ancillary data in a single classification procedure. These methods range from using probabilities to combine continuous spectral and categorical ancillary data, to using the logit classifier, which incorporates all types of data in a single step in classification.

Nonspatial ancillary data, which can be used in calibration or modeling, can also be used to aid in the process of extracting information. The automated use of crop calendars and yield models in agriculture, similar to those developed for the LACIE program, are examples. Again, the need here is for algorithms which merge spectral with ancillary data.

Research Issues--Ancillary Data

1. Research, development, and testing of both categorical and continuous models which combine remotely sensed and ancillary data should continue. Much progress has already been made in this area, and only one to two years should be necessary to produce significant results with refereed publications in two to three years. Because this research is needed for applications to geo-based information systems, this task should be supported at Priority I.

2. Fundamental research into advanced models, algorithms, and procedures which directly utilize both remotely sensed and ancillary data and their spatial and temporal variations, is also needed in the process of extracting information. An example is a procedure which exploits ancillary information to segment multi-temporal images. The models used here are specific to remote sensing rather than for general purposes, distinguishing this issue from 1. above. Since this research is a suitable follow-up to that described in 1. above, funding should be at the Priority II or III level depending on the time schedule for 1. Results should be achieved and published in two to three years.

2.3 Object Scene Inference

Here we address incorporating digital image representations into systematic methods for inferring the attributes of object scenes. For a specific application, this generally involves two phases: (1) determining the values of the model parameters in order to particularize the general model to the object scene at hand, and (2) performing the calculations which, based on the model, will yield the quantitative or qualitative information desired about the object scene. Mapping, inventory, and monitoring of natural resources are the primary objectives of inference. Mapping shows the location of classes, objects, items, or types of interest; it includes both hardcopy and display. Inventory is concerned with the counting, aggregation, census, or planimetry of scene objects without explicitly retaining spatial coordinate information. Monitoring refers to detecting change, discovering unusual conditions, and other operations of limited spatial and temporal scope.

Included in the process of inference are classification, categorization, identification, recognition, clustering, partitioning, taxonomy, and segmentation. We will be concerned with supervised and unsupervised learning, teaching, or training, with estimating parameters, distributions, and error rates, with assigning identities, labels, or symbols by either automatic or interactive means, and with evaluating the accuracy, dependability, and robustness of the entire process. Of particular interest is the role of the human and of the ancillary data, including those which have their source in images and those which do not. Techniques based on statistical as well as on structural, syntactic, relational, and other

deterministic approaches are germane. We are concerned with algorithms for multisource data, including multisensor observations, multitemporal observations, and combinations of multiimage data and non-image data.

In contrast to map displays or statistical inventory which forms the final product of the recognition process and benefits the "end user," data displays are intermediate products intended to improve the recognition process itself; they provide the opportunity for human interaction. The scope of the displays may range from simple histograms, which allow the users to judge the overlap between statistical distributions, to digital images which provide the means for assigning labels by photointerpreters.

2.3.1 Image Partitioning

Image partitioning is the process of delineating subsets of pixels of a digital image, where pixels belonging to the same subset possess similar characteristics and those in different subsets possess dissimilar characteristics. The definition of similarity depends in a complex way upon the taxonomy of the object scene, the required attributes, and the digital images and ancillary data available. In general, similarity is defined in terms of both the measurement values of the pixels and the relative location of the pixels within the digital image. One example of image partitioning is assigning object scene class labels to pixels in the digital image. Another example is identifying those pixels in the digital image which closely resemble (spectrally) their four nearest neighbors (spatially) to the North, South, East and West.

Clustering. One method of image partitioning is clustering, or forming subsets of similar objects. Much of the research activity in developing techniques of pattern recognition for remote sensing has been in

the area of discriminant analysis (or classification), or the problem of making new observations about known groups. A more difficult and perhaps more important area for research lies in developing methods of clustering for discovering the groups in the first place.

Methods for clustering used in the LACIE made use of only the spectral aspects of the digital images. Toward the end of the LACIE several clustering algorithms (AMOEBAS, ECHO, BLOBS) were developed which incorporated the use of spatial information. These algorithms are currently being considered for applications to remote sensing in agriculture.

Research Issues--Clustering

1. One of the critical research questions is how to evaluate clustering algorithms. The work of Fisher and Van Ness (1971) provides a general framework for comparing clustering algorithms. They test whether or not a particular algorithm produces clusters satisfying certain "conditions" for every possible data set. Admissibility criteria need to be defined which will assist in selecting appropriate clustering algorithms for use in applications of remote sensing. (Priority I).
2. There is a need to define performance measures for clustering algorithms in particular applications. (Priority I).
3. We need to develop appropriate models for pure and mixed pixels to use in spatially-oriented clustering algorithms. (Priority I).
4. We should determine how other characteristics of digital images (texture, for example) might be used to develop new clustering algorithms. (Priority I).

Classification. By classification we mean the process of assigning to a pixel a label, corresponding to an information class. The label identifies the pixel or otherwise describes its attributes, which are inferred from the multi-image data available for that pixel. The process of inference may consist of a combination of arithmetic and logical computations.

A classification method is considered effective if it is computationally feasible and produces reliably accurate results. In general, accurate results are achievable if:

(i) the multi-image data contain information sufficient for characterizing the information classes of interest, and this information is preserved by the processes of image representation used to form the data base.

(ii) an effective training procedure has been devised. A training procedure is a sequence of operations used to partition the multi-dimensional feature space defined by multi-image into disjoint regions having a one-to-one correspondence with the information classes (labels) of interest. The training procedure is effective if it can be readily carried out by machine and/or by a human analyst, and if it reliably produces feature space partitioning that results in accurate classifications.

(iii) an effective decision rule for deciding to which region of the partitioned feature space an "unknown" pixel should be assigned is available.

The processes in image representation have been discussed at length earlier. Here it simply will be reemphasized that a great variety of characteristic features may be extracted from the multi-image. The features may convey spatial information such as shape, size, or texture. They may convey temporal or topographic information (Fleming et al., 1979), or information about syntactic or structural relationships among scene components. The variety of features is limited only by the practical size and complexity of the data base and by the ingenuity and success of the researchers concerned with the discernment and representation of digital image characteristics. As more and different forms of data become available for use in conjunction with remote sensing data, continued research is required to better understand how these various forms of data interact in meaningful and informative ways.

The progress in developing effective training procedures involving spectral and temporal multi-image data was greatly advanced by research and development in conjunction with the LACIE. The most appropriate methods for partitioning the feature space into regions corresponding to the information classes are dictated largely by the amount and quality of ground truth information available. Under the stringent conditions of the LACIE, the problem of training with very limited ground truth (even no current verifiable ground observations) was explored. The results demonstrated that it is indeed possible to extract useful information from the data. The results also indicated that the role of the human analyst in the process is crucial and that the less certain the supporting data, the less reliable the results from the analysis of the data will be. An

important related problem is characterizing mixed pixels from scene areas of fixed but unknown combinations of cover types. There is still considerable potential for improving the overall process by developing improved unsupervised methods of partitioning (clustering) and improved interaction between data and analyst and between analyst and machine.

To date, rules for classifying object scenes, once training has been completed, largely have been limited to very straightforward statistical decision rules based on simple parametric assumptions concerning the features used. The most familiar is the Gaussian maximum likelihood rule and its variations, as used in the LACIE. More effective techniques are available, but they are more difficult to carry out (Kettig and Landgrebe, 1976), and most of them rely on simple parametric assumptions about the features. As more complex data bases involving diverse forms of data from widely disparate sources and of greatly varying quality become available, the familiar decision rules and classification procedures become outmoded. More general techniques are required for decision-making under such circumstances.

Training Procedures. To be most effective, training procedures should take into account the limited availability of concurrent ground observations and the availability of ancillary information, and should consider the most effective role in the training process for the human data analyst. Specifically, research in this area should aim to:

1. Develop techniques which efficiently use numerous sources of data, account for variability in both the information content and reliability of the data, and tolerate conflicting and missing data.

(Priority I).

2. Develop effective methods for displaying high dimensional data for evaluation and use by data analysts. (Priority I).

3. Investigate applying techniques of artificial intelligence in exploring subjective reasoning processes. This approach may be useful for determining how human analysts integrate diverse sources of information. It has been used previously for medical diagnosis and geological image interpretation. (Priority III).

4. Develop capabilities for learning pattern grammars which describe scene characteristics of renewable resources. (Priority III).

Decision Rules. The complexity of the decision rule used for classification is related to the logical and statistical complexity of the data base. There is a pressing need to develop more flexible and more powerful decision-making schemata. Specifically, research is needed to:

1. Develop effective procedures for making decisions which do not depend on restrictive parametric assumptions concerning the interactions of diverse data sources. (Priority I).

2. Develop multistage procedures for making decisions which, by successively using more sources of information, can produce increasingly refined classifications. Thus, for example, multi-temporal procedures should be able quantitatively to use past results with new data to produce an up-to-date, more detailed, and more reliable classification. (Priority II).

3. Investigate formulating generalized discriminant functions which can appropriately weight data features according to their relative

reliability and their information content. The results of using these discriminant functions should include an indication of the reliability of the classification produced. (Priority I).

4. Develop decision rules which tolerate missing data. (Priority II).

5. Investigate techniques of syntactic pattern recognition which use local and global structural features to identify significant scene attributes for applications to renewable resources. (Priority III).

2.3.2 Proportion Estimation

Proportion estimation is determining the fraction of the total acreage in a given area which contains material of interest. For example, of the total acreage in an area of 5 by 6 nautical miles, consider the problem of estimating accurately the proportion which will be harvested as winter wheat. Much of the research toward the application of satellite (MSS) data to proportion estimation has been sponsored by NASA and USDA. Summaries of this research are given in Heydorn et al. (1978), Feiveson (1978), and Hanuschak et al.

Approaches that have been taken can be categorized as follows:

a. Enumeration of classifications. In this approach an entire area is classified and the proportion of the pixels in a given class is the proportion estimate for that class. A variation on this method is one in which the area is randomly sampled and only the sample points averaged to obtain the estimate.

b. Stratified Areal Estimation. As with the procedure above, the area of interest is again classified. Here, however, the resulting

classification map (in a "classification map" each pixel is assigned to a given class and the area is thus partitioned into object classes) is treated as a stratification of the area. The proportion estimate is then obtained from a separate random sample using methods for stratified areal estimation. This approach is discussed in Heydorn et al. (1978) and Tanenbein (1970, 1971, 1972).

c. Regression Estimators. The approaches that have been tried are based on obtaining a linear regression of crop proportion estimates, derived from a sample survey, onto proportion estimates derived from classification. In a typical approach developed by USDA (Hanuschak et al.) a ground sample survey of crop acreages is taken using 1 x 1 mile primary sampling units. These units are also classified using Landsat data to derive a second proportion estimate for each sampling unit. The ground sample estimates are then regressed onto the classified estimates. A total area estimate is then obtained by first classifying the whole area and then projecting that number using the regression to obtain the final estimate. This method, as the stratified areal estimation described above, can reduce the variance of the estimator based on the ground sample alone.

d. Direct Estimators. Several methods (Feiveson, 1978) have been developed for estimating proportions directly (i.e., the methods do not depend upon an intermediate classification step). The following problem has been of interest recently. We are given a "mixture" density f , whose component densities are members of some parametric family \mathcal{F} so that f can be uniquely represented (some for positive integer M) as

$$f = \sum_{i=1}^M \lambda_i f_i, \quad (1)$$

where

$$\lambda_i \in [0,1] \text{ and } \sum_{i=1}^M \lambda_i = 1$$

$$f_i \in \mathcal{F}.$$

We want to estimate $M, \lambda_i, f_i, i = 1, 2, \dots, M$. Here the mixing proportions, λ_i , are taken to be the crop proportions in the area. Redner (1980) summarizes much of the theoretical work that has been done on estimation problems associated with this model. Lenington and Rassbach (1978) discuss an application that has been successfully applied to crop area estimation. Finally, Teicher (1961, 1963), Yakowitz (1968), and Goodman (1974) present the underlying theory of such a model--the identifiability of statistical distributions.

Research Issues

The research issues are listed below. After each statement of the issue a projection is given for the length of time required to obtain significant results and an attempt is made to rank the issue in terms of its importance.

1. For methods that depend upon a classification of the scene there is a need to develop improved classification methods which
 - a. require only a small number of training samples;
 - b. can deal with a large number of object classes;
 - c. deal with the fact that the samples to be classified come from a nonstationary object class distribution (i.e., the distributions change over geographical coordinates);

- d. can account for mixed pixels--a phenomenon that is a result of a sensor with finite resolution.

Judging from the research that has already been done with agricultural data derived from Landsat 1, 2, and 3, it would appear that, at least for crop discrimination, an improved sensor will be needed before the problem can be satisfactorily solved. It would appear, therefore, that in about five to ten years (depending upon the rate of development of satellite sensor systems) significantly better classification methods could be derived. Because of the dependence on sensor improvement and because of the fact that promising developments in deriving direct methods of proportion estimation, this issue is given a Priority III rating.

2. We need to formulate object class distribution models that could separate the predictable variables from the random variables and to specify a parametric family of laws of probability which will account for the random variables. Such a model would minimize the training sample requirements for supervised classifiers, aid in the development of clustering methods whose clusters have an explainable correspondence to object classes, and solve many of the estimation problems related to the mixture model approach. Work in this area is just beginning. It appears that approximately 2-3 more years is required before significant models in this area will evolve. It is, however, an effort that is basic to this general area. (Priority I).

3. In the mixture model approach there are two major problems. There is a need to derive estimators for the number of mixing distributions (an estimator for M in equation (1)). Second, there is a need to develop con-

sistent estimators of the mixing distributions. If the mixing distributions are known to have come from a given parameterized family, this requires obtaining consistent estimators for the parameters. In addition, small sample estimators of the bias and variance of these estimators are needed.

A satisfactory solution to this problem will probably depend upon the development of models as discussed in 2. above, and therefore it will probably be 3 or more years before satisfactory solutions appear, although significant progress could conceivably be made within the next 1 or 2 years. This effort seems the next logical step after developing object class models. (Priority II).

4. For methods that depend upon either clustering or estimating mixing distributions, there is generally the need to associate an object class name to a result (distribution or cluster). This is often called the labeling problem. Often the labeler makes mistakes in assigning an object name to a given set of pixels. Thus, there is a need to derive labeling methods which are reasonably robust to labeling errors. Again, labeling could benefit from a more effective distribution model, so approximately 2-3 years are needed before satisfactory solutions can be obtained. (Priority II).

5. The partial success of linear regression estimators that use estimates from machine classifiers and ground sample surveys suggests that more general regression or stratified areal estimators should be considered, at least when attempting to decrease the variance of an estimate from a ground sample survey. An estimator of this kind may work well with Landsat data. Moreover, when ground survey data is available,

satisfactory estimators could possibly be delivered in 1 or 2 years without a knowledge of the distribution model mentioned above. Immediate results would benefit domestic crop area surveys. (Priority I).

2.3 Error Models

An error model is a function or a family of functions that maps an estimate and its true value to a real number or a set of real numbers in order to measure the discrepancy between the two values. Typical measures of discrepancy are bias, mean square error, and the probability that the estimator will assume one value when the true value is different.

In remote sensing, error models have been used to evaluate the performance of estimators, to correct for bias, to determine sample allocations, and to reduce the variance of a given estimator. Some of the specific applications are discussed below.

a. Performance Evaluation. In inventory mapping, classifiers have been used to extend classification results from a small sample based on ground truth or an analyst interpreter to a large area. At least with Landsat data, experience indicates that a substantial classification error can result. The spectral similarity of confusion objects, object size (relative to the resolution of the sensor), the number of observations over time, and the number of training samples all can cause errors in classification. In cases where a classifier has been applied to each randomly allocated segment, the variance and bias due to classification error in the final estimates has been studied by Houston et al. (1978). Also, the performance of both machine and human classifiers applied to small areas (e.g., 5x6 n. mile

segments) (Wheeler et al., 1978; Chittineni, 1979, 1980) and on Landsat full frames (Bauer, 1977) has been studied. For the most part the error models considered in these studies have been elementary.

b. Bias correction. When an area is inventoried by counting the number of pixels classified into a given class, a bias can result if classification errors occur. This bias can be expressed in terms of the omission and commission error rates of the classifier and the proportion of the object class present in the scene. Attempts have been made to estimate these errors and to correct the results accordingly (Grey and Schucany, 1972). Quite often it is desirable to compute these errors with the same sample that was used to train the classifier. In this way, efficient use is made of the observations acquired from ground truth or by an analyst interpreter. Unfortunately this can result in biased error estimates unless special estimators are considered. The techniques related to this notion of "reusing" or "recycling" data have been called "jackknifing" (Gray and Schucany, 1972; Glick, 1978), "cross validation" (Stone, 1973), and "bootstrapping" (Efron, 1977).

c. Sample Allocation. As discussed in section 2.3.2 (Proportion Estimation), classifiers have been used to stratify an area into object strata. Errors in classification lead to impure strata. A priori knowledge of this impurity or current estimates of it can be used to allocate samples in an inventory survey and thereby increase the efficiency of the sample. A Neyman allocation, for example, would require a knowledge or an estimate of the proportion of the object in each stratum; this estimate is a measure of stratum impurity. Sequential methods based on Bayesian allocation are considered by Pore (1979). In this approach error models (in terms of updated mean square error estimates) are considered.

d. Variance Reduction. Inventory estimators using post-stratification have been discussed in the literature (Cochran, 1963; Fuller, 1966). Here no attempt is made to allocate samples in a special way (generally, a simple random sample is allocated to the union of all the strata) but the proportion estimate is made by averaging the class proportion estimates within each stratum across strata. When classifiers are used to generate the strata, classification errors determine the efficiency of the estimator. Error models which make use of cross validation have been studied by Myers and Wheeler (1979) in an attempt to design efficient (i.e., low variance) estimators. Many of the concepts discussed above related to "reusing" data apply here.

Research Issues

The research issues are described below. After each statement of the issue an estimation of the time required to obtain significant results is given. In addition, an attempt is made to rank each issue in terms of its importance.

1. Much of the research on the evaluation of classifier performance is based on empirical studies in which a given classifier has been tested on specific data. The general question, "how much information is in Landsat data" has not been addressed. An answer to a question of this kind could presumably be used to establish an upper bound for classification accuracy in a given application. Specifically, error models should be developed which could predict the performance of the classifier, perhaps in terms of omission and commission error rates, in a given region for a given application.

At least under suitable restrictions, such as using only point spectral values from Landsat (and not spatial relationships), it is estimated that significant results may be achieved in 1 to 2 years. The more general solution will probably depend upon developing better methods for "scene understanding" which are still being investigated in research in pattern recognition and artificial intelligence. Significant results should be available in 4-6 years. A good understanding of this issue would determine future sensor requirements, which in turn would greatly influence the course of pattern recognition research. (Priority I).

2. Even though it is sometimes known that certain factors influence classification errors, it is often difficult to assemble a data set in which each of these factors varies over a range of interest. Moreover, it is nearly impossible to find a data set in which only one factor at a time varies. Therefore, much more complete evaluative studies could be designed if real data could be simulated. To date little progress has been made in developing simulated data. Significant progress could be made in two years with 4-6 years required for realistic simulations to be developed. Since many programs now being considered by NASA involve foreign countries where no ground truth measurements are available, the use of simulated data may provide the only realistic tool for evaluation. (Priority I).

3. In inventory applications, an estimate of the classification error can be used more efficiently to allocate samples and to derive methods which can use a given sample for both this error estimation and for other functions related to estimation, such as classifier training. Some studies have been done using the so-called recycling methods as discussed above. While these

methods may reduce the bias of the estimate, they have a tendency to increase variance. Therefore, more research into the best ways of obtaining error estimates along with other estimates is needed.

A fair amount of research has already been done on problems of this type so good solutions may be proposed in a year or two. Current programs exist which could make immediate use of significant results in this area. (Priority I).

4. Many of the applications of remote sensing to mapping and inventory must be done without any ground truth sample for calibration or training purposes. One way to supply such information is to obtain it through manual image interpretation processes. However, because analyst interpreters are likely to make errors, methods are needed which are resistant to such errors. There has been some work done on this problem (Chittineni, 1979, 1980) but, for the most part the assumptions required in those methods do not always apply to real situations. Research is needed first to understand or model analyst errors and then to develop methods of automated pattern recognition which take advantage of the statistical properties of those errors.

One of the problems in modeling analyst errors is that interpretation procedures tend to be subjective and therefore inconsistent across analysts. While this effort is indeed important, significant accomplishments may not be possible until effective procedures for analyst interpretation are developed which can minimize or eliminate these inconsistencies. It would therefore seem that good solutions to this problem would take about 4-6 years. (Priority III).

2.4 Computational Structure

Computational structure refers both to the method of representing digital image data (data structure) for subsequent analysis and the architecture of the computer system used to perform the analysis.

The issue of appropriate data structure design for pattern recognition and for image processing should be clearly separated from that of data base design for resource management purposes. The object of data structure design is to render possible the efficient execution of algorithms for preprocessing, modeling, and object scene inference. The role of data base definition, on the other hand, is to facilitate the retrieval of the processed information in a manner conducive to its manipulation in conjunction with extrinsic sources of information. Data structures are thus primarily concerned with machines and algorithms, while data bases are primarily concerned with the user.

Research over the years has shown that many of the methods used in pattern recognition can be arranged in a parallel structure. As a simple example, consider the familiar linear discriminant. Here each element (feature) of the pattern vector is multiplied by a weight, and the resulting weighted elements are summed to obtain the discriminant value. The multiplications and partial sums can be performed in parallel. The fact that large amounts of data often need to be processed in applications of pattern recognition is a major motivation for considering parallel methods. Indeed, clever formulations with parallel processing concepts can greatly increase the processing speed. In fact, they may render feasible image processing methods which otherwise could not even be considered.

Research Issues by Area

2.4.1 Parallel Processing

This section discusses alternatives to conventional single-instruction, single data stream architectures for preprocessing and classification algorithms. Only stored-program digital computer systems are considered here, although optical, electro-mechanical, and hard-wired digital systems may eventually prove economical in specific high-volume operational applications.

There are about three dozen special-purpose (parallel) machines for pattern recognition currently under various stages of development throughout the western world--about the same number as a decade ago. Most of these machines are built at the chip level, with gate-level design. Since the development of special purpose LSI and VLSI (Very Large Scale Integration) chips is still extremely expensive, greater returns can be expected from designs based on commercial microprocessors, which are now available at a cost of a few dollars each. Bit-slice architectures, in particular, permitting extension to arbitrary word lengths, are promising. It should be noted, however, that advances in the speed of general-purpose digital computers historically have consistently outpaced the improvements in performance offered by special-purpose machines, and there is no real indication that the situation has changed.

Among special purpose digital computer configurations of interest in pattern recognition, the following are noted:

- a. Multiple-instruction, single data stream machines. These are essentially pipe-line machines whose programming and behavior are for

programs with low branching factors not radically different from those of conventional machines. They seem to require no research in pattern recognition other than occasional benchmarking for price-performance index.

b. Single-instruction, multiple data stream machines. Most of the array processors fall into this category, with essentially a single control unit and multiple arithmetic and logic units. These machines are suited for classical pattern recognition.

Research Issues--Parallel Processing

Since the composition of the Working Group and of the group of invited participants did not include specialists in this area, research issues should be elaborated further by appropriate specialists. The following issues represent issues identified by the Working Group.

1. The applicability of special purpose processors to sets of images (multi-spectral, multi-temporal, and multi-sensor) needs to be evaluated. (Priority I).

2. Operating or supervisory systems ("kernels") for applications of pattern-recognition need to be evaluated. (Priority III).

3. Storage-hierarchy configurations matched with both algorithms and data volume need to be investigated. (Priority III).

4. The possibility of applying special purpose processors to interactive processing and displays needs to be considered. (Priority I).

5. Special I/O devices for cartographic applications need to be interfaced with new processor configurations. (Priority II).

2.4.2 Image Data Structures

Data structures are largely dependent on the storage hierarchy selected. Different structures are appropriate for random-access memories, block-access memories (such as disks), and sequential-access memories such as magnetic tape. The characteristics of the processor itself, such as word-size, internal bus configuration and data transfer paths, direct memory access, and the operating system, must also be taken into account. Parallel and special purpose machines impose special structural considerations of their own.

The following examples give some idea of the diversity of data structures already proposed or used in pattern recognition and image processing:

a. Bit-plane structures. Originally developed for Illiac III, these structures store separately each power of two of the intensity levels. This method is used most notably in the PAX image processing packages implemented at the University of Maryland.

b. Pixel-by-pixel storage. In this straightforward method, successive rows or columns of an image are stored sequentially. Unfortunately, no standard format exists, and most programs do not have the flexibility to process arrays of variable size. Powers of two are becoming increasingly popular as preferred dimensions. Adaptive delta-modulation may reduce the total number of bits required at the expense of increased storage complexity.

c. Chain encoding. Proposed by Freeman in the early sixties, chain encoding provides an efficient means of encoding the boundaries of blocks

of homogenous areas. Although the original formulation was restricted to vectors connecting adjacent pixels, "long" vectors were introduced which made possible studies of the trade-off between accuracy of boundary representations and storage costs. Vector-encoded images are particularly appropriate for shape recognition. Numerous algorithms exist for converting vector-coded data to grid-cell coded data.

d. Contour coding. If the variations in intensity are relatively smooth, contour coding is a viable alternative to pixel-by-pixel storage of grey scale images. Contours are generally represented in the form of vectors.

e. Tightly closed boundary (TCB) structure. In this scheme, proposed by Merrill, the points on the boundaries separating homogenous areas (or contours) are sorted by one of their cartesian coordinates. This structure leads to fast algorithms for many operations involving several images.

f. Pyramid or quad-tree structures. The subject of numerous recent papers, these hierarchical data structures divide the image into successive quadrants. Only quadrants containing non-homogeneous information, however, are so divided, resulting in considerable storage savings. Again, algorithms exist for converting pyramid-encoded data to one of the standard forms, but many operations can be performed directly on the encoded data.

g. Two-dimensional polynomial approximation. Smoothly varying levels of intensity can be encoded on a sparse regular or irregular grid structure using, for example, spline functions. Such encoding results in

considerable storage savings and may also result in improved classification because the pixel-to-pixel correlations would automatically be taken into account. Such data structures would also be directly compatible with digital terrain models.

h. Computational geometry. Recent mathematical advances have extended the theory of linear sorting and searching to two-dimensional geometric structures such as points, lines, and areas. Many common operations such as nearest neighbor location can be executed with an order-of-magnitude faster than conventional approaches.

i. Symbolic encoding. Entities which occur frequently in an image or a set of images may be assigned a symbolic label, and the information preserved in the form of symbol-coordinate pairs. Image information in such symbolic form may then be used in contextual, syntactic, or relational classification methods, and to develop models at successively higher levels of abstraction.

Research Issues--Image Data Structures

1. For lossless (information-preserving) data structures, efficient interconversion methods need to be developed. (Priority I).

2. Time/space trade-offs must be developed for the various classification methods and data structures mentioned above. Appropriate ways for applying the methods developed in theoretical computer science ("algorithmic computational complexity") to this area are best exemplified by recent work in computational geometry. (Priority I).

3. For lossy, non-invertible transformations, the effects of encoding on the utility (accuracy, continuity, bias, etc.) of the final classification product need to be investigated. (Priority II).

2.5 Continuing Studies

One fruitful aspect of conducting workshops was the identification of interesting new topics being considered by the scientific community which appear to be useful in remote sensing applications. Several of these topics were discussed and proved helpful in identifying research issues. Additional topics which the Working Group felt were not adequately addressed are discussed below.

2.5.1 Polarization Data in Pattern Recognition and Image Analysis

Polarization data is recorded energy for which the polarization angle of the illuminating or reflected energy is also recorded. In active sensors such as radar, illuminating energy is transmitted in a known polarization state (e.g., vertically polarized) and a particular polarization component of the reflected energy is recorded (e.g., the horizontally polarized component). In passive sensors such as optical or thermal sensors only the polarization angle of the recorded energy is known, and the polarization angle of the radiation source must be modeled or assumed.

In addition to the research issues already discussed, digital images composed of polarization data will present another set of research problems. Polarization data will be strongly related to the specular components of the object scene and to the geometry of the sensor, the target, and the illuminating source. Thus, polarization data will present a special set of issues in preprocessing as well as in mathematical representation of the digital image. It is recommended that this area be the subject of future studies.

2.5.2 Computer Architectures and Parallel Processing

The Working Group and the group of invited participants did not include specialists in this area. Nevertheless, it was apparent from workshop presentations and discussions that analyzing multiple digital images will require specially-designed computers with unique processing capabilities. Although a few research issues were tentatively identified (see 2.4.1) this area should be studied further by the appropriate specialists.

2.5.3 Applicability of "Expert" Systems to Interactive Analysis

Interactive analysis implies the use of complex ancillary information in a suitably organized, computer-stored form, by a human specialist working in concert with a computer system. "Expert" systems, on the other hand, form useful judgments from incomplete, uncertain evidence. Although to our knowledge no expert system has yet been developed to assist an analyst in applications of pattern recognition to renewable resources, this topic clearly deserves further attention.

REFERENCES

- Anuta, P. E., Geometric correction of ERTS-1 digital multi-spectral scanner data, Information note 103073 LARS, Purdue University, November (1973).
- Bauer, M. E., Crop identification and area estimation over large geographic areas using Landsat MSS data, LARS technical report 012477, prepared for Goddard Space Flight Center, January (1977).
- Bryant, Jack, AMOEBA: a spatial clustering algorithm for multitemporal Landsat data, Report No. 15, NASA Contract NAS-9-14689-4S, Texas A&M University, Department of Mathematics, College Station, Texas, August (1977).
- Bryant, Jack, On the clustering of multidimensional pictorial data, Pattern Recognition 11, 115 (1979).
- Chittineni, C. B., Learning with imperfectly labeled patterns, Proceedings of the IEEE Conference on Pattern Recognition and Image Processing, Chicago, Illinois, 52 (1979).
- Chittineni, C. B., Maximum likelihood estimation of label imperfection probabilities and its use in the identification of mislabeled patterns, Presented at Machine Processing of Remotely Sensed Data Symposium, LARS, Purdue University (1980).
- Cochran, W. G., Sampling Techniques, John Wiley and Sons, New York (1963).
- Colvocoresses, A. P., Space oblique mercator, a new map projection of the earth, Photogrammetric Engineering, August (1974).
- Dave, J. V., Contrast attenuation factors for remote sensing, IBM Journal of Research and Development 23, March (1979).
- Decell, H. P., Jr. and L. F. Guseman, Jr., Linear feature selection with applications, Pattern Recognition 11, 55 (1979).
- Efron, B., Bootstrap methods: another look at the jackknife, Technical Report No. 32, Division of Biostatistics, Stanford University (1977).
- Feiveson, A. H., Estimating crop proportions from remotely sensed data, Proceedings of the LACIE Symposium (1978).
- Fisher, L., and J. Van Ness, Admissible clustering procedures, Biometrika, 58, 91-104 (1971).

- Fleming, M. D. and R. M. Hoffer, Machine Processing of Landsat Data and DMA Topographic Data for Forest Cover Type Mapping, Proc. 1979 Symp. Machine Processing of Remotely Sensed Data, IEEE Cat. No. 79CH1430-8 MPRSD, 377-390 (1979).
- Fu, K. S., Syntactic Methods in Pattern Recognition, Academic Press, New York (1974).
- Fuller, W. A., Estimation employing post strata, JASA 61, 1172 (1966).
- Glick, N., Additive estimators for probabilities of correct classification, Pattern Recognition 10, 211 (1978).
- Goodman, L. A., Exploratory latent structure analysis using both identifiable and unidentifiable models, Biometrika 61, 215 (1974).
- Gray, H. L. and W. R. Schucany, The Generalized Jackknife Statistic, Marcel Dekker, Inc., New York (1972).
- Hanuschak, G., R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno, and C. Miller, Obtaining timely crop area estimates using ground-gathered and Landsat data, Technical bulletin No. 1609, United States Department of Agriculture, Economics, Statistics and Cooperative Service, August (1979).
- Haralick, R. M., Edge and region analysis for digital image data, Computer Graphics and Image Processing 12, 60 (1980).
- Haralick, R. M., Statistical and structural approaches to texture, Proceedings of the IEEE 67, 786 (1979).
- Haralick, R. M., Structural pattern recognition, arrangements, and theory of covers, Proceedings of the IEEE Conference on Pattern Recognition and Image Processing, Troy, New York, 112-117 (1977).
- Haralick, R. M. and L. Watson, A facet model for image data, Technical report, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (1980).
- Heydorn, R. P., R. M. Bizzell, J. A. Quirein, K. M. Abotteen, C. A. Sumner, Classification and mensuration of LACIE segments, Proceedings of the LACIE Symposium (1978).
- Horn, Berthold K. P. and R. J. Woodham, Landsat MSS coordinate transformations, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University (1979).
- Houston, A. G., A. H. Feiveson, R. S. Chhikara, E. M. Hsu, Accuracy assessment: the statistical approach to performance evaluation in LACIE, Proceedings of the LACIE Symposium (1978).

- Kanal, L., Patterns in pattern recognition: 1968-1974, IEEE Trans. Pat. Recog. IT-20, 697 (1974).
- Kauth, R. J., P. F. Lambeck, W. Richardson, Feature extraction as applied to agricultural crops as seen by Landsat, Proceedings of the LACIE Symposium (1978).
- Kettig, R. L. and D. A. Landgrebe, Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects, IEEE Trans. Geoscience Electronics, Vol. GE-14, 19-26 (1976).
- Konecny, G., Mathematical models and procedures for the geometric restitution of remote sensing imagery, ISP Commission III Congress, Helsinki (1976).
- Lambeck, P. F. and J. F. Potter, Compensation for atmospheric effects in Landsat data, Proceedings of the LACIE Symposium (1978).
- Leberl, F. and W. Kropatsch, Automatic registration of satellite images with a digital map data base, International Society of Photogrammetry, Commission III Symposium, Moscow (1978).
- Lenington, R. K. and M. F. Rassbach, CLASSY: an adaptive maximum likelihood clustering algorithm, Proceedings of the LACIE Symposium (1978).
- McEwen, R. B. and H. R. Jacknow, USGS digital cartographic data base, presented at Auto-Carto IV, November (1979).
- Mikhail, E. M., J. R. Baker and J. C. McGlone, Metric information from aircraft multispectral scanner (MSS) data, Proceedings, American Society of Photogrammetry Spring Convention, Washington, D.C., March (1979).
- Mikhail, E. M. and M. M. Ethridge, Positional information from single and multiple coverage multispectral scanner data, Proceedings of the ASP Fall Convention, Little Rock, Arkansas, October (1977).
- Moayer, M. and K. S. Fu, A tree approach for fingerprint pattern recognition, IEEE Trans. Comput. C-25 (1976).
- Myers, R. M. and S. G. Wheeler, Fisher linear classification and cross-validation bias correction comparison with the LACIE Procedure I small grains estimation, IBM technical memo RES 23-70, Contract No. GBS NAS9-14350, Federal Systems Division, Houston, Texas, August (1979).
- Panton, D. J., A flexible approach to digital stereo mapping, Photo. Engr. and Remote Sensing, December (1978).
- Pavlidis, T., Structural pattern recognition: primitive and juxtaposition relations, in Frontiers of Pattern Recognition, M. S. Watanabe, ed., Academic Press, New York, 421-451 (1972).

- Pavlidis, T., Structural Pattern Recognition, Springer, New York (1977).
- Pore, M. D., Bayesian techniques in stratified proportion estimation
Lockheed Electronics Co., Inc., Technical Report LEC-13490, August
(1979).
- Potter, J. F. and M. Shelton, Effect of atmospheric haze and sun angle
on automatic classification of ERTS-1 data, Proceedings of the 9th
International Symposium on Remote Sensing of Environment, 865-
874, April (1974).
- Proceedings of the 1979 IEEE Computer Society Conference on Pattern
Recognition and Image Processing, IEEE Cat. No. 79CH1428-2C,
IEEE Service Center, 445 Hoes Lane, Piscataway, New Jersey (1979).
- Proceedings of the Technical Sessions of the LACIE Symposium, Vol. II,
NASA Report No. JSC-16015, NASA Johnson Space Center, Houston,
Texas (1978).
- Redner, R. A., Maximum likelihood estimation for parameters in a mixture
model, To appear as a NASA technical memo.
- Rosenfeld, A. and L. S. Davis, Image segmentation and image models,
Proceedings of the IEEE 67, 764, May (1979).
- Synder, J. P., The space oblique mercator projection, Photo. Engr. and
Remote Sensing, May (1978).
- Stone, M., Cross-validatory choice and assessment of statistical
predictions, Presented to the Royal Statistical Society, December
(1973).
- Strahler, A. H., The use of prior probabilities in maximum likelihood
classification of remotely sensed data, in Remote Sensing of
Environment, in press.
- Strahler, A. H., J. E. Estes, P. F. Maynard, F. C. Mertz, and D. A. Stow,
The use of collateral data in digital image modeling and classifica-
tion, Proceedings of the Fourteenth International Symposium on
Remote Sensing of the Environment, San Jose, Costa Rica, in press.
- Teicher, H., Identifiability of finite mixtures, Ann. of Math. Stat. 34,
1265 (1963).
- Teicher, H., Identifiability of mixtures, Ann. of Math. Stat. (1961).
- Tenenbein, A., A double sampling scheme for estimating from binomial data
with misclassifications, JASA 65, 1351 (1970).

- Tenenbein, A., A double sampling scheme for estimating from binomial data with misclassifications: sample size determination, Biometrics 27, 935 (1971).
- Tenenbein, A., A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection, Technometrics 14, 187 (1972).
- Tubbs, J. D. and W. A. Coberly, Spatial correlation and its effect upon classification results in Landsat, Proceedings of the 12th International Symposium on Remote Sensing of the Environment, Manila, Phillipines (1978).
- Vanderbilt, V. C., Simulated response of a multispectral scanner over wheat as a function of wavelength and view/illumination directions, Paper prepared for publication, LARS, Purdue University, April (1980).
- Wheeler, S. G., R. P. Heydorn, P. N. Misra, and R. T. Smart, An evaluation of Procedure 1, Proceedings of the LACIE Symposium (1978).
- Yakowitz, S. and J. Spragins, On the identifiability of finite mixtures, Ann. of Math. Stat. 39, 209 (1968).

Appendix

WORKSHOP ON REGISTRATION & RECTIFICATION OF REMOTE SENSING DATA

Texas A&M University
January 10-11, 1980
Room 402, Rudder Tower

January 10, 1980

8:30 - 9:00	Coffee & donuts
9:00 - 9:15	Basic Research Program - Overview R. B. MacDonald, NASA/Johnson Space Center
9:15 - 9:30	Pattern Recognition & Image Analysis - Overview L. F. Guseman, Jr., Texas A&M University
9:30 - 10:00	Workshop Overview & Guidelines E. M. Mikhail, Purdue University
10:00 - 10:30	Break
10:30 - 12:00	Registration/Rectification: Which Comes First? D. J. Panton, CDC
12:00 - 1:30	Lunch
1:30 - 3:00	Registration of Multitemporal/Multisource Data P. E. Anuta, Purdue/LARS
3:00 - 3:30	Break
3:30 - 5:00	Registration/Rectification Considerations for Radar Imagery R. Marque, Goodyear Aerospace Corp.
5:00 - 5:30	The Resampling Problem R. Dye, ERIM

Dinner at Texan by Arrangement

January 11, 1980

8:00 - 8:30	Coffee & Donuts
8:30 - 10:00	Digital Terrain and Remote Sensing R. McEwen, USGS/DAT
10:00 - 10:30	Break
10:30 - 12:00	Registration/Rectification for Weather Satellite Data B. Remonti, NOAA
12:00 - 1:30	Lunch
1:30 - 4:30	Working Group Review (Workshop Participants welcome)

WORKSHOP ON REGISTRATION AND RECTIFICATION OF REMOTE SENSING DATA

January 10-11, 1980

Texas A&M University

Workshop Coordinator:

Dr. Edward M. Mikhail, Professor
School of Civil Engineering
Purdue University
West Lafayette, Indiana 47907

Attendees:

Mr. Paul Anuta
Laboratory for Applications of
Remote Sensing
Purdue University
West Lafayette, Indiana 47907
(317) 749-2052

Dr. Ralph Bernstein
IBM Scientific Center
1530 Page Mill Road
Palo Alto, California 94304
(413) 855-3126

Mr. Robert H. Dye
ERIM
3300 Plymouth Road
Ann Arbor, Michigan 48107
(313) 994-1200

Mr. Richard Juday
NASA/Johnson Space Center
Mail Code SF3
Houston, Texas 77058
(713) 483-3611

Mr. Robert Marque
Goodyear Aerospace Corp. - Bldg. 13
Litchfield Park, Arizona 85340
(602) 932-7202

Dr. Robert McEwen
Team Leader, DAT
519 National Center
U. S. Geological Survey
Reston, Virginia 22092
(703) 860-6294

Mr. Dale J. Panton
Control Data Corp. - Station HQM 909
2800 East Old Shakopee Road
Bloomington, Minnesota 55440
(612) 853-6929

Mr. Ben Remondi
Bldg. FOB-4
NOAA/NESS
Suitland, Maryland 20031
(301) 763-2516

Dr. Sam Rifman
TRW-Defense and Space Systems Group
One Space Park
Redondo Beach, California 90278
(213) 536-2340

Dr. Al Zobrist
Jet Propulsion Laboratory
Pasadena, California 91106
(213) 354-3237

WORKSHOP ON DIGITAL IMAGE MODELING

Texas A&M University
February 21-22, 1980
Holiday Inn South, College Station, Texas

February 21, 1980

- 8:30 - 9:00 Coffee & Donuts
- 9:00 - 9:15 Basic Research Program - Overview
R. B. MacDonald, NASA/Johnson Space Center
- 9:15 - 9:30 Pattern Recognition & Image Analysis - Overview
L. F. Guseman, Jr., Texas A&M University
- 9:30 - 9:45 Digital Image Modeling - Overview
W. A. Coberly, University of Tulsa
- 9:45 - 10:45 Problems of Digital Image Modeling for Landsat
Q. A. Holmes, ERIM
- 10:45 - 11:00 Break
- 11:00 - 12:00 Digital Image Analysis
A. Rosenfeld, University of Maryland
- 12:00 - 12:30 Discussion
- 12:30 - 1:30 Lunch
- 1:30 - 2:30 Spatial Features and Data Compression
R. Mitchell, Purdue University
- 2:30 - 3:30 Facet Model
R. Haralick, VPI
- 3:30 - 3:45 Break
- 3:45 - 4:45 Terrain Models and Their Uses
R. Woodham, University of British Columbia

Dinner at Texan by Arrangement

February 22, 1980

- 8:00 - 8:30 Coffee & Donuts
- 8:30 - 9:30 Texture
Shin-Yi Hsu, State University of New York, Binghamton
- 9:30 - 10:30 Model Validation
D. S. Simonett, University of California, Santa Barbara
- 10:30 - 10:45 Break
- 10:45 - 12:30 Discussion of Research Questions
A. H. Strahler, University of California, Santa Barbara
- 12:30 - 1:30 Lunch
- 1:30 - ---- Working Group Meeting

WORKSHOP ON DIGITAL IMAGE MODELING

February 21-22, 1980

Texas A&M University

Workshop Coordinators:

Dr. William A. Coberly,
Associate Professor and Chairman
Division of Mathematical Sciences
University of Tulsa
Tulsa, Oklahoma 74104
(918) 592-6000, Ext. 228

Dr. Alan Strahler, Assistant Professor
Department of Geography
University of California
Santa Barbara, California 93106
(805) 961-3772

Attendees:

Dr. Norman Griswold, Associate
Professor
Department of Electrical Engineering
Texas A&M University
College Station, Texas 77843
(713) 845-7441

Dr. Robert Mitchell
Department of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907
(317) 493-3362

Dr. Forrest Hall, Division Scientist
for Earth Observations Division
NASA/Johnson Space Center, Code S
Houston, Texas 77058
(713) 483-4775

Dr. Emanuel Parzen, Distinguished
Professor of Statistics
Institute of Statistics
Texas A&M University
College Station, Texas 77843
(713) 845-3141

Dr. Robert Haralick,
Department of Electrical Engineering
Virginia Polytechnic Institute
Blacksburg, Virginia 24060
(703) 961-5961

Dr. Azriel Rosenfeld
Computer Science Center
University of Maryland
College Park, Maryland 20742

Dr. Quentin Holmes
Environmental Research Institute of
Michigan
P. O. Box 8618
Ann Arbor, Michigan 48107
(313) 994-1200

Dr. David S. Simonett
Department of Geography
University of California
Santa Barbara, California 93106
(805) 961-3139

Dr. Shin-Yi Hsu
Department of Geography
State University of New York
Binghamton, New York 13901
(607) 798-6502

Dr. Robert Woodham
Faculty of Forestry
2357 Main Mall
University of British Columbia
Vancouver, British Columbia V6T 1W5
Canada
(604) 228-4918

WORKSHOP ON DIGITAL IMAGE PATTERN RECOGNITION

Texas A&M University
March 26-28, 1980
Room 402, Rudder Tower

March 26, 1980

Meet in lobby of Holiday Inn, 7:45 a.m., for transportation to Rudder Tower

Morning Session: Chairman - George Nagy, University of Nebraska

8:00 - 8:30 Coffee & Donuts

8:30 - 8:45 Basic Research Program - Overview
R. B. MacDonald, NASA/Johnson Space Center

8:45 - 9:00 Pattern Recognition and Image Analysis - Overview
L. F. Guseman, Jr., Texas A&M University

9:00 - 9:45 Mapping and Monitoring
A. H. Strahler, University of California, Santa Barbara

9:45 - 10:45 Inventory
R. P. Heydorn, NASA/Johnson Space Center

10:45 - 11:00 Break

11:00 - 12:00 Classification Based Upon Multiple Sources of Data
P. H. Swain, LARS/Purdue University

12:00 - 1:00 Lunch

Afternoon Session: Chairman - R. P. Heydorn, NASA/Johnson Space Center

1:00 - 2:30 Context and Consistent Labeling
R. Haralick, V.P.I.

2:30 - 3:45 Image Texture Analysis
L. Davis, University of Texas at Austin

3:45 - 4:00 Break

4:00 - 5:30 Feature Selection, Extraction and Combinations
H. P. Decell, University of Houston

Dinner at Texan by Arrangement

March 27, 1980

Meet in lobby of Holiday Inn at 8:15 for transportation to Rudder Tower

Morning Session: Chairman - P. H. Swain, LARS/Purdue University

8:30 - 9:00 Coffee & Donuts

9:00 - 10:00 Pictorial Data Bases and Data Structures
S. K. Chang, University of Illinois at Chicago Circle

10:00 - 11:00 Image Data Structures and Relation to Image Analysis
S. Tanimoto, University of Washington

11:00 - 11:15 Break

11:15 - 12:00 Interactive Pattern Recognition
Y. T. Chien, University of Connecticut

12:00 - 1:00 Lunch

Afternoon Session: Chairman - R. P. Heydorn, NASA/Johnson Space Center

1:00 - 2:30 Clustering
J. VanNess, University of Texas at Dallas

2:30 - 3:45 Estimators for Probability of Correct Classification
N. Glick, University of British Columbia

3:45 - 4:00 Break

4:00 - 5:30 Working Group Meeting

March 28, 1980

Meet in lobby of Holiday Inn at 8:15 for transportation to Rudder Tower

Morning Session: Chairman - George Nagy, University of Nebraska

8:30 - 9:00 Coffee & Donuts

9:00 - 10:30 Short Presentations - Attendees

10:30 - 10:45 Break

10:45 - 12:00 Workshop Wrap-up - Attendees

12:00 - 1:00 Lunch

Afternoon Session: Chairman - R. P. Heydorn, NASA/Johnson Space Center

1:00 - ---- Discussion - Research Objectives
Working Group and Workshop Participants

WORKSHOP ON DIGITAL IMAGE PATTERN RECOGNITION

March 26-28, 1980

Texas A&M University

Workshop Coordinators:

Dr. Richard Heydorn
Earth Observation Division
NASA/Johnson Space Center (SF3)
Houston, Texas 77058
(713) 483-5305

Dr. Philip Swain
LARS/Purdue University
1220 Potter Drive
West Lafayette, Indiana 47906
(317) 749-2052

Dr. George Nagy, Professor and
Chairman
Department of Computer Science
University of Nebraska 68588
(402) 472-3200

Attendees:

Dr. S. K. Chang
Department of Information Engineering
University of Illinois at Chicago
Circle
Chicago, Illinois 60680
(312) 996-5494

Dr. R. M. Haralick
Department of Electrical Engineering
Virginia Polytechnic Institute
Blacksburg, Virginia 24060
(703) 961-5961

Dr. Y. T. Chien
Department of Electrical Engineering
and Computer Science
University of Connecticut
Storrs, Connecticut 06268
(203) 486-4816

Dr. William MacFarland
Department of Electrical Engineering
University of Missouri
Columbia, Missouri
(314) 882-6387 or 3379

Dr. Larry Davis
Computer Science
University of Texas at Austin
Austin, Texas 78712
(512) 471-7316

Dr. Emanuel Parzen, Distinguished
Professor of Statistics
Institute of Statistics
Texas A&M University
College Station, Texas 77843
(713) 845-3141

Dr. Ned Glick
Department of Mathematics
University of British Columbia
Vancouver, British Columbia
(604) 228-6621

Dr. Sam Shanmugan
Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas 66045

Dr. Forrest Hall
Division Scientist for Earth
Observations Division
NASA/Johnson Space Center
Houston, Texas 77058
(713) 483-4776

Dr. Steve Tanimoto
Computer Science Department
University of Washington
Seattle, Washington
(206) 543-4848 or 1595

Dr. John VanNess
Mathematical Sciences Department
University of Texas at Dallas
Richardson, Texas 75080
(214) 490-2166

MATHEMATICAL PATTERN RECOGNITION AND IMAGE ANALYSIS

SCHEDULE OF MEETINGS

- OCTOBER 4-5; WORKING GROUP - COLORADO STATE UNIVERSITY
BRIEFING BY APPROPRIATE NASA PERSONNEL ON SELECTED APPLICATION
RESEARCH PROJECTS AND LONGER TERM PLANNING
- NOVEMBER 5-6; WORKING GROUP - TEXAS A&M UNIVERSITY
TECHNICAL OVERVIEW OF GENERAL RESEARCH AREAS; ORGANIZATION
OF FUTURE WORKSHOPS
- DECEMBER 17-18; WORKING GROUP - TEXAS A&M UNIVERSITY
TECHNICAL OVERVIEW OF GENERAL RESEARCH AREAS; ORGANIZATION
OF FUTURE WORKSHOPS
- JANUARY 10-11; WORKSHOP - TEXAS A&M UNIVERSITY
REGISTRATION & RECTIFICATION
- FEBRUARY 21-22 (+23); WORKSHOP - TEXAS A&M UNIVERSITY
DIGITAL IMAGE MODELING
- MARCH 26-28 (+29); WORKSHOP - TEXAS A&M UNIVERSITY
DIGITAL IMAGE PATTERN RECOGNITION
- JUNE 9-10; WORKING GROUP - TEXAS A&M UNIVERSITY
REVISE BASIC RESEARCH PLAN AND IMPLEMENTATION/COORDINATION
PLAN