LARS Information Note 110474

NASA CR-
141860

(NASA-CR-141860) AN INTRODUCTION TO   N75-26477
QUANTITATIVE REMOTE SENSING (Purdue Univ.)
                                    CSCL 20A
                                              Unclas
                              G3/43   26666

AN INTRODUCTION TO QUANTITATIVE REMOTE SENSING

John Lindenlaub
James Russell

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1974   PRICES SUBJECT TO CHANGE

Information Note 110474

AN INTRODUCTION TO QUANTITATIVE REMOTE SENSING

by

John Lindenlaub

and

James Russell

The Laboratory for Applications of Remote Sensing

Purdue University

West Lafayette, Indiana

This information note supersedes
Remote Sensing Analysis: A Basic Preparation
by John Lindenlaub (Information Note 110471)

i a

PREFACE

This booklet discusses the quantitative approach to remote sensing and provides an introduction to the analysis of remote sensing data. It stresses the application of pattern recognition in numerically oriented remote sensing systems. The specific purpose of this material is to provide a common background and orientation for users of the LARS computer software system. The booklet is the first in a series of educational units dealing with the analysis of remote sensing data. However, the booklet can be utilized in many other ways. Newcomers to remote sensing will be introduced to basic concepts and terminology. Remote sensing veterans can increase their understanding of numerical analysis and automatic data processing.

The format of this booklet resembles that of a programmed text. It is not necessarily intended that you read the materials from cover to cover. Spread throughout the manual are questions about the material. Do your best to answer these questions as you come to them. Depending upon your responses, you will be directed to various sections of the manual. In this way your study may be tailored to your individual needs.

The material presented here has been drawn from five LARS Information Notes:

"On Pattern Recognition" by G. Cardillo and D. Landgrebe (101866)

"Systems Approach to the Use of Remote Sensing" by D. Landgrebe (041571)

"Remote Sensing Analysis: A Basic Preparation" by J. Lindenlaub (110471)

"Machine Processing for Remotely Acquired Data" by D. Landgrebe (031573)

"Identification of Agricultural Crops by Computer Processing of ERTS MSS Data" by M. Bauer and J. Cipra (030173)

Editorial assistance was provided by Tom Builta, Tina Cary, Shirley Davis and Philip Swain. The authors wish to express their appreciation to all who contributed to this booklet.

John Lindenlaub

James Russell

TABLE OF CONTENTS

SECTION I

AN INTRODUCTION TO REMOTE SENSING

Objectives

Upon completion of Section I, you should be able to:

1. Define "remote sensing."

2. Identify three types of measurable electromagnetic field variations that are used in remote sensing.

3. Describe the electromagnetic spectrum from 0.3 to 15.0 micrometers. Your description should include the approximate location of the ultraviolet, visible and infrared portions of the spectrum.

4. Describe a remote sensing system, such as an earth survey system, and its application.

5. Distinguish between image-oriented and numerically oriented remote sensing systems. Your discussion should include one example of each type.

6. State at least one advantage and one disadvantage of each of the following types of sensors:

> a) photographic
> b) television
> c) scanner

Pretest for Section I

You can check your understanding of the information contained in Section I by attempting to answer the following questions:

1. Define "remote sensing."

2. Remote sensing is dependent largely upon certain types of measurable variations in electromagnetic fields. These are

    _____, _____ and _____ variations.

3. Match the following parts of the electromagnetic spectrum to the corresponding wavelength bands.

    _____ visible   A. 0.30 to 0.40 micrometers

    _____ ultraviolet   B. 0.40 to 0.70 micrometers

    _____ thermal infrared   C. 0.70 to 3.00 micrometers

    _____ reflective infrared D. 3.00 to 15.0 micrometers

                                 E. 15.0 to 30.0 micrometers

4. Describe an application of remote sensing.

Pretest for Section I (cont.)

5. Complete the following:

Type of Sensing System	Characteristics	Example
IMAGE-ORIENTED		
NUMERICALLY ORIENTED		

6. Complete the following:

TYPE OF SENSOR	ADVANTAGE	DISADVANTAGE
Photographic		
Television		
Scanner		

Check your answers to the pretest questions in the Appendix on page 57.

If you missed four or more questions, we suggest you read the entire section.

If you missed fewer than four questions, you might want to refer just to those parts of Section I pertaining to the questions you missed.   Then take the posttest for Section I on page 11.

If you did not miss any questions, go directly to Section II on page 13 and try the pretest.

THE PHYSICAL BASIS OF REMOTE SENSING

Imagine that you are high above the surface of the earth.
You want to survey the earth's surface in order to learn about
its resources in hope of managing them better. How could
information about these resources be obtained? First, you will
need to obtain measurements of some natural phenomena. From
these data you can form a hypothesis about what is on the ground.
By definition, you will be engaged in remote sensing, "the
science of acquiring information about objects from measurements
made at a distance, without physical contact with the objects."
These measurements are, in fact, measurements of variations in
field strengths; electromagnetic, force and acoustical wave
fields are common examples. This booklet is restricted to
considerations of variations only in electromagnetic fields.
Three important types of electromagnetic field variations are:

- Spectral variations -- changes in the intensity of
  radiation at given wavelength(s);
  i.e., differences in color

- Spatial variations -- changes in radiation from one
  location to another; i.e.,
  differences in shape and position

- Temporal variations -- changes in radiation from one
  time to another; i.e., dif-
  ferences over time

In order to derive information from these field variations, one
must:

- Measure the variations and

- Compare these measurements to those
  of known natural or man-made objects.

For example, suppose you need a map showing all of the
bodies of water in a certain region. Water at the surface of
the earth can not be sensed directly from spacecraft altitudes;
however, the manifestations of these bodies of water do exist at
those heights. These manifestations, in the form of electromag-
netic radiation, can be measured and the measurements analyzed to
determine which points on the earth are water and which are not.

In order to understand the use of spectral variations of
the electromagnetic spectrum in remote sensing, you need some
basic information about the electromagnetic spectrum.

| COSMIC RAYS | GAMMA RAYS | X-RAYS | U-V | | INFRARED | MICRO-WAVES | T-V | RADIO | ELECTRIC POWER |

VISIBLE SPECTRUM   INFRARED

0.40   0.50   0.60   0.70   0.80   13.5
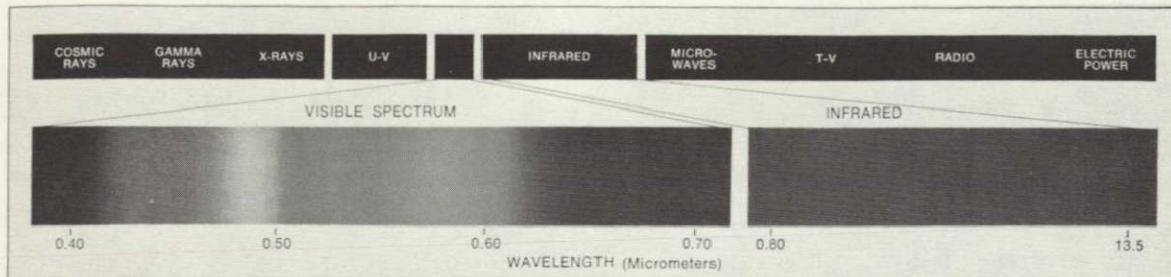
WAVELENGTH (Micrometers)

Figure 1. The Electromagnetic Spectrum

The portion of the electromagnetic spectrum from 0.3 to 15.0 micrometers, referred to as optical wavelengths, is the range of wavelengths used by remote sensing devices which depend upon the sun as the primary energy source.

The portion of the electromagnetic spectrum visible to the eye extends from 0.4 to 0.7 micrometers. The wavelengths shorter than 0.4 micrometers are in the ultraviolet region. The portion above the visible spectrum is the infrared region, with 0.7 to approximately 3 micrometers referred to as the reflective infrared and the region from 3 to 15 micrometers called the emissive or thermal infrared region. In the thermal infrared region energy is emitted from an object as a result of its heat rather than being reflected from it.

The reflected and emitted electromagnetic radiation can be measured by an earth survey system such as that shown in Figure 2. There is a certain amount of data processing on board the sensor vehicle, including the merging of data concerning sensor calibration and the direction in which the sensor was pointed.

The data is then transmitted to earth for analysis and further processing. This may be done through telemetry, as in the case of the Earth Resources Technology Satellites (ERTS), or through direct package return as with SKYLAB. There is usually a need then for certain preprocessing of the data followed by one or more stages of data analysis. At this point, when data is converted to information, it is usually helpful to merge ancillary information (perhaps derived from sources on the surface of the earth) with the remotely sensed information.
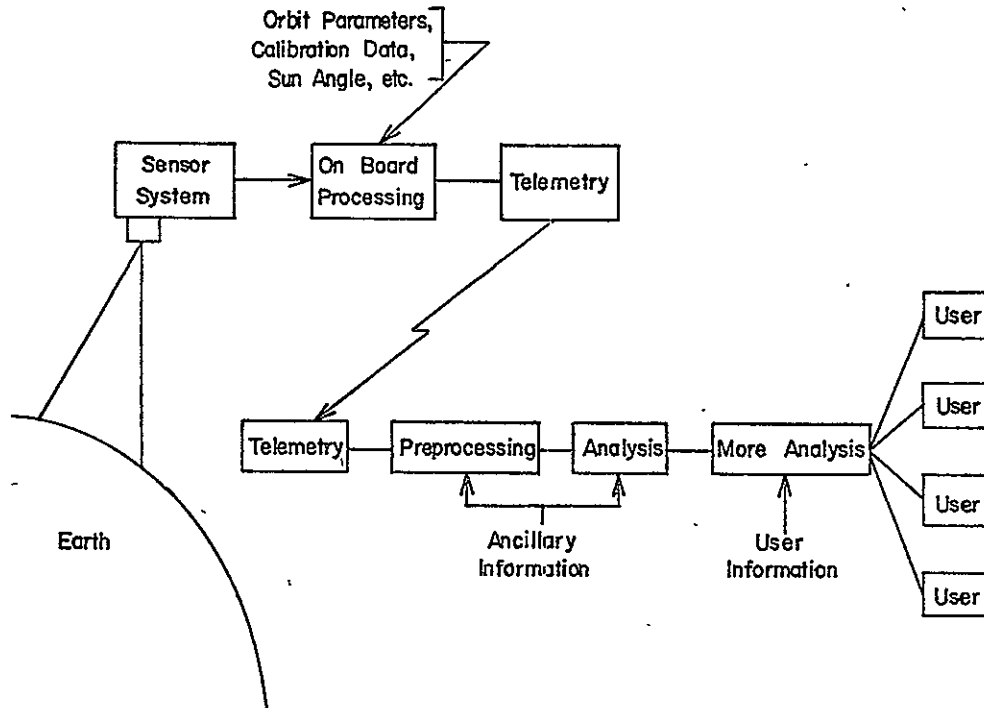
Figure 2. Organization of an Earth Survey System

An important part of the system is indicated by the right-hand blocks in Figure 2: the users. In the case of an earth resources information system, utilization of the information can prove to be a challenge since many potential consumers of the information are not accustomed to receiving it from a space system and may indeed know very little about the information-providing capabilities of such systems.

Summarizing, it is possible to derive information about the earth's surface and the condition of its resources by measuring the spectral, spatial, and temporal variations of the electromagnetic fields emanating from points of interest and then analyzing these measurements to relate them to specific classes of materials. To do so profitably, however, requires an adequate understanding of the materials to be sensed and, in order to make the information most useful, a precise knowledge about how the information will be used and by whom.
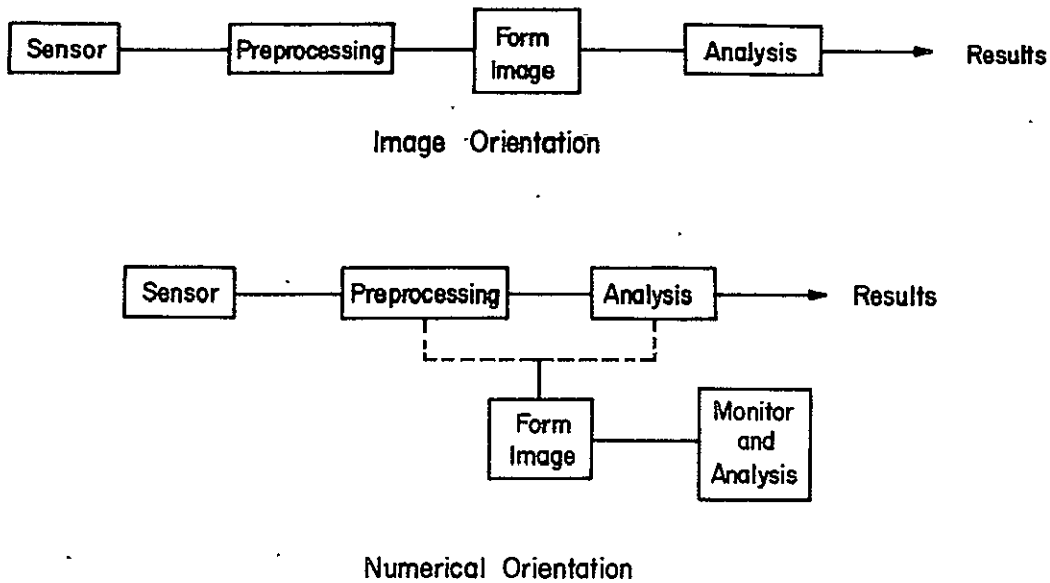
THE DUALITY OF REMOTE SENSING SYSTEMS

Figure 3. Organization of Image-Oriented and Numerically
Oriented Systems.

At the present state of remote sensing technology, there
are two basic types of systems, which will be referred to here
as

·Image-oriented, and

·Numerically oriented.

An example of an image-oriented system is an aerial camera
and a photointerpreter. Photographic film is used to measure
the variations of the electromagnetic fields, and the photo-
interpreter relates these variations to specific classes of
surface cover.

An example of a numerically oriented system is a multi-
spectral scanner and a digital computer. The multispectral
scanner is used to measure variations in electromagnetic fields,
and the resulting data is analyzed with the aid of a computer.

. Comparing the two systems as shown in Figure 3, both systems need a sensor and some preprocessing. The distinction between the two can be brought out most clearly by noting the location of the Form Image block in the two diagrams. In the image-oriented system, forming the image is a step in the primary data stream and must precede the analysis. Numerically oriented systems, on the other hand, need not necessarily form an image. If they do, and in earth resources studies they often do, it may be somewhere other than in the main data stream. Images may be used to monitor the operation of the system and perhaps to do some special analysis.

Since an image is a very efficient way to communicate large quantities of information, let us arbitrarily restrict ourselves to available sensors capable of creating images. Table I shows a categorization of sensors into three broad classes: photographic, television, and scanner. The table also provides examples of the advantages and disadvantages of each.

Table I

Sensor Type	Advantages	Disadvantages
Photographic	Spatial Resolution	Data Return
Television	Size and Weight	Spectral Range
Scanner	Spectral Range	Mechanical Complexity

In the case of photography, the great advantage is the very high spatial resolution (that is, to see small objects distinctly) which can be achieved, but to maintain this high resolution, the data must be returned directly, i.e., the film must be physically carried from the sensor platform to the analyst. Also, photography as a sensor is useful only in the visible and a small part of the reflective infrared portions of the spectrum.

Television sensors operate over approximately the same spectral region as photography. However, television has the advantage that the signal occurs in electrical form and thus is immediately ready to be transmitted to the earth via a radio telemetry link. Thus, one is not faced with the task of carrying along a large quantity of the storage medium (photographic film for photography) when using a television sensor. One may view this as an advantage of size and weight or as one of efficiency in that a television-equipped satellite may be operated a very long time with a single servicing.

Scanners can be built to operate over the entire optical wavelength range and over a large range of intensity levels. In order to achieve these advantages, however, they tend to be mechanically more complex than other sensors.

An important point to realize is that the advantages and disadvantages mentioned above must be considered only as examples since the advantages and disadvantages in any specific instance will depend upon the precise details of the system. General statements are difficult to apply with regard to the type of sensors which are best for image-oriented and numerically oriented systems. Photography tends to be favored for image-oriented systems due to its high spatial resolution capability, while multiband scanners tend to be favored for numerically oriented systems because of their greater spectral and dynamic ranges.

The technology of image-oriented systems is relatively well developed. Reference to the use of aerial photography for military purposes dates back over 100 years. Its use by civil engineers for surveying and mapping represents one of the first civilian uses. Image-oriented systems have the advantage of being easily understood by the layman, an advantage important in the earth resources field with its many new data users. Image-oriented systems are well suited for producing subjective information and are especially suited to circumstances where the classes to be identified in the analysis cannot be precisely decided upon beforehand. Thus, man, with his intuitive and evaluative insights, can be deeply involved in the analysis activity. Pictorially oriented systems can be relatively simple and low-cost, but on the other hand, they are difficult to use for large-scale surveys involving very large amounts of data.

The technology of numerically oriented systems has developed within the last 20 years or so. The technology is based on large scale digital computers (1950's), solid state sensors (early 1960's), and pattern recognition theory (1960's). Because of their greater complexity, numerically oriented systems are not as easily understood by the layman. Numerically oriented systems are especially well suited to situations which require quantitative output (area measurements, disease level, amount of temperature gradient) or where large areas are to be surveyed.

In conclusion, there are two general types of systems. This duality exists primarily for historical reasons; the technologies developed independently in different disciplines. One type is based on image interpretation, and therefore a key goal of an intermediate portion of the system is the generation of a high-quality image. In the case of numerically oriented systems, imagery is less important and indeed may not be necessary at all. These two types of systems should not be viewed as competitive with one another since they have different capabilities and each has advantages under certain circumstances. Hybrid systems using the best characteristics of both types of systems are now evolving.

Posttest for Section I

1. A friend just learned that you are interested in remote sensing and asks, "What is remote sensing?" How would you answer his question briefly and non-technically?

2. Remote sensing systems depend on one or more types of variations in electromagnetic fields. List three of these variations.
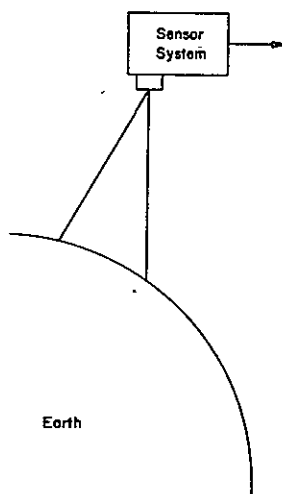
    a. _____

    b. _____

    c. _____

3. The portion of the electromagnetic spectrum from 0.3 to 15.0 micrometers contains the ultraviolet, the visible and the infrared. What are the approximate wavelengths of these three portions of the electromagnetic spectrum?

    ultraviolet = _____

    visible = _____

    infrared = _____

4. Complete the diagram of an Earth Survey System shown below from the sensor to the user.
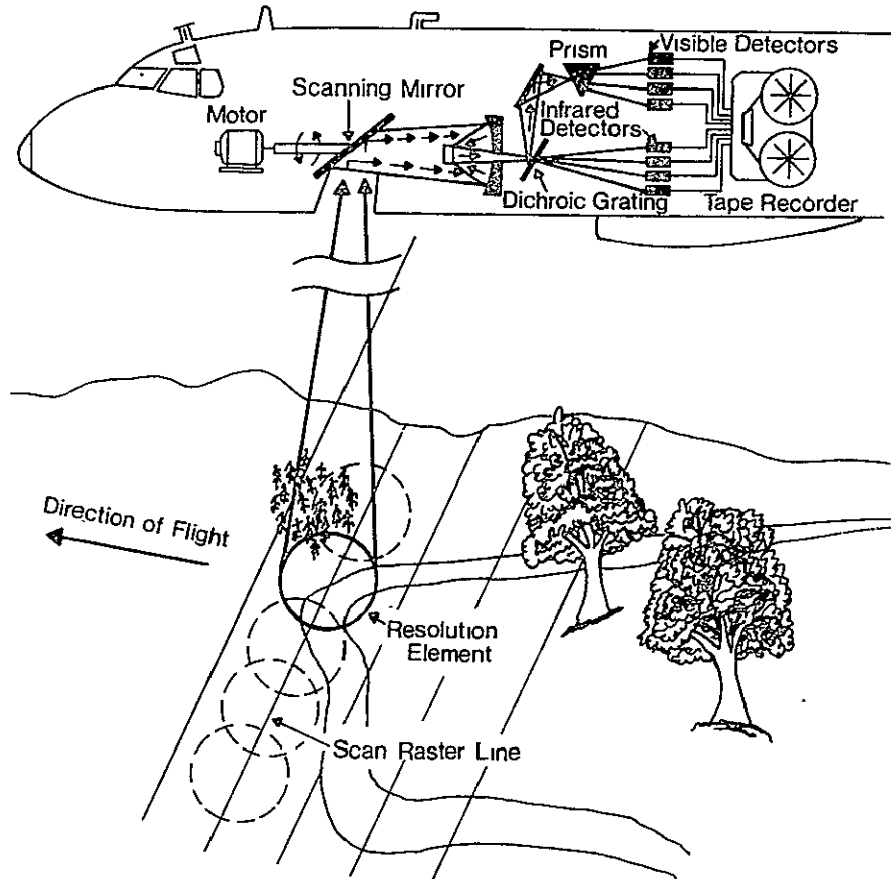
Posttest for Section I (cont.)

5. As they have developed to date, remote sensing systems tend to fall into two categories: image-oriented and numerically oriented systems. Briefly describe how they are different and cite an example of each type of system.

6. State one advantage and one disadvantage of each sensor: photographic, television and scanner.

CHECK YOUR ANSWERS TO THE POSTTEST FOR SECTION I IN THE APPENDIX ON PAGE 58.

If you missed any questions, you may want to go back to that portion of the booklet discussing the questions you missed.

If you answered all the questions correctly, or have checked the text for the ones you missed, you are ready to go on to Section II.

SECTION II

MULTISPECTRAL ANALYSIS

Objectives

Upon completion of Section II, you should be able to:

7. Outline and/or discuss the basic operation of an airborne multispectral scanner.

8. Describe how black and white film may be used to identify various colors.

9. Given a graph showing relative spectral response vs. wavelength for various ground covers, plot their associated points on the λ_1, λ_2 plane.

10. Discuss two uses of temporal information in the identification of crop type.

11. Discuss the use of a multispectral scanner system for crop identification.

12. List three applications of multispectral analysis other than crop identification.

Pretest for Section II

You can check your understanding of the information contained in Section II by attempting to answer the following questions.
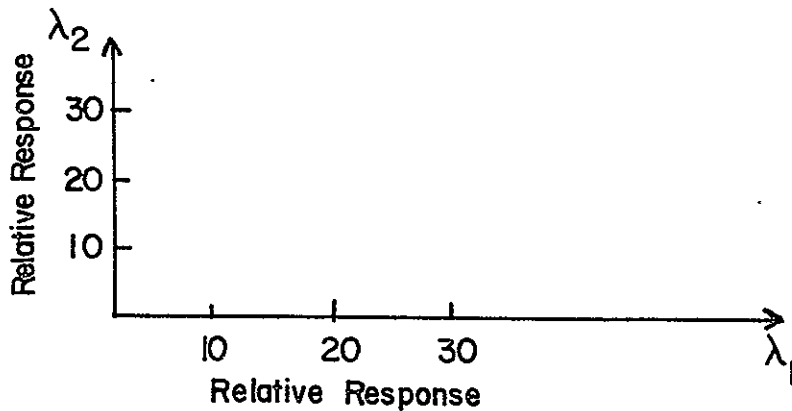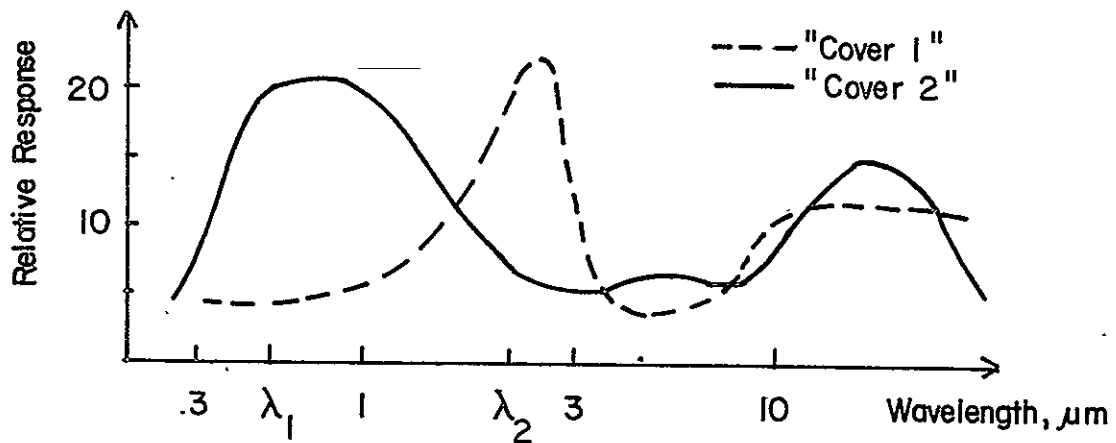
7. Imagine you are to brief a group of visitors on the operation of an airborne multispectral scanner. The figure below is available for reference. List in outline form the main points that you would include in your briefing.

Pretest for Section II (cont.)

8. Briefly describe how black and white photographic film can be used to identify various colors.

9. Let the relative response at wavelength λ_1 and λ_2 be the features used in a remote sensing system. Plot the points in the λ_1, λ_2 feature plane that represent "cover 1" and "cover 2."

Pretest for Section II (cont.)

10. Temporal information can be used to assist in crop identification. Briefly explain how this is possible.

11. Discuss the use of multispectral scanner data for crop identification.

12. List three applications of multispectral analysis other than crop identification.

Check your answers to the pretest questions in the Appendix on page 59-60.

If you missed four or more questions, we suggest you read the entire section.

If you missed fewer than four questions, you might want to refer just to those parts of the section pertaining to the questions you missed. Then take the posttest for Section II on page 29.

Question 7	Page 19
Question 8	Page 21
Question 9	Page 23
Question 10	Page 23-24
Question 11	Page 25
Question 12	Page 25

If you did not miss any questions, go directly to Section III on page 33 and try the pretest.

Visible Detectors

Prism

Scanning Mirror

Motor

Infrared
Detectors

Dichroic Grating

Tape Recorder

Direction of Flight

Resolution
Element

Scan Raster Line
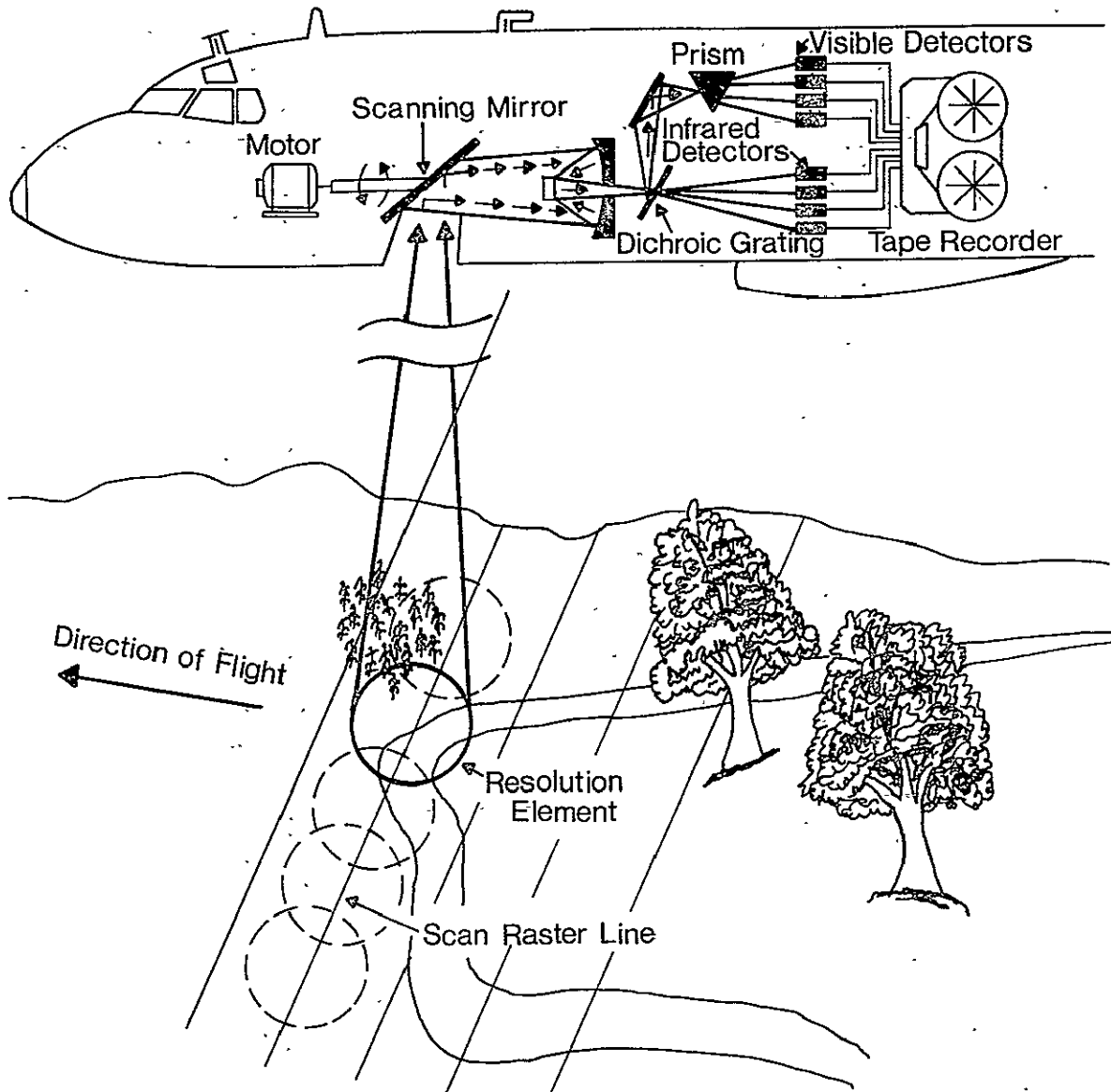
Figure 4. An Airborne Multispectral Scanner

A TYPICAL MULTISPECTRAL SCANNER SYSTEM

Figure 4 illustrates a multispectral scanner mounted in an aircraft. The scanner can also be mounted on other platforms, such as a spacecraft.

Basically, the scanner receives reflected and emitted radiation from a ground resolution element. The radiation is separated by a dichroic grating which reflects the visible wavelengths and transmits wavelengths in the infrared portion of the spectrum. The visible wavelengths are then reflected from a mirror and pass through a prism which differentially refracts the wavelengths in different portions of the visible spectrum. The dichroic grating disperses the infrared radiation. The intensity of the radiation at different wavelengths is then measured by detectors and produces electrical signals which can be recorded on magnetic tape or transmitted directly to the ground by telemetry.

Since all energy from a given element of the earth's surface passes through the same optics simultaneously, the response in each spectral band for that element of the earth's surface can be determined.

In order to view a large area of the earth, the instantaneous field of view of the multispectral unit must scan across the scene and continue this scan pattern as the aircraft moves forward. The scanning is accomplished by a motor-driven mirror as shown in Figure 4. The rotating mirror scans across the scan raster line perpendicular to the forward direction of the aircraft. The motion of the aircraft provides for scanning of the scene in the direction of flight so that eventually every element in the scene has been in the instantaneous field of view of the instrument.
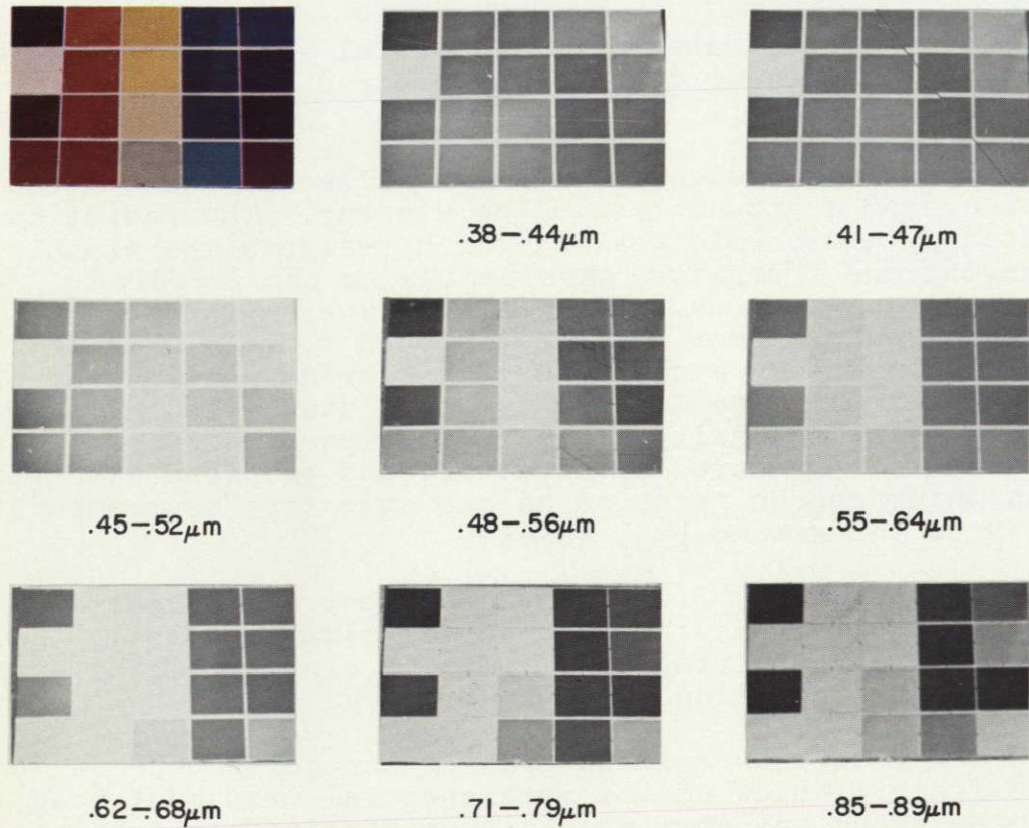
.38−.44μm   .41−.47μm

.45−.52μm   .48−.56μm   .55−.64μm

.62−.68μm   .71−.79μm   .85−.89μm

Figure 5.   Multispectral Photography of Color Cards.

CHARACTERISTICS OF MULTISPECTRAL DATA

Quantitative remote sensing relies heavily upon the "multispectral" approach to data analysis. A basic understanding of the multispectral concept can be obtained by referring to Figure 5. The image in the upper-left corner was produced by photographing a set of color cards with conventional color film-- film designed to give a "natural" color rendition. The remaining images were obtained by photographing the same set of color cards with black and white film using several different optical filters. The wavelengths transmitted by each filter are shown below each gray tone image.

Concentrate on the yellow color cards in each image. A qualitative multispectral description of the yellow cards is given in the middle column of the table below. By arbitarily assigning a 0 to 10 numerical brightness scale a quantitative multispectral description may be obtained as shown in the right hand column of the table.

.38 - .44 μm	dark gray	2
.41 - .47 μm	medium gray	4
.45 - .52 μm	light gray	6
.48 - .56 μm	very light gray	8
.55 - .64 μm	almost white	9
.62 - .68 μm	white	10
.71 - .79 μm	almost white	9
.65 - .89 μm	medium light gray	5

Multispectral Description of the Yellow Color Cards

A similar procedure could be used to obtain a multispectral description for each of the other color cards. If you were to assign numerical values carefully and accurately enough you would find that each color card has a unique "spectral signature." You might want to pick another color and try it below. In summary, the multispectral approach permits the identification of any color card by noting the intensities of gray produced on the various black and white photographs by that particular color card.

.4 - .7μm   .7 - .9μm   4.5 - 5.5μm
                                                          (Artist's Conception)
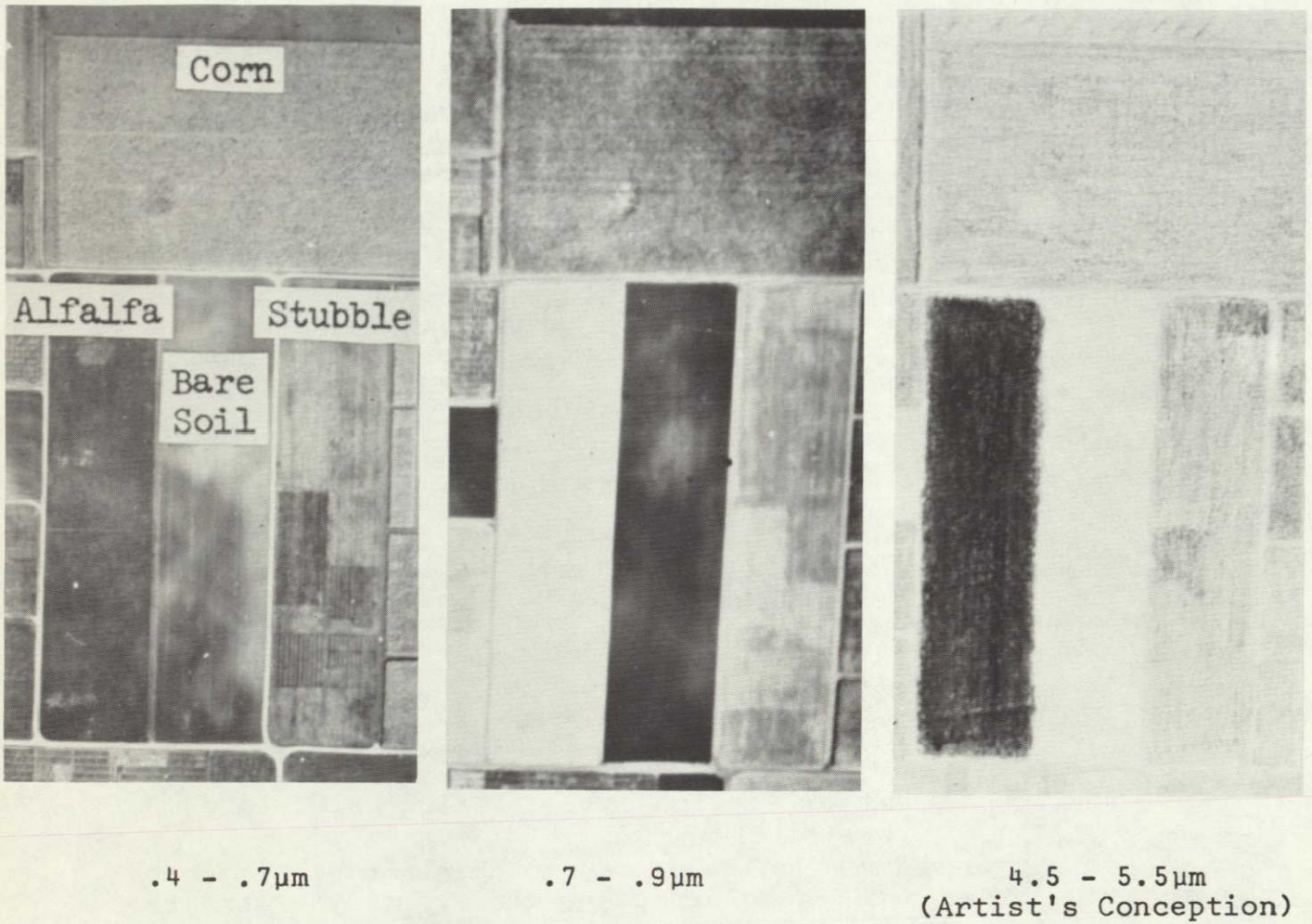
Figure 6. Multispectral Response of Corn, Alfalfa, Stubble
           and Bare Soil

     As a very simple example of the multispectral approach,
Figure 6 shows images of an agricultural scene taken in three
different portions of the spectrum. An examination of Figure 6
will reveal that alfalfa is the only cover type with dark--light--
dark responses in the three spectral bands. On the other hand,
bare soil is gray--dark--white. Thus, alfalfa can be discriminated
from bare soil, stubble and corn by identifying the fields which
are dark--light--dark in these three spectral bands.
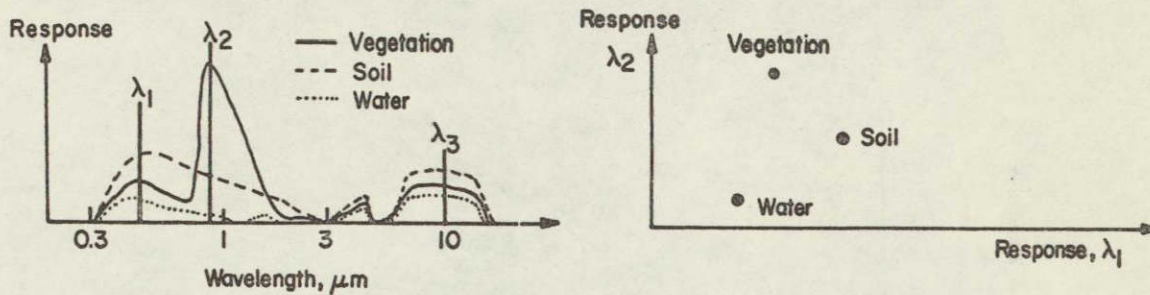
Figure 7.   Spectral Data in Two-Dimensional Feature Space.

To better understand this approach and to see how a numerically oriented system may be based upon it, consider Figure 7. Shown at the left is a graph of relative response (reflectance) as a function of wavelength for three types of earth-surface cover material: vegetation, soil and water. Let two wavelengths be selected, labeled λ_1 and λ_2. Shown at the right side of this figure are the data for the three materials at these two wavelengths, plotted relative to one another. For example, in the left-hand graph, soil has the largest response at wavelength λ_1. This manifests itself in the right-hand graph in the fact that soil has the largest abscissa value (the greatest displacement to the right). Note that a larger number of bands can be used. The response at λ_3 could be used and the data plotted in three dimensions. Four or more dimensions indeed have meaning and utility in numerically oriented systems even though a graphic plot of the data is not possible.

Temporal information can also be useful since time is always a parameter of the spectral response of maturing vegetation. As an example, consider the problem of discriminating between soybeans and corn. Under cultivation, these two plants have approximately 140-day growing cycles. Figure 8 illustrates what the two-dimensional response plot might be for fields of these two plants with time as a parameter. Upon planting and for some time thereafter, fields of soybeans and corn appear as bare soil when viewed from above. Eventually, both plants emerge from the soil, develop a canopy of green vegetation, mature to a brownish dry vegetation, and are harvested. But in addition to the vegetation of the two plants having a slightly different
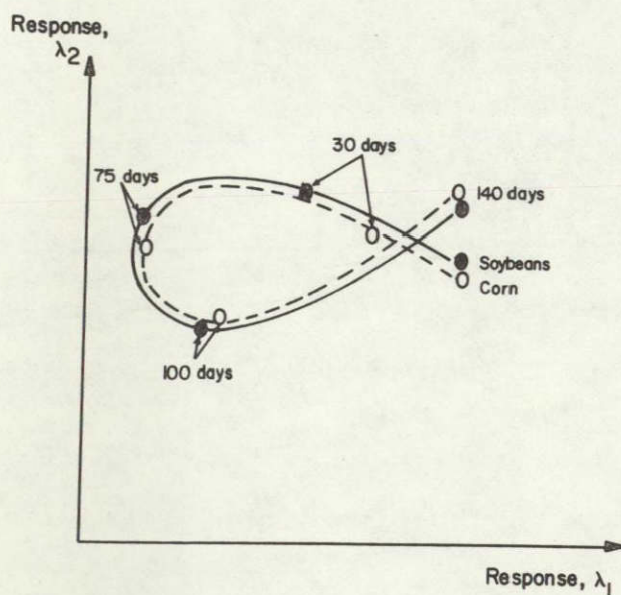
Figure 8.   Temporal Change in Two-Dimensional Feature Space.

response as a function of wavelength, the growing cycles are
different.   Figure 8 shows a rather large difference between
them 30 days from planting (partial canopy) as compared to 75
days (full canopy).   Thus, one way in which temporal information
is used is simply for determining the optimal time at which to
conduct a survey of given materials in order to get optimal
differentiation between the materials.

A second use of temporal information is perhaps less
obvious.   Consider the situation in Figure 8 at the 75- and 100-
day points.   In each case, the separation of the two materials
is relatively slight.   However, if the data are replotted in
four-dimensional space, with λ_1 and λ_2 at 75 days as dimensions
one and two and λ_1 and λ_2 at 100 days as dimensions three and
four, the small separabilities at the two times may reinforce
one another.

A third use of temporal information is for change detection.
In many earth resources problems it is useful to have an accurate
historical record of the changes taking place in a scene as a
function of time.   Urban development, for example, can be
monitored by remote sensing, and land-use analysts are turning
to this technology for assistance.

AN EXAMPLE OF MULTISPECTRAL ANALYSIS

As an example of the use of a multispectral scanner system of the type shown in Figure 4, let's turn now to a sample classification of a one-mile by four-mile area into agricultural classes. A conventional panchromatic air photo of the scene is shown at the left in Figure 9 (next page). The identification of each field as determined by ground observation has been added to the photo. The symbols on the air photo refer to the follow-ing cover types: S-soybeans, C-corn, O-oats, W-wheat, A-alfalfa, T-timothy, RC-red clover, R-rye, SUDAN-sudan grass, P-pasture, DA-diverted acres and H-hay. Four-dimensional data (four spectral bands) were used for the classification. Responses in these four bands are shown in Figure 9. Figure 10 (page 27) shows the results of the classification in two different ways. In the center of the figure, all points of the scene classified as row crops (either corn or soybeans) are printed. On the right, all points classified as cereal grains (either wheat or oats) are printed.

A quantitative evaluation of the accuracy was obtained by tabulating the percentage of correct classification for a number of test fields in the flight line. Test fields are fields known from reference data to be a certain crop or cover type. The results of this tabulation are shown in Figure 11 on page 28. In this example the results for all classes are above 80% correct.

The same procedures have been used for many different classification tasks in addition to crop species identification. Some of these are: determining and mapping soil types, mineral content, organic content and moisture content; geologic feature mapping; water quality mapping; identification of forested areas and tree species delineation; and delineation of geographic and land-use categories.
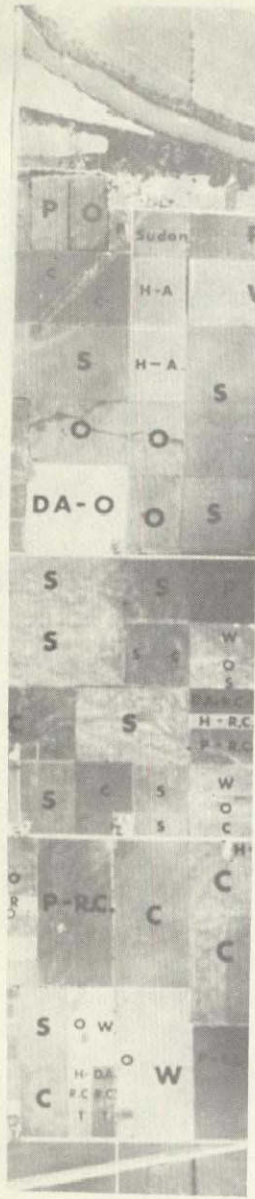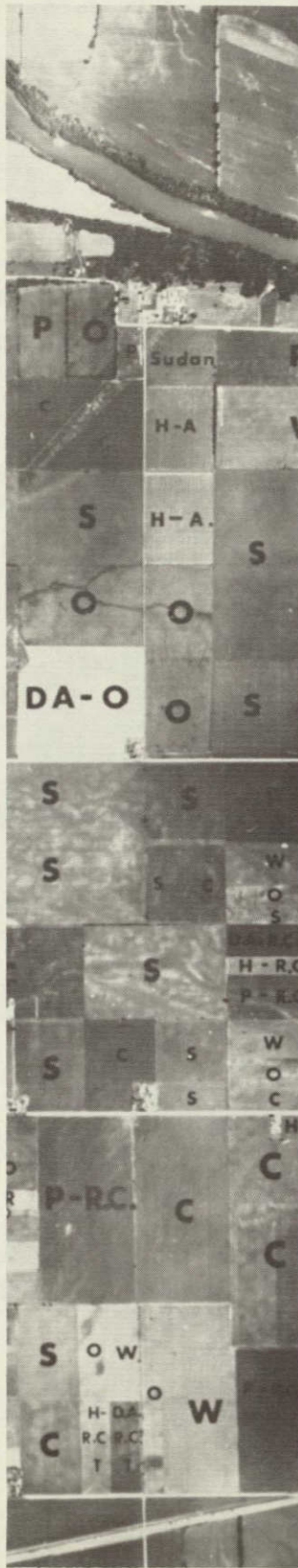
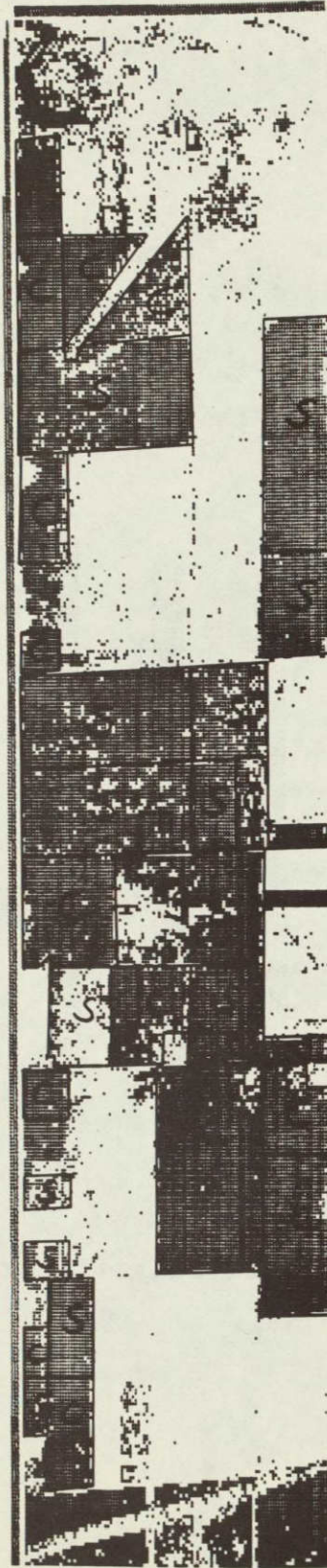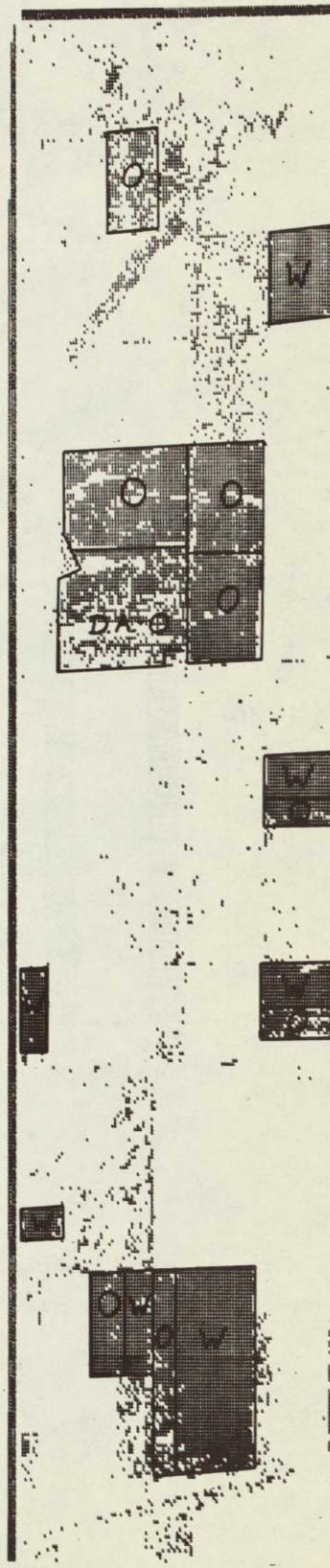Air Photo   .40-.44μm   .52-.55μm   .66-.72μm   .72-.80μm

Figure 9.   Data in Four Spectral Bands

Air Photo   Row Crops   Cereal Grains
(Corn & Soybeans)   (Wheat & Oats)

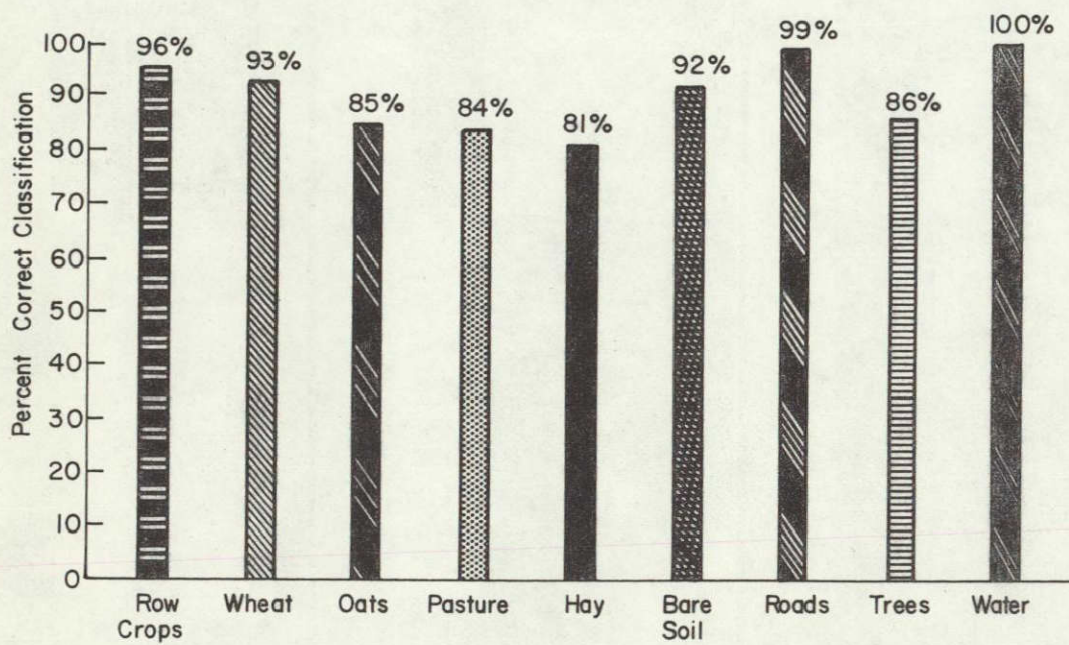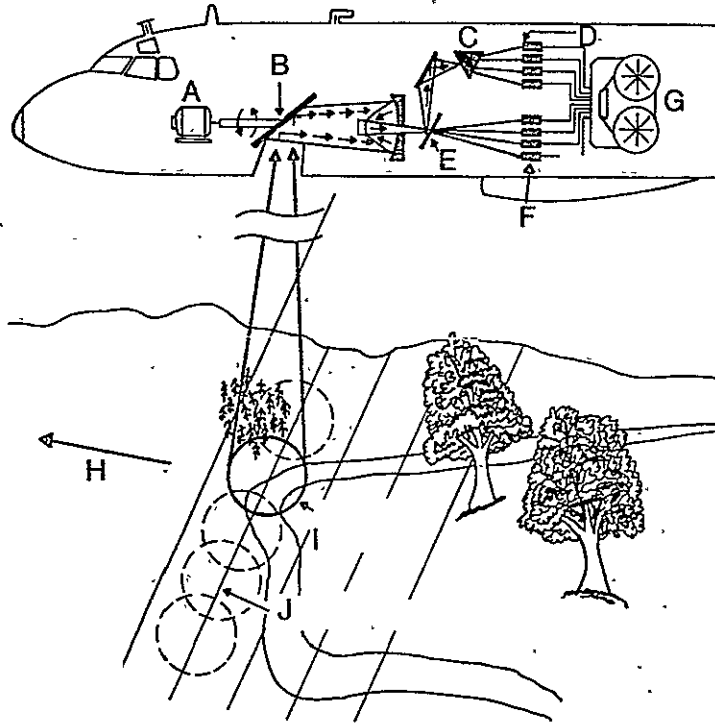Figure 10.   Spectral Pattern Recognition of Row Crops and Cereal Grains.

Figure 11. Evaluation of Classification Results for
Test Samples.

Posttest for Section II

7. Identify the letters in the diagram below and briefly describe the function of each part of the airborne multispectral scanner.

A. _____

B. _____

C. _____

D. _____

E. _____

F. _____

G. _____

H. _____

I. _____

J. _____

Posttest for Section II (cont.)

8. If you were limited to using black and white film and a board of known colors, how could you identify unknown colors in a scene?

9. Select two wavelengths λ_1 and λ_2 to serve as features for distinguishing "cover 1" from "cover 2" and plot the "cover 1" and "cover 2" points on the λ_1, λ_2 plane.
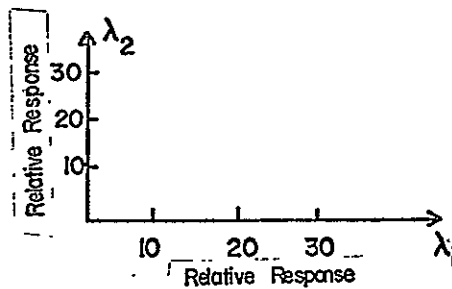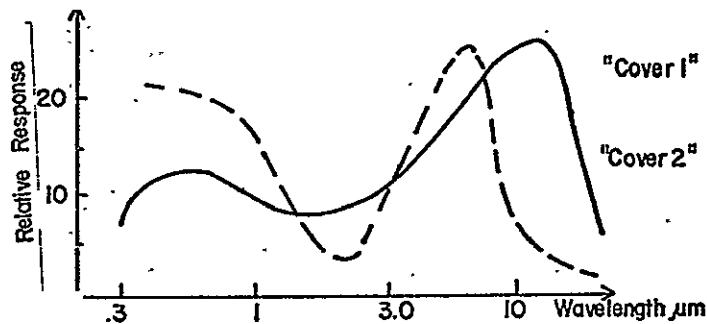
10. Identify two ways in which temporal information can be used in crop identification.

Posttest for Section II (cont.)

11. How can a multispectral scanner be used to identify crops?

12. What are some applications of multispectral analysis other than crop identification?

CHECK YOUR ANSWERS TO THE POSTTEST FOR SECTION II IN THE APPENDIX ON PAGES 60-61.

If you missed any question, you may want to go back to that portion of the booklet discussing the questions you missed.

Question 7	Page 18
Question 8	Page 21
Question 9	Page 23
Question 10	Page 23-24
Question 11	Page 25
Question 12	Page 25

If you answered the questions correctly, or have checked the text for the ones you missed, you are ready to go to Section III.

SECTION III

THE ROLE OF PATTERN RECOGNITION IN REMOTE SENSING

Objectives

Upon completion of this section, you should be able to:

13. Define "pattern recognition" and list three examples of it. They need not be in the area of remote sensing.

14. Given a block diagram of a pattern recognition system, discuss the following terms:

>   a) receptor,
>
>   b) feature vector,
>
>   c) categorizer, and
>
>   d) decision rule.

15. Name two decision rules.

16. Given a plot of training samples associated with different classes of ground cover, classify an unknown point using a specified decision rule.

17. State two conditions a class must meet in order to be useful.

18. Describe briefly the procedures by which the accuracy of multispectral scanner data classification can be verified.
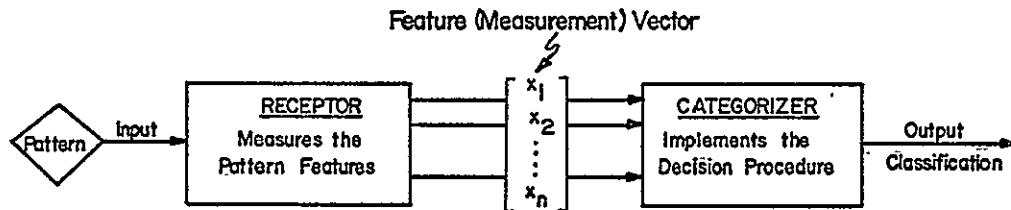
Pretest for Section III

You can check your understanding of the information contained in this section by attempting to answer the following questions.

13. Define "pattern recognition" and give three examples of ways it can be used.

14. A block diagram of a pattern recognition system is shown below. With this figure as an aid, explain the meaning of the terms receptor, feature vector, categorizer and decision rule.

Feature (Measurement) Vector

15. State two decision rules for classification of remotely sensed data.

Pretest for Section III (cont.)

16. Below are shown the training samples associated with three
different classes of ground cover as they appear in the
λ_1, λ_2 plane. U is a data point of unknown classification.
Using the classification rule: "Assign any data point U
to that class for which the distance between U and the mean
(center of gravity) of the training samples is minimum,"
classify the unknown point U.
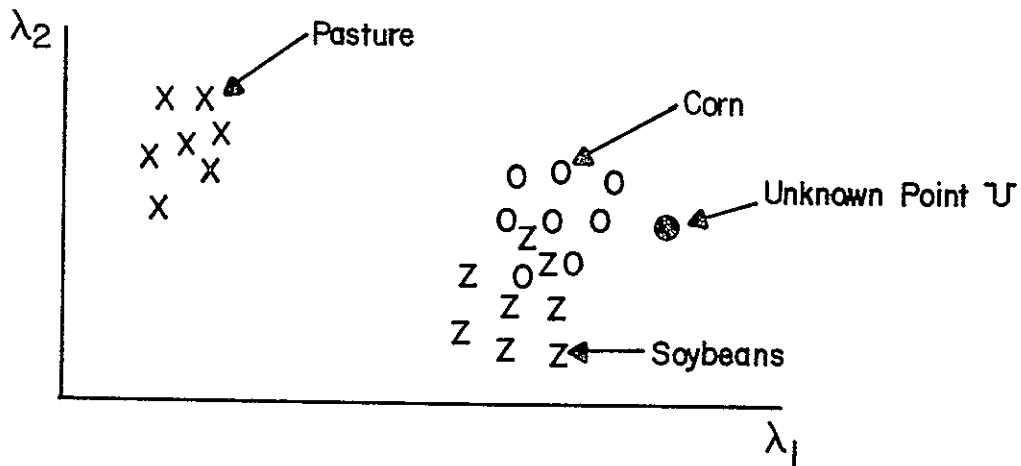
Point U is probably _____.

17. In order for classes to be useful, they must meet two
conditions. Name these two conditions.

18. Outline a procedure for determining the accuracy of remote
sensing data classification.

Check your answers to the pretest questions in the Appendix on page 62.

If you missed four or more questions, we suggest you read the entire section.

If you missed fewer than four, you might want to refer just to those parts of the section pertaining to the questions you missed.

If you did not miss any questions, you have satisfactorily completed An Introduction to Quantitative Remote Sensing.

PATTERN RECOGNITION

The basic concepts underlying the numerical processing of remotely sensed data are those of pattern recognition. Pattern recognition is the process of discerning and classifying objects or configurations. In order to further clarify what is meant by the terms pattern and pattern recognition, let's look at a number of examples.

Probably the first thing that comes to mind upon hearing the term "pattern recognition" is the problem of recognizing various geometrical patterns. Examples of this type of problem are:

1) Reading of typed, printed or handwritten text

2) Identification of a person from his photograph

3) Distinguishing man-made objects from natural objects on aerial photographs.

But pattern recognition is not limited to these cases, as evidenced by the following examples:

4) Recognition of the spoken word for various speakers, e.g. human voice-to-computer communication

5) Identification of a speaker regardless of the words spoken

6) Character or signal recognition over lines of communication, e.g. communication between computers

7) Target identification of aircraft, submarines and missiles, and distinction from decoys using radar and sonar

8) Recognition of fields of agricultural crops, their condition, and state of growth from aerial observation.

To help you understand the role of pattern recognition in remote sensing, we will introduce a conceptual picture or model of a pattern recognition system. This will be followed by a discussion of a number of examples of pattern recognition rules or algorithms that might be used in remote sensing applications. These example algorithms introduce the concept of training samples. After some elaboration on this topic, an example of the type of results that can be obtained by applying pattern recognition techniques to ERTS data is presented.

A PATTERN RECOGNITION MODEL

The problem of designing devices which recognize and classify patterns may conveniently be divided into two steps. The first step involves the problem referred to as feature extraction i.e., operations on the pattern which determine its identifying characteristics. The second step involves the decision-making procedure which classifies the pattern on the basis of the comparison of its characteristics (both similarities and differences) with those of a reference set of patterns.

Generally, in a pattern recognition problem, a number of measurable quantities are used to characterize the patterns. The choice of these quantities (called features) represents the feature extraction problem mentioned above. This problem, by no means a trivial one, is a major stumbling block to the total unification of all the applications of pattern recognition. Often the designer must use his intuition based on some prior experience to choose what seems to be a suitable set of features. On the basis of these features, studies are then undertaken to determine the best decision making strategy to employ for classifying the patterns.

The pattern recognition system is generally designed in two parts, one part called the receptor and the second called the categorizer. A simple block diagram is shown in Figure 13.
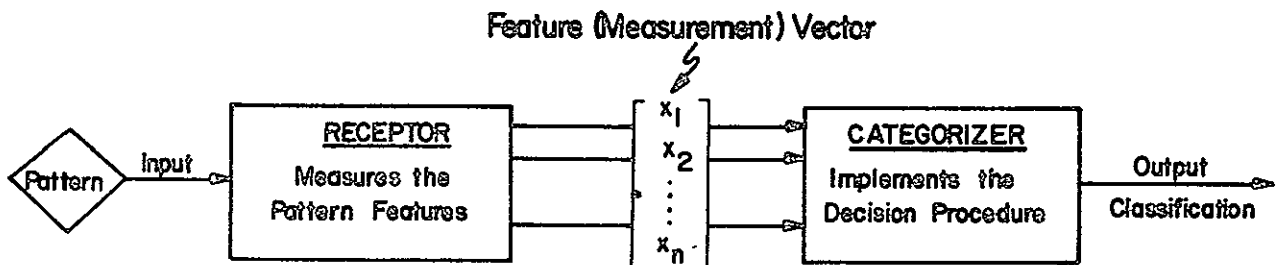
Feature (Measurement) Vector

Figure 13. Block Diagram of a Pattern Recognition System.

The input to the receptor is the physical pattern to be recognized. The receptor, using various sensors, performs the task of measuring the chosen features. The output of the receptor is a set of numbers. Each number in the set represents the numerical value of one of the features. For instance, X_1 is the numerical value of feature 1, X_2 is the numerical value of feature 2, and so on. This set of numbers is usually referred to as a vector, called the measurement vector or feature vector, whose components are the various feature measurements.

The categorizer portion of the pattern recognition device is responsible for assigning a given input pattern to a class by comparing the feature vector with a set of reference patterns, called "training patterns," for each of the possible classes. The designer constructs the categorizer to obtain the "best" possible recognition of the patterns to be classified. The term "best" in this case refers to the best performance as indicated an appropriate performance. For example, a commonly used performance measure is "minimum probability of error." The optimal design of the categorizer in a particular problem is carried out with respect to a specific set of features; to obtain the best possible system, it is also necessary to optimize over all possible sets of features.

CLASSIFICATION ALGORITHMS: EXAMPLES

Let us consider how one might devise a procedure for analyzing multispectral data. The context of the examples is the problem of classifying data obtained from an agricultural area into one of three categories - wheat, corn or alfalfa. Assume we have decided that the percentage reflectance of electromagnetic energy in certain selected regions of the spectrum will be used as features. This choice of features could have been made, for example, by examining the characteristics of the various crops as they appear from the air. The receptor portion of the pattern recognition system then measures the percentage reflectance in the selected spectral bands.

Let x_1, the first component of our feature vector, be the percentage reflectance in band one, x_2 the percentage reflectance in band two, . . ., x_n in band n, where n is the total number of features measured. The ordered set of these measurements forms our feature vector $X = (x_1, x_2, . . ., x_n)$, and on the basis of this vector the categorizer is to decide if the data from a particular ground resolution element is wheat, corn or alfalfa.

In order to represent a feature space easily on paper, let us consider the situation in which we only measure two features, i.e., reflectance in two spectral bands. Thus, the feature vector contains only two components: $X = (x_1, x_2)$. The receptor then represents each data point examined (at its input) by two numbers (at its output). To start the process we might examine five data points each selected from known fields of wheat, corn and alfalfa, then plot and label the features of each crop type (see Figure 14).

These patterns of known classification constitute the training patterns against which the patterns from unknown fields are compared. Obviously, these training patterns must be large enough and carefully enough selected that the patterns are typical of all future patterns to be classified. In practice, the selection of training patterns is crucial and requires great care and judgment. The job of the categorizer is to take the feature vector for data of unknown classification (such as point U in Figure 14) and determine its classification. That is, the job is to classify unknown data points on the basis of their representation in the feature space. The net result of applying most classification rules is to "carve up" the feature space into non-overlapping regions, one region corresponding to each of the categories. A data point of unknown classification is classified according to the region in which the data point falls. Many decision rules for making the classification have been proposed and studied in the technical literature. We will mention only three here to illustrate the approach.
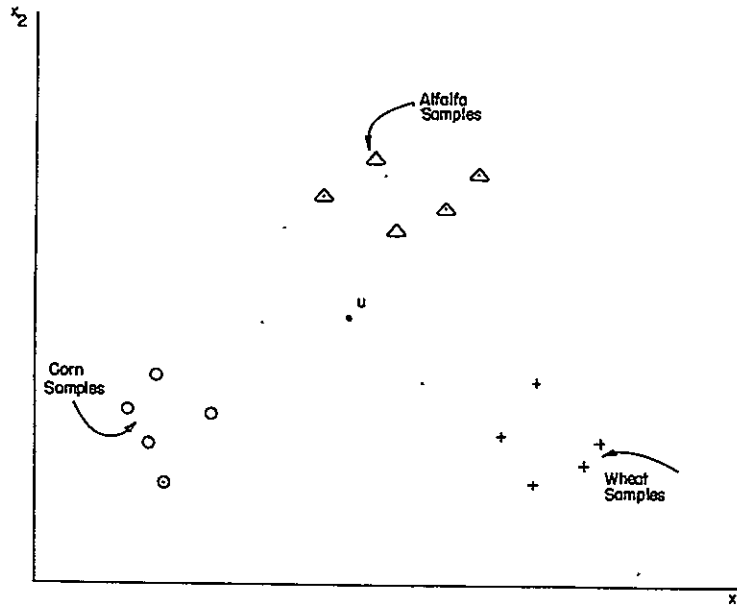
Figure 14.   Samples in Two-Dimensional Feature Space.

1. <u>Minimum Distance to the Means</u>   In this approach the
mean vector of each known class is found and represented as a
point in the feature space (see Figure 15).   The feature space
is divided into three regions for classification, as shown in
Figure 15.   Then, depending on which region of the feature space
the point of unknown classification falls in, the unknown point
is classified as alfalfa, corn or wheat (alfalfa for the example
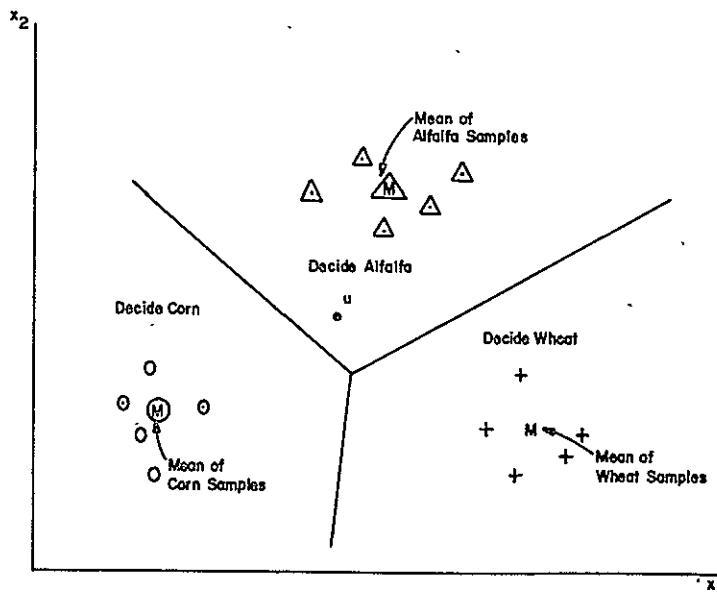U. shown).

Figure 15.   Decision Regions Based on Minimum Distance to Means.

2. Minimum Distance to the Nearest Member of a Class
According to this criterion, the unknown is classified into the
same class as that of the training point nearest it. As before,
this decision criterion divides the feature space into decision
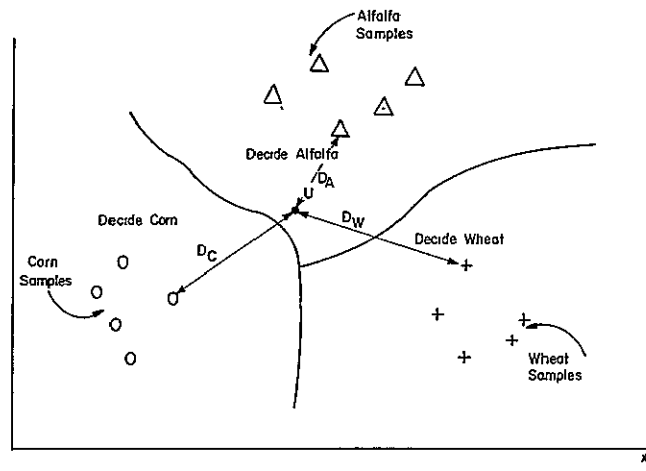regions. A graphical representation of this is shown in Figure 16.

Figure 16. Decision Regions Based on Minimum Distance to
Nearest Member of a Class. D_A is less than D_W
or D_C and hence U is classified as alfalfa.

In these first two classification schemes, the assumption
is made that the classes are sufficiently represented (charac-
terized) by the limited number of known training points. The
justification of this assumption is a problem in its own right.
As will be seen, this same assumption is employed in the third
method of classification to be discussed, but in a slightly
different way.

3. Statistical Pattern Recognition In the previous two
examples, classification was based on a geometric distance concepts
(the decision boundaries were determined based on the distance
from the unknown point to the mean value of the training samples or
the nearest training point). In this example, training data is used
to estimate statistical properties of each pattern or class. The
three classes - wheat, corn and alfalfa - are described statistic-
ally by their multivariate probability density functions.[1] Each
density function represents the probability that the associated
pattern falls in a given region of the feature space. A typical
situation is shown in Figure 17. A common statistical pattern rec-
ognition algorithm classifies an unknown data point according to
which class has the largest probability density at that point. Figure
18 shows decision regions which might be obtained by this approach.

---

[1]The adjective"multivariate"signifies that we are talking
about more than one dimension, i.e., a data vector. For simplicity
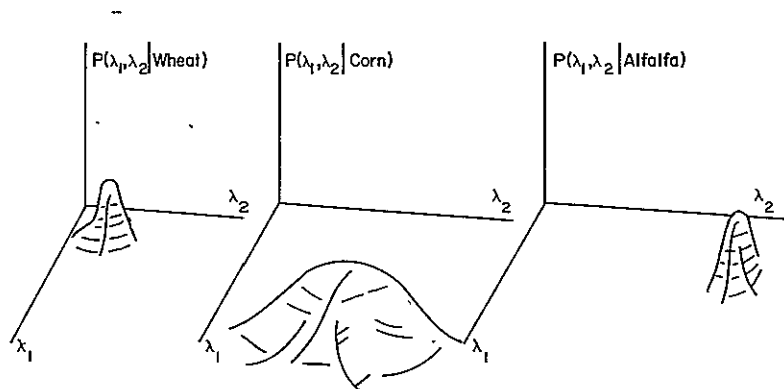we have been using two dimensions in these examples.

Figure 17. Probability Density Functions for Each Class.

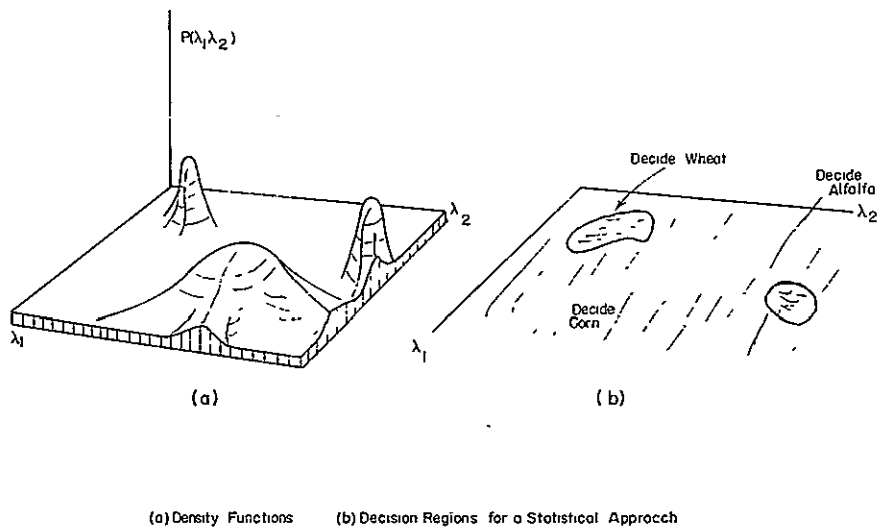(a) Density Functions   (b) Decision Regions for a Statistical Approach

Figure 18. Density Functions (a) and Decision Regions (b) for a Statistical Approach.

The knowledge of the density functions could come about in one of two ways: 1) the probability densities are actually known, 2) the probability densities are estimated from samples of each class. In the latter case, which is most common, the basic assumption again employed is that the samples are sufficient to characterize the classes. In this case this means that the sample size is large enough to accurately estimate the probability densities required.

To summarize, in each of these three classification algorithms, training samples from each class are used to establish boundaries in feature space. It is assumed that the actual patterns are adequately described by the training samples. The rule used for establishing the boundaries may be based on geometric properties (examples 1 and 2) or statistical properties (example 3) of the training samples. An unknown data point is classified by assigning it to the class associated with the region in which the data point falls. In practice, considerable judgment and experimentation may be required to select the best classification algorithm for a given pattern recognition task. A given algorithm may work well in one situation and not in another.

The situation shown in Figure 14 is a bit artificial in the sense that in practice classes are often not as obviously separable as indicated there. Figure 19 illustrates a more typical two-dimensional situation.
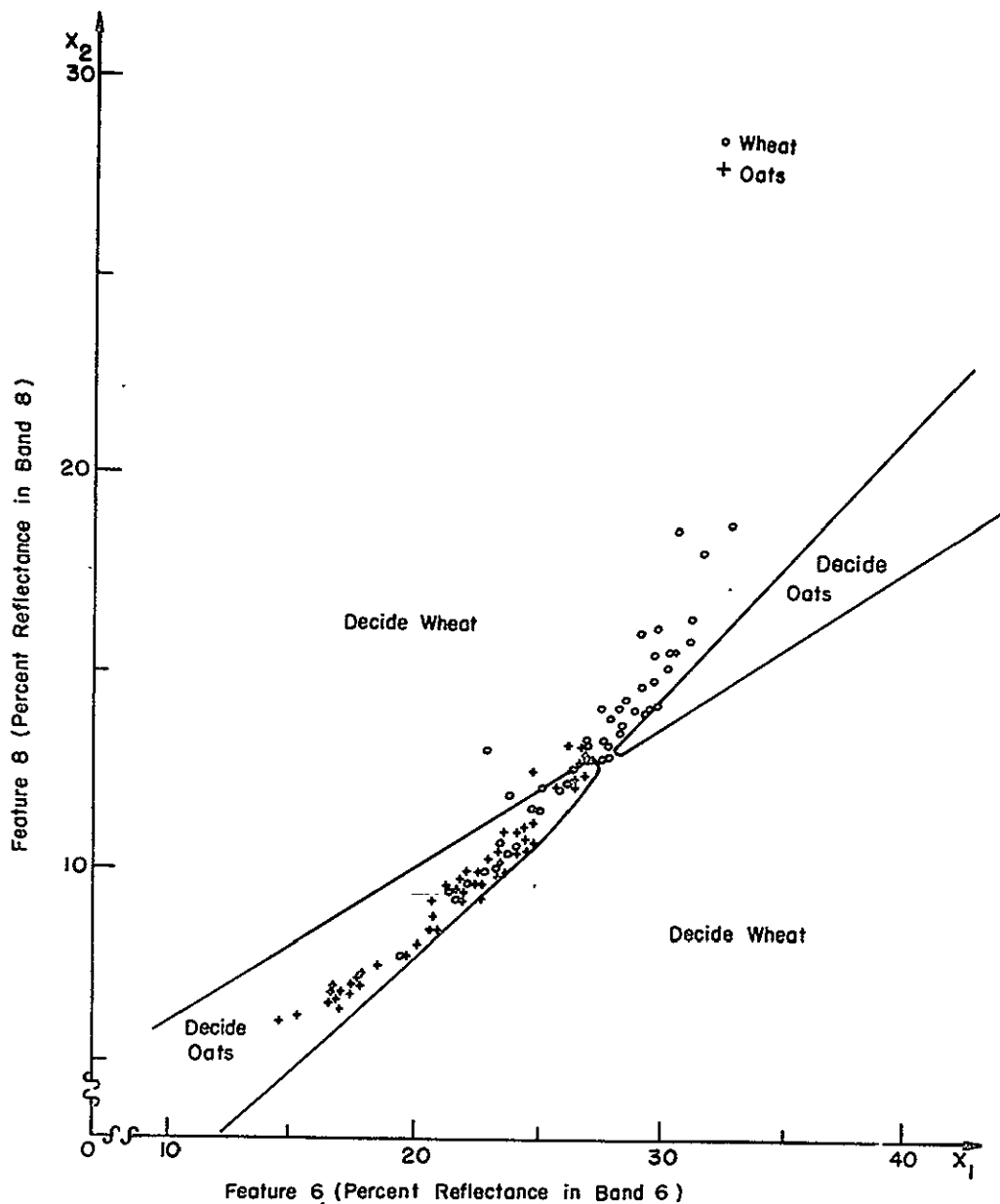
Figure 19. Overlapping Classes in Feature Space

The data for this illustration were obtained from actual laboratory measurements of the reflectance of wheat and oats in two spectral bands. Notice that the two pattern classes are somewhat overlapping in two-dimensional space, indicating that perfect classification will not be possible.

SELECTION AND DEFINITION OF CLASSES

The selection of the classes used in a particular analysis is a very important part of the process. There are two conditions that a class must meet in order to be useful. The class must be separable (spectrally discriminable) from all others and it must be of informational value. For example, it does no good to define a class called "iron ore deposits" if the spectral response which iron ore provides is not sufficiently distinct from all other earth-surface materials over which data is to be gathered. On the other hand, if no one is interested in locating the iron ore deposits within the region to be surveyed, there is no reason to define such a class. One may name classes of informational value and then check their separability, or vice versa.

Once the classes have been selected, the next step in the identification of classes is arriving at accurate definitions of them. In an agricultural survey, simply naming a class soybeans does not define it precisely enough. For example, what percent ground cover is required before a given resolution element should have its classification changed from bare soil to soybeans? What percentage of the field may be covered with weeds? The fact of the matter is that the class becomes precisely defined only by the training samples used to define it. Thus, an important step in the analysis procedure is the selection of training samples which are typical of the whole class in question.

One should also recognize that the definition of a class has a direct effect on the decision boundaries. Referring to Figure 20, you will notice that the region corresponding to the soybean class includes the point U. Defining the corn class in a slightly different manner leads to different training samples for the corn class as shown in Figure 21. As a result the decision boundaries shift and point U now falls into the corn class. We see that the location of the training sets will determine the classification of a given point. Because every point in feature space becomes associated with one of the named classes, it is important that the list of classes be exhaustive so that there is a logical class to which every point in the scene may be assigned.
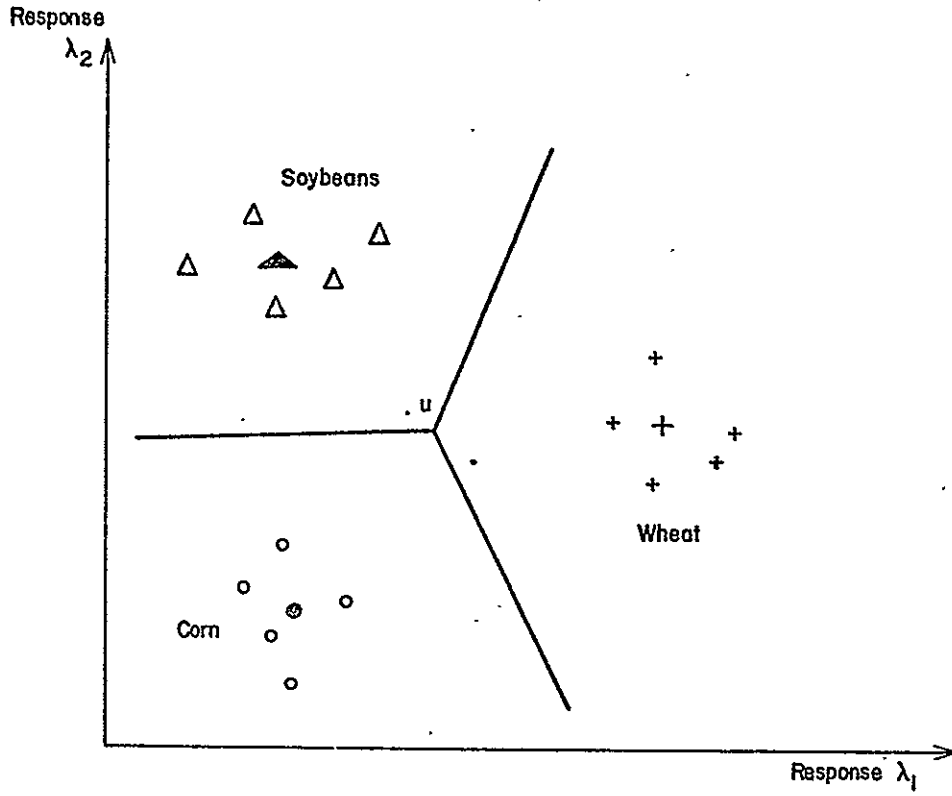
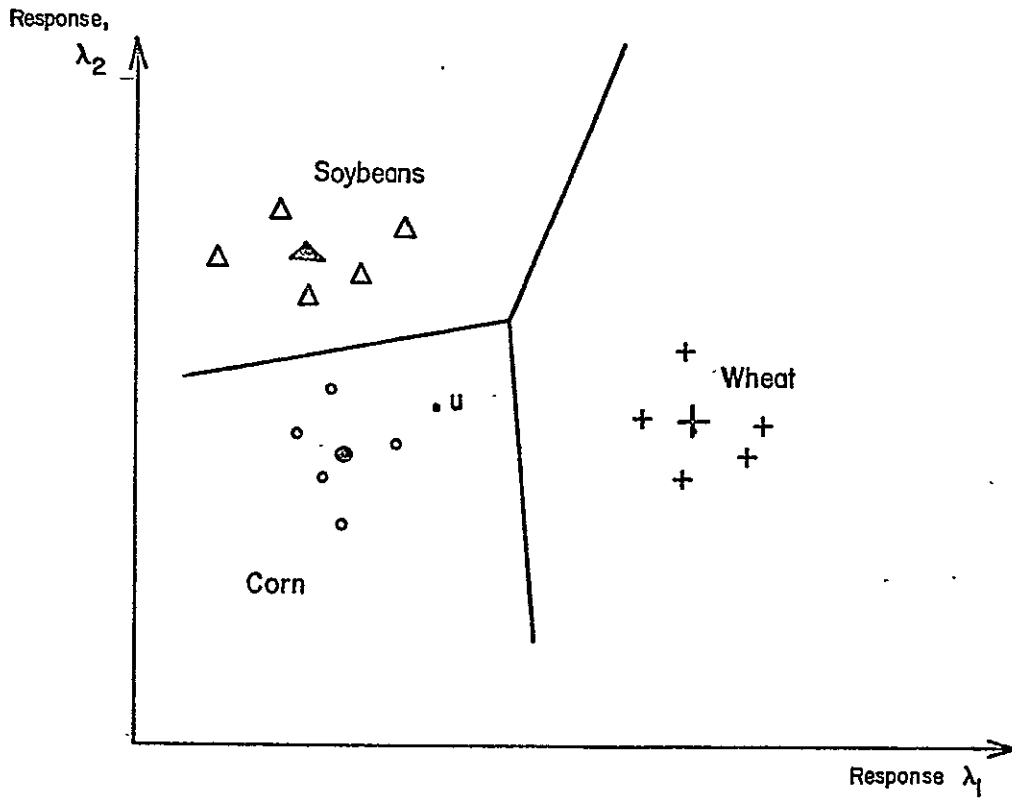Figure 20.   Minimum Distance to Means Classification

Figure 21.   Minimum Distance to Means Classification with Corn
Training Set Shifted Upward from Figure 20.

AN ERTS EXAMPLE[2]

To illustrate the quantitative analysis procedures and to show the potentials of the techniques we have been discussing, an example follows in which pattern-recognition was used to classify a relatively large amount of data. Data for this experiment were collected by the first Earth Resources Technology Satellite (ERTS-1). In addition to illustrating the pattern recognition steps described earlier, the example shows how these techniques can be used with satellite-collected data in a practical situation.

ERTS data acquired in August 1972 were used to study three agricultural counties in Northern Illinois (DeKalb, Lee and Ogle). The goal was to classify the area into three cover types: "corn," "soybeans" and "other." The "other" category was composed of subclasses which included other agricultural land uses (hay, pasture, small grain stubble) as well as towns and highways.

The selection of training samples was aided by reference data which included ground identification of crop type for over 500 fields in four parts of the three-county area. For training the classifier, a few fields (3 to 12) were chosen for each county from the available fields of known ground cover. The training data were used to estimate the multivariate probability density functions discussed in conjunction with Figure 17. The remaining fields for which ground observations were available were designated as "test fields" and were used later to evaluate the accuracy of the classification.

The DeKalb County area was classified first, using a training set with 12 corn fields, 12 soybean fields and two to three fields for each of the "other" classes. An overall test performance (total test points correctly classified/total test points classified) of 83 percent was achieved for the three classes of corn, soybeans, and "other." Quantitative results from this classification are shown in Table II. Similar classification accuracy was obtained for the other two counties.

Estimates from ERTS data of the percentage of the three-county area covered by corn and soybeans agree well with ground-based estimates by the United States Department of Agriculture, particularly for corn (see Table III).

---

[2]Text and figures adapted from LARS Information Note 030173, "Identification of Agricultural Crops by Computer Processing of ERTS MSS Data" by Marvin E. Bauer and Jan E. Cipra, 1973.

Table II.   Classification of corn, soybean and "other" test
fields by computer analysis of MSS data (DeKalb Co.,
Illinois).

Crop	No. Points	No. Points Classified As			Percent Correctly Classified
		Corn	Soybeans	"Other"	
Corn	3968	3367	357	244	85
Soybeans	1113	115	855	133	77
"Other"	295	16	50	234	79
Total	5376	3498	1262	611	

Overall Performance: 83 percent

Table III.   Comparison of USDA acreage estimates with estimates
derived from computer analysis of ERTS data (DeKalb,
Ogle, and Lee Counties, Illinois).

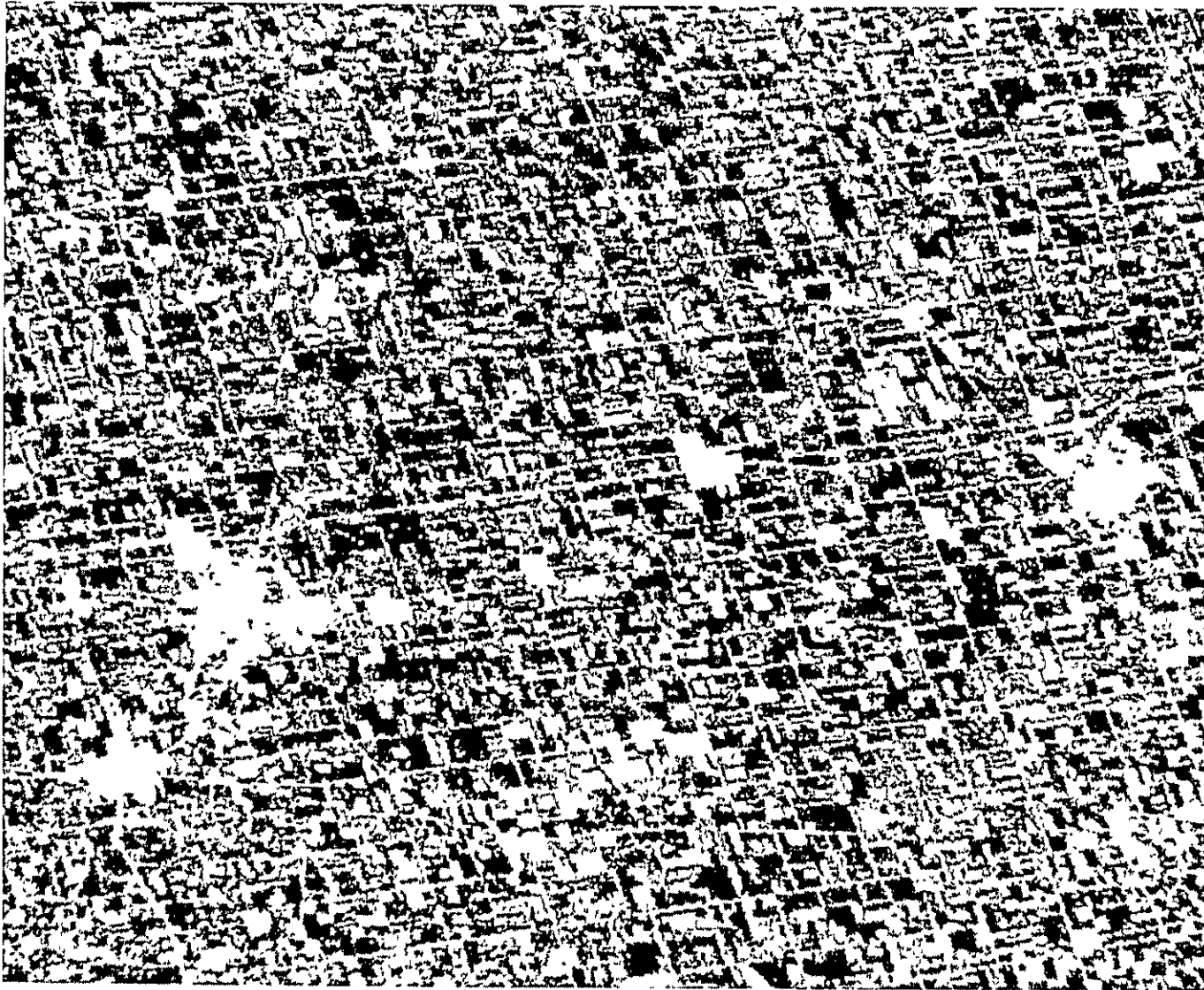|	USDA	ERTS
	(Percent of Total Area)	
Corn	40.2	39.6
Soybeans	18.0	17.8
"Other"	41.8	42.6

50

Figure 22.   Computer classification map of corn, soybeans,
             and "other" for Dekalb, Ogle, and Lee Counties,
             Illinois.   Corn-black, Soybeans-gray, and
             "other"-white.

A classification map of areas of corn, soybeans, and "other" in part of the three-county area is shown in Figure 22. As this example shows, computer-aided analysis of ERTS multi-spectral scanner data is an effective method of identifying agricultural crops over a three-county area. Other applications in which this type of analysis has proved effective include urban land use analyses, forest cover type mapping, water quality studies, and soil maps.
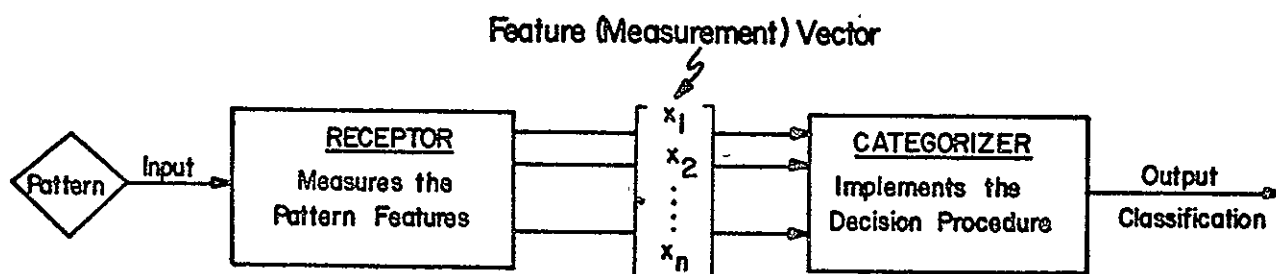
The LARS Information Note from which this example was drawn goes into considerably more detail. The interested reader should consult the original reference (see footnote on page 48).

The techniques described here are experimental in nature and still evolving. Many questions about the utility of machine-aided training procedures remain to be answered. The Earth Resources Technology Satellite provides for the first time a data base with which answers to these question can be sought.

Posttest for Section III

13. What are three examples of pattern recognition?
(They do not need to be in the area of remote sensing.)

14. Discuss the following diagram including mention of each of the components and the concept of a "decision rule."

Feature (Measurement) Vector

```
                  ┌─────────────┐        ┌ x₁ ┐        ┌─────────────┐
  ◇ Pattern ──Input──┤  RECEPTOR   ├───────┤ x₂ ├───────┤ CATEGORIZER ├──Output──►
                  │ Measures the│        │ ⋮  │        │ Implements the│  Classification
                  │Pattern Features│     └ xₙ ┘        │Decision Procedure│
                  └─────────────┘                      └─────────────┘
```
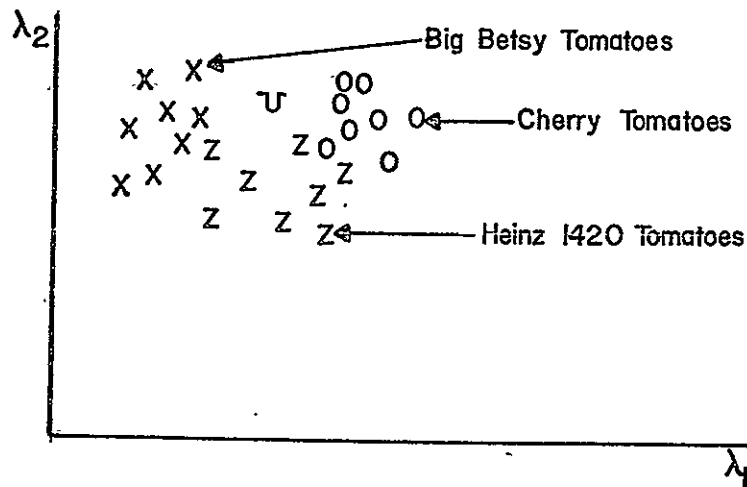
Posttest for Section III (cont.)

15. Given the location in a feature space of several training samples for each of three crops, what are two different methods by which an unknown point may be classified as one of the three crops. (Diagrams may be useful in answering the question.)

16. Below is a hypothetical pattern recognition system using two features and the classification rule: assign a data point to the class for which the distance between the point being classified and the closest training point is the smallest. A typical situation showing training points for three classes and an unknown data point U are shown below.

Using the classification rule just stated, predict which class name the pattern recognition system would assign to U. (Circle one).

A) Big Betsy Tomatoes

B) Cherry Tomatoes

C) Heinz 1420 Tomatoes

Posttest for Section III (cont.)

17. The choice of classes is a very important part of multi-spectral data analysis. There are two conditions that a class must meet in order to be useful. The class must be _____ and _____.

18. In the sample classification using ERTS-1 data, it was stated that "An overall test performance of 83 percent was achieved for the three classes of corn, soybeans and other." How was this percentage determined?

CHECK YOUR ANSWERS TO THE POSTTEST FOR SECTION III IN THE APPENDIX ON PAGE 63.

If you missed any questions, you may want to go back to that portion of the booklet discussing the questions you missed.

If you were correct in answering all the questions, or have checked the text for the ones you missed and are satisfied that you understand the key points presented, you have satisfactorily completed this introduction to remote sensing data analysis.

APPENDIX

Answers to Pretest for Section I (pages 2-3)

1. Remote sensing is the science of acquiring information about objects from measurements made at a distance, without physical contact with the objects.

2. Spectral, Spatial and Temporal (in any order)

3. B - visible, A - ultraviolet, D - thermal infrared, C - reflective infrared.

4. Remote sensing may be used to derive information about the earth's surface and the condition of its resources. Your answers may have included a discussion of how an earth survey system could be used to provide information to users or you may have described a specific application of remote sensing in an area such as agriculture, forestry, geology, land use, etc.

5.

| TYPE OF SENSING SYSTEM	CHARACTERISTICS	EXAMPLE
IMAGE-ORIENTED	Photographic film used to measure variations in electromagnetic fields. Variations related to specific surface cover by photointerpreter.	Aerial camera and photointerpreter
NUMERICALLY ORIENTED	Multispectral scanner used to measure variations in electromagnetic fields. Data analyzed with the assistance of a computer.	Multispectral scanner and digital computer

6.

TYPE OF SENSOR	ADVANTAGE	DISADVANTAGE
Photography	Spatial Resolution	Data Return
Television	Size and Weight	Spectral Range
Scanner	Spectral Range	Mechanical Complexity

Answers to Posttest for Section I (Page 11-12)

1. Remote sensing is the science of gathering information about objects, often on the surface of the earth, usually by examining the radiation reflected from or emitted by the objects. It can be done from a low-flying aircraft or spacecraft up to several hundreds of miles away and may cover very large surface areas at one time.

2. Spectral, Spatial and Temporal (in any order)

3. ultraviolet = less than 0.4 micrometers
   visible   = 0.4 to 0.7 micrometers
   infrared   = greater than 0.7 micrometers

4. Your drawing should resemble the figure below.
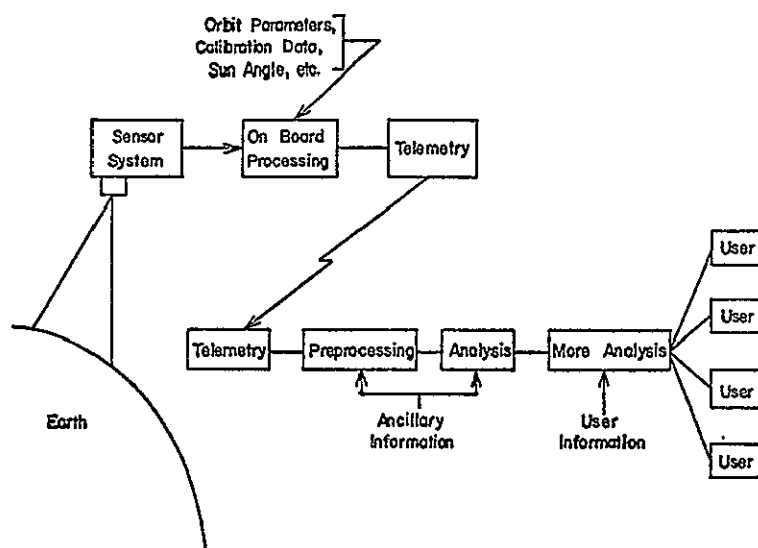
5. The image-oriented system forms an image in the primary data stream while the numerically oriented does not necessarily form an image. If it does, it is an ancillary process. An example of an image-oriented system is an aerial camera which forms an image on photographic film and a photointerpreter who, with the aid of equipment, analyzes the photograph. On the other hand, a numerically oriented system involves a multispectral scanner which gathers the data and an analyst who interprets the information with the aid of a computer.

6. The most common advantages and disadvantages of the sensors are:

	Advantage	Disadvantage
Photographic	Spatial Resolution	Data Return Necessary
Television	Size or Weight	Spectral Range Limited
Scanner	Broad Spectral Range	Mechanical Complexity

Answers to Pretest for Section II

7. Here is one possible outline:

A. Measurement of a Single Ground Resolution Element
1. Radiation (reflected or emitted) strikes scanning mirror
2. Reflected to Dichroic Grating
a. Infrared radiation dispersed by grating
b. Visible radiation reflected by grating and dispersed by prism
3. Radiation at Various Wavelengths measured by detector
4. Converted to Electrical Signals
5. Recorded on Magnetic Tape (or Transmitted to Ground)

B. Measurement of Large Area (many ground resolution elements)
1. Scanning Mirror driven by Motor
2. Mirror scans across Scan Raster Line perpendicular to direction of flight
3. Plane moves forward
4. Mirror scans adjacent Scan Raster Line usually with slight overlap

8. In order to identify colors with just black and white film, the same scene or objects must be taken with several different colored filters on the camera. By comparing the intensity of gray in each of the black and white photographs with a known standard, it is possible to determine the color of an unknown object. (Refer to Figure 5).

9.

10. Your answer should have indicated that temporal information is used to determine when during the growing season that it is easiest to discriminate between specific types of crops. You might also have mentioned that temporal information may be used to plot data in four (or more) dimensions using data gathered at different times for several types of crops. Another possibility is the measurement of change over time.

Answers to Pretest for Section II (cont.)

11. The data from several spectral bands (wavelengths) are analyzed for the area under consideration. The data from fields of known crops (training fields) are used to train the classifier, which in turn will "identify" the unknown fields. Accuracies of greater than 80% have been obtained in some cases.

12. The question asks for three applications. Here are several possibilities:
mapping soil types, geologic feature, mapping, water quality mapping, identification of forested areas, tree species delineation and land-use classification.

Answers to Posttest for Section II

7. A. Motor to turn scanning mirror

    B. Scanning Mirror to move across scan line and reflect radiation to dichroic grating

    C. Prism to separate radiation in visible portion of spectrum

    D. Visible Detectors to convert visible radiation to an electrical signal

    E. Dichroic Grating to reflect visible radiation and separate infrared radiation

    F. Infrared Detectors to convert infrared radiation to an electrical signal

    G. Tape Recorder to record electrical signals from visible and infrared detectors

    H. Direction of Flight to cover large surface area with scanning mirror

    I. Resolution Element is amount of ground scene detected at any instant

    J. Scan Line is path of scan for scanning mirror

8. Take several photos of the scene including a board of known colors using a variety of filters on the camera. Compare the intensities of gray for the objects of unknown color with the intensities of gray of the known colors.

9. Your answer will depend upon the wavelengths which you selected. If you selected λ_1 at .8 and λ_2 at 10 micrometers, your plot should appear similar to the one shown below.
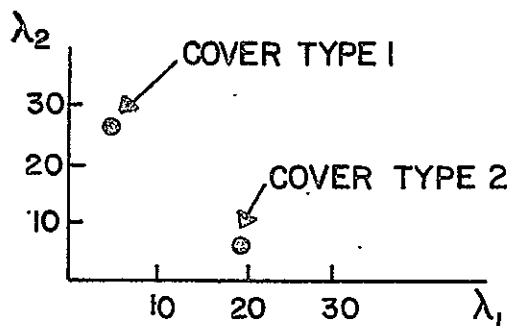
10. Temporal information can be used since most crops change color and configuration over time. Temporal information also allows the analyst to determine at what time during the growing season it is easiest to discriminate between specific types of crops. Data gathered at different times may be used to distinguish between crops.

11. A multispectral scanner can gather data from several spectral bands (different wavelengths) and the data for unknown fields can be compared with data from known fields in determining the crop types in the unknown fields.

12. Some applications of multispectral analysis other than crop identification are:

> soil type mapping
>
> geologic feature mapping
>
> water quality mapping
>
> tree species identification within forests
>
> land use mapping

Answers to Pretest for Section III

13. Pattern recognition is the procedure of distinguishing and classifying features, configurations and designs.

    Some of the examples which you might have mentioned are:
    identifying a person or object from a photograph, reading, recognizing a person from just hearing his voice, understanding the message of a speaker, identifying targets using radar and sonar, and recognizing agricultural crops from aerial photographs or multispectral scanner data.

14. The receptor receives the input from the object (pattern) being "observed" and measures the features using various sensors.

    The feature vector is composed of the values of the measurements of the pattern. In remote sensing, these are usually the intensities of the radiation at different wavelengths.

    The categorizer receives the input from the receptor in the form of the feature vector and then implements the decision procedures by comparing the input with training data and providing an output in the form of classification of the input.

    The decision rule is the criterion which determines how the data values (feature vector) will be classified.

15. Any two of the following would be an acceptable answer:

    a) Minimum Distance to the Means

    b) Minimum Distance to the Nearest Member of a Class

    c) Statistical Pattern Recognition using Probability Density Functions

16. Point U is probably corn.

17. They must be separable and have informational value.

18. Classify "test" samples of known type and use the following equation:

$$\text{Accuracy (in \%)} = \frac{\text{Test samples classified correctly}}{\text{Number of test samples classified}} \times 100$$

Answers to Posttest for Section III

13. Examples of pattern recognition include: reading, identifying people or objects in a photo, recognizing a person from hearing his voice, identifying sounds, comprehending the message from a speaker, identifying targets using radar and sonar.

14. The radiation from the pattern (input) is received by the receptor which measures the radiation (pattern features) and converts it into the feature (measurement) vector. The feature vector may be the intensity of radiation at various wavelengths. The feature vector is then received by the categorizer which uses a decision rule to classify the data. The classification is the output of the categorizer.

15. The unknown point may be classified using the a) minimum distance to the means, b) the minimum distance to the nearest member of a class or c) statistical pattern recognition using probability density functions. Any two of the three is acceptable.

16. Correct answer is C) Heinz 1420 Tomatoes

17. Separable and of informational value

18. The classifier was used on "test" samples of known crop type. The accuracy of the classification was determined by dividing the number of test samples classified correctly by the total number of test samples classified. A possible calculation might have been:

$$\text{Accuracy (in \%)} = \frac{\text{Number of test points classified correctly}}{\text{Total number of test points classified}} \times 100$$

Number of test points classified correctly =

  number of test points of corn classified as corn

+ number of test points of soybeans classified as soybeans

+ number of test points of other classified as other

= 3367 + 855 + 234 = 4456

$$\text{Accuracy (in \%)} = \frac{4456}{5376} \times 100 \approx 83\%$$