Reprinted from

Institution of Engineers, Australia' Electrical Engineering Transactions

June 1981

Characterization and Extraction of Information in Earth Observational Image Data

by D. A. Landgrebe and P. H. Swain

LARS TECHNICAL REPORT 060281

Characterization and Extraction of Information in Earth Observational Image Data

D. A. LANDGREBE and P. H. SWAIN

SUMMARY Use of imaging devices from space for gathering information about the earth has developed greatly in the past two decades. In this paper some fundamentals for information systems based upon such devices are discussed. The discussion is divided into two parts: how information is contained in remote sensing image data and how information may be extracted from such data. It is concluded that a firm foundation of fundamental principles exists as a suitable point of departure for extending our understanding of such systems and for discovering new ways to make practical application of the remote sensing technology.

1 INTRODUCTION

Use of imaging devices from space for gathering information about the earth has developed greatly in the past two decades. Yet, the realisation of the full potential for gathering earth resources information in this way is in its infancy. Rapidly developing abilities in a number of nations to construct imaging sensors, to process large data volumes and to incorporate machine intelligence into systems assures a growing capability for such information-gathering systems.

In this paper we will review some of the fundamentals upon which such information systems are based. Frequently such a discussion is given in terms of the images the system produces. In this paper we choose to discuss satellite-based earth observational systems in terms of information flow. The treatment is, therefore, divided into two parts: (1) how information is represented or coded into data via the design and operation of the sensors and other data sources; and (2) how information is extracted from such data.

In order to treat this subject adequately it is necessary that we specify the meanings we wish to associate with such words as *data, image data, image*, and *picture*. The term *data* will refer to a set (usually an ordered set) of numbers which contain, or are hypothesised to contain, information. The term *image data* is intended to imply data which, by associating either gray values or colors with the numbers, could be displayed as an image. An *image*, then, is the result of carrying out this association; it is a means for presenting image data to a human observer for purposes of manual interaction with the image data or manual analysis of it. The term *picture* is used to refer to an image when the image is primarily to be utilised based upon its (humanly interpretable) spatially structured characteristics.

To illustrate the use of these terms, Landsat multispectral scanner (MSS) data is image data. Images are frequently created from it either for the purpose of manual analysis or as an interactive tool in computer analysis of the data. Some Landsat image data (e.g., Landsat data containing pictorially identifiable features such as coastlines, cities, large rivers, major highways with a contrasting background, major geologic features, etc.) are pictorial in character and can be successfully processed using either manual or computer-implemented picture processing techniques. However, Landsat MSS data from many regions of the earth are often not very pictorial in character (e.g., desert or rangeland with few spatially structured features, agricultural land particularly where field size is small or moderate, etc.); in these cases the data may need to be processed by image processing techniques using other than pictorial characteristics.

Thus, in order to discuss earth observational data from an information system standpoint our purposes are best served by viewing earth observational data as image data rather than pictures. Though picture processing techniques, both manual and computer-implemented, are very valuable, they are only one family of a larger set of techniques to be considered. We wish to treat the subject in this broader context.

Satellite collection of earth observational data has increased greatly both in volume and in sophistication over the past two decades. Table 1 shows a list of U.S. satellites of the last twenty years and their major sensor systems. The history has been for the early satellites of the series to be designed primarily to produce pictorial products; however, there has been a steady evolution toward the production of image (and some non-image) data.

There has also been a steady evolution towards operational status for such capabilities. For example, in the case of low orbit weather satellites the early TIROS series evolved to the ESSA series and then to the ITOS/NOAA series and finally to the currently operational TIROS N design. At geostationary orbital altitudes the ATS satellites evolved through SMS to the current GOES (Geostationary Operational Environmental Satellites). The Nimbus series has served as an ongoing experimental platform for developing new capability.

As compared to weather/atmospheric observations for the U.S. alone for which more than 40 satellites have been launched in 20 years, only four satellites have been launched in the past eight years which are specifically intended to observe the solid earth surface. The Landsat satellites were the first

Submitted to The Institution of Radio and Electronics Engineers Australia on 11 March 1981.

TABLE 1

A SUMMARY OF U.S. EARTH OBSERVATIONAL SATELLITES OF THE 1960'S AND 1970'S

Satellite(s)	Launch Dates	Major Sensors
TIROS I-X	1960-65	Television, IR, auto picture transmission
Nimbus 1-7	1966-78	Television, IR scanner, image dissector, UV, data relay system, IR, temperature profile, humidity and pressure modulated radiometer, microwave radiometers
ESSA 1-9	1966-69	Television, radiometers
ATS I, III	1966-67	Spin-scan radiometer
ITOS 1	1970	Television, radiometers
NOAA 1-6	1970-79	Television, radiometers, proton monitor, temperature profiling radiometer
Landsat 1-3	1972-78	Television, multispectral scanner, data relay
SMS 1-2	1974-75	Spin-scan radiometer, data relay
GOES 1-3	1975-78	Spin-scan radiometer, data relay
HCMM	1978	Radiometers
Seasat	1978	Visible, IR, microwave radiometer, radar altimeter
TIROS-N	1978	Radiometer, vertical sounder
Magsat	1979	Magnetometers

designed for that purpose. In this case operational status is expected to be achieved within the next two years, with the multispectral scanner to be launched in 1982 aboard Landsat-D. Landsat-D is also expected to carry a more advanced experimental multispectral scanner known as Thematic Mapper.

Although both weather and land observation satellites have been used to some degree for ocean observation, Seasat was the first designed specifically for this purpose.

In addition there were the valuable collection of photographs from the manned programs of Mercury, Gemini and Apollo and the photographs and data from Skylab.

For our purposes it is helpful to visualise such earth observational systems, not in terms of space hardware as implied in table 1, but rather by concentrating upon the information flow in the entire process of collecting and disseminating information about earth surface variables. We shall focus specifically upon land observation systems.

2 HOW INFORMATION IS CONTAINED IN EARTH OBSERVATIONAL DATA

The hypothesis that spaceborne sensors can be useful implies, of course, that information is, in fact, flowing from earth to satellite altitudes by some means and may be captured by the proper kind of satellite sensor system. The key to devising effective sensor systems, and indeed to assessing the ultimate potential of such technology, is in understanding the form in which this information exists.

A fundamental principle of remote sensing is that information is conveyed in the force field and electromagnetic fields emanating from the earth's surface and, in particular, in the spatial, spectral and temporal variations of those fields (see reference 1, Ch. 1). It is necessary to be specific about this information flow in some fashion, devising a suitable engineering model of it to serve as the basis for engineering decisions and system designs. It might be possible to do this based on the information theory principles attributed originally to Shannon [2]; however, this has not been the favored approach in the field. Rather the use of intrinsic dimensionality concepts, the origins of which are found in the work of Gabor [3], have more frequently been used. The works of the 1960's and 1970's to create quantitatively based systems were centred on use of spectral variations, and the use of dimensionality as a basis for information measure arose quite naturally from the originally presumed correspondence of dimensionality with spectral bands.

Considering the entire system, consisting of the earth's surface and the atmosphere, the sensor system, the processing and information extraction system and the means for information dissemination, it is possible to classify the variables significant to information content in the system into the following five categories:

(1) Spectral sampling scheme
(2) Spatial sampling scheme
(3) Signal-to-noise ratio
(4) Ancillary data
(5) Information desired at the output

Use of the sensor system at more than one time adds a sixth category to this list:

(6) Temporal sampling scheme.

The first, second and sixth of these categories include directly the attempt to sample and represent the spectral, spatial and temporal variations of the force and electromagnetic fields from the earth's surface. Based upon the fundamentals stated above, if one could devise a sampling and representation scheme such that from the data produced by the sensor one could completely and precisely reconstruct the signals entering the sensor, then one could be assured that one has an ideal sensor in that all information is retained in passing through it. To approximate this ideal sampling and representation operation as closely as possible, with the available technology, must be the goal of every sensor designer and operator.

We shall next examine further the information content parameters of such systems by discussing each of these six categories in turn.

2.1 Spectral Sampling Scheme

Information extraction techniques, be they manual or machine-based, rely upon making identifications based upon features in the data. The scheme by which these spectral, spatial and temporal variations are sampled and represented, therefore, provides the basis for the devising of these features. Traditionally, the spectal sampling scheme has involved measuring the energy emanating from each scene element in a given set of non-overlapping spectral bands. For example, the bands used in the Landsat MSS are those from 0.5-0.6, 0.6-0.7, 0.7-0.8 and 0.8-1.1 micrometres and some pains were taken in the construction of the MSS to ensure that these bands, though immediately adjacent, had minimum overlap. The relatively simple systems now in operation utilise these spectral band measurements directly as the features. The matter of precisely which scheme of sampling the spectral distribution of energy is optimal remains an open question. Recent work indicates that there is considerable further potential for exploiting the spectral distribution of energy as a source of information [4].

2.2 Spatial Sampling Scheme

Spatial sampling refers to the scheme by which the spatial distribution of energy emanating from the scene is sampled and represented in the data. The usual method for accomplishing this is to digitally raster (sequentially sweep) the scene, preserving a measure of energy from each ground resolution element. However, the degree of adjacency of the resulting "pixels" (picture elements) can certainly be varied to permit either overlap or underlap, the sensor sensitivity distribution over a pixel may be varied and the pixel size may be varied from spectral band to spectral band, to name only a few possibilities. Furthermore, the term spatial sampling scheme is meant in a broad enough sense to include the dwell time of the sensor on a given pixel and the use of polarization measures as features. Consideration of polarization characteristics has apparently not advanced very far as yet, although some early results indicate promise [5,6]. The use of linear solid state array sensors such as that to be used on the French SPOT satellite permit a large increase in pixel dwell time, as compared to line-scanning devices and this, in turn, permits an improved signal-to-noise ratio for a given spatial and spectral resolution. The use of spatial characteristics in information extraction is now beginning to be intensively researched and later in this paper we shall introduce some of the lines of thought being pursued.

2.3 Signal-to-Noise Ratio

The third of the six categories mentioned, signal-to-noise ratio, is included to make direct accounting for the impact of the various sources of noise in the system. From an information representation standpoint a multispectral vertical view of the surface of the earth represents very complex subject matter. Signal versus noise in this application is not a very simple matter. It has been found useful in this case to use a stochastic process model for the signal as well as the noise and we are led to the following definition of signal and noise. The *signal* is all spectral, spatial and temporal variations which are information-bearing with respect to the user classes desired, whereas *noise* is all such variations which are not information-bearing. In these terms it is clear that given variations from the earth's surface may be signal for one user and noise for another user. For example, electromagnetic variations due to soil patterns might be signals for the soil scientist but noise for the crop scientist.

The original signal model used in remote sensing sprang from the concept that different earth surface covers each have a unique spectral response, sometimes referred to as a spectral signature and that deviations from this unique response represent noise. Such a signal model has served well in the early stages of the science, particularly where the data complexity is not great and where the applications are not too complex. However, as the demand by users for more detailed earth resource information has grown and as the complexity of data sets, in terms of signal-to-noise ratio and dimensionality, has grown, more detailed models have been desirable. In particular, it is recognised that a given ground cover is not characterised by a single spectral response but an ensemble of responses. For example, as human observers we call wheat in a number of different states or conditions still wheat. Indeed, some of the spectral response variation from the "norm" is information-bearing; the details of this variation may be made to assist in discriminating between different materials with similar mean spectral responses. For example, green vegetation in a naturally occurring situation may tend to have a higher variance than green vegetation which is highly cultured.

Unlike signal models, those for noise in remote sensing systems have from the beginning been thought of in terms of stochastic models. Here, too, some significant progress has been made in understanding the sources of noise, thus improving the stochastic models but there is a great deal of potential for further progress. Noise sources can conveniently be discussed in terms of two categories, those whose origin are within the scene and those generated on-board the spacecraft. Noise from the scene originates from background variability characteristics, variations in illumination or view of scene and in the atmosphere.

As compared to scene noise, sensor system noise is characterized by having its origin in the portion of the system built by man and, therefore, is to some degree under his control. Typical sources of noise of this type [7] are optically induced noise in the sensor system, detector processes such as thermal noise, modulation noise and shot noise, quantization noise and noise induced by computation. Each of these sources of noise has different characteristics and a suitable mathematical representation must be used for each. For example, thermal noise is customarily modeled as independent white gaussian noise with fixed mean and variance while shot noise is a signal-dependent noise whose magnitude increases with the magnitude of the signal level. Perhaps the greatest effort in noise study has been with regard to atmospheric effects but the relative deleterious contribution of each of these sources has never been well established.

As indicated above, from an information flow viewpoint, it is the role of the sensor system to serve as a transducer, converting information content as contained in electromagnetic variations into data and to do so with minimum information loss. Taken as a group, then, the first three categories listed above, the spectral and spatial sampling schemes and the signal-to-noise ratio, define the information entering the system via the sensor at a single data collection time. It is clear that these three are directly interrelated with one another; they define the fashion in which the energy available at the sensor is to be partitioned. For example, for a specified dwell time, simultaneously requiring very high spatial and spectral resolution leaves relatively less energy to produce high signal-to-noise ratio. Thus, one cannot expect to specify sensor parameters from all three categories independently. A determination of the information content of the data necessarily implies examining these three as a set.

2.4 Ancillary Data

In information systems based upon remote sensing, information is frequently added to the data stream either (1) directly from sources other than the spaceborne sensor or (2) in a somewhat less direct way during the information extraction process itself. An example which illustrates the former would be the registration of geographically distributed data, such as topography, previous land use, governmental boundaries, etc., directly onto the sensor image data; an example of the latter would be through the use of ancillary information such as training samples in the analysis process. The less direct case has been the traditional means by which ancillary data is incorporated into a remote sensing information system. The air photo-interpreter uses his knowledge of a region to draw conclusions about what he sees in imagery, conclusions which he would be unable to draw without this knowledge. Indeed, the more knowledge he has of the region, the more conclusions he can correctly draw about imagery from it.

The same is true of computer-based data analysis systems. Training samples are essential to pattern classification algorithms. They are the analog to the knowledge which a photo-interpreter uses in drawing conclusions or, perhaps more precisely, they are the means by which the knowledge of the scene which the human analyst possesses is entered into the system information flow. It has been found, for example, that the more data of this type which is utilised in such computer analysis of remote sensing data, the greater accuracy which can be expected and the more complex the sensor data that can be successfully utilised (see reference 1, Ch. 7).

In contrast to this, the intention of the more direct incorporation of ancillary data, such as registration of geographically distributed data with remotely sensed data, is to add directly to the information content of the data stream by increasing its dimensionality. As remote sensing data continues to become more routinely used and as the use of geo-reference data bases becomes more and more common, the importance of this type of ancillary data use will continue to grow.

2.5 Information Desired at the Output

As compared to the other categories, all of which characterize information flow *into* the system, this category is the output. In many problems it can be specified in terms of a list of discrete classes into which the image data is to be divided, e.g., crop species, urban classes and the like. In some more advanced applications the degree of membership in a class may also be required. This might be the case, for example, if crop condition, projected yield, projected crop maturity date and the like are required. In all cases it is to be noted that there is a relationship between this category and the other categories as well. Broadly stated, this relationship comes down to the more or less obvious fact that simple information output may be produced by using a simple information extraction algorithm on simple data but correspondingly more complex data and algorithms must be used if more complex information products are required.

2.6 Temporal Sampling Scheme

The sampling and representation of temporal variations in the scene may be assumed to be just as information-bearing as the spectral and spatial dimensions. It does require that the space platform must be scheduled for passes at the appropriate times, thus, perhaps complicating the mission profile and these passes must be done in relationship to the cloud cover. However, one can certainly add to the intrinsic dimensionality of the data stream by the construction of multitemporal data sets. At the same time one can add enormously to the number of classes and subclasses which exists in the data. A given (uniform) earth cover type at one time can usually change in a number of different ways with time; thus, a bitemporal observation may have many more classes present, a tritemporal still more and so on. This serves to broaden the number of types of information products which can be produced, assuming, of course, that the information extraction algorithms of adequate complexity and the needed ancillary data are available.

3. INFORMATION EXTRACTION

Given, then, that earth-observational remote sensing image data are a rich source of information about the earth and its environment, the following sections survey the methods which may be used to extract that information, i.e., convert it to forms useful for specific applications. Of principal interest are image processing methods which can be implemented on high-speed digital computers. Computer analysis of remote sensing image data was given great impetus by the launching of the first Landsat satellite by the U.S. in 1972. There are several compelling reasons for employing computerised analysis techniques, for example:

Data Volume: Modern sensor technology has produced remote sensing instruments having very high spatial resolution. The degree of detail available in data of this resolution can be used to good advantage for many earth observational applications but requires processing of extremely large quantities of raw data. High speed digital computers are needed to achieve the required processing rate.

Dimensionality: Each MSS pixel from Landsat consists of four measurements made on a ground resolution element. In addition, the satellite's periodic passes over each ground point make available a virtually unlimited number of measurements per point. Treating each measurement as a component of a vector, computer-implemented numerical methods can deal with these high-dimensional measurement vectors more quantitatively than would be possible using manually oriented techniques, thereby facilitating the utilization of the numerous pieces of data about the ground scene.

Timeliness: The value of the product resulting from remote sensing data analysis tends to decrease as the time required to produce the product increases. Computerized numerical methods can produce analysis results far more rapidly than manual image interpretation methods.

TABLE 2

IMAGE PROCESSING OPERATIONS TYPICALLY APPLIED TO EARTH-OBSERVATIONAL IMAGE DATA

Preprocessing	Information Extraction
Radiometric calibration	Supervised classification
Illumination angle correction	Clustering
Scene-to-scene registration	Texture analysis
Scene-to map registration	Change detection
Projection conversion	Image segmentation
Noise suppression	Factor analysis

Furthermore, the computer is an inherently quantitative device, which enhances the objectivity and repeatability of the analysis results which can be produced from the quantitative remote sensing data.

Although the computer plays a central role in the implementation of image analysis methods, this is not necessarily to the exclusion of human participation in the analysis. At the present state of computer science and artificial intelligence, there remain tasks which are performed most efficiently by human faculties and still other tasks which we simply do not yet know how to accomplish by machine. Our design goal is to use the human analyst and the computer as effectively as possible, admittedly, though, with a bias toward achieving analyst-assisted computer processing as opposed to computer-assisted manual processing.

Finally, it is worth emphasising that information extraction is presumed to be the objective and that this implies distillation of concentrated information from possible vast quantities of raw data. There are often many computerized preprocessing steps required prior to initiation of the actual information extraction steps (see table 2) but a characterizing aspect of the latter is that the output of an information extraction step is of greatly diminished volume compared to the input. Classification is a typical information extraction step, often mapping a high-dimensional data vector (pixel) into a single real number (the "class" of the pixel). The following discussion will outline several ways that classification can be accomplished, effectively utilizing and extracting the information in the data.

3.1 Information from the Spectral Domain: The Basic Approach to Classification

In simplest terms, the task of classifying multispectral remote sensing data can be described as follows (see reference 1, Ch. 3).

Each multispectral measurement defines a point in a Cartesian coordinate system or "measurement space". From the presumption that a specific ensemble of spectral measurements characterizes the ground cover types of interest, it follows that the points in the measurement space corresponding to a given ground cover type occupy a well-defined region of the measurement space and, furthermore, that the regions associated with different cover types are, or tend to be, distinct or "separable". Using information available from the ground scene to be classified, the measurement space may be partitioned into regions corresponding to the different ground cover types (see fig. 1). To classify each of the pixels in the scene, it is then only necessary to determine in which of the regions of the measurement space the measurement vector from a given pixel lies.

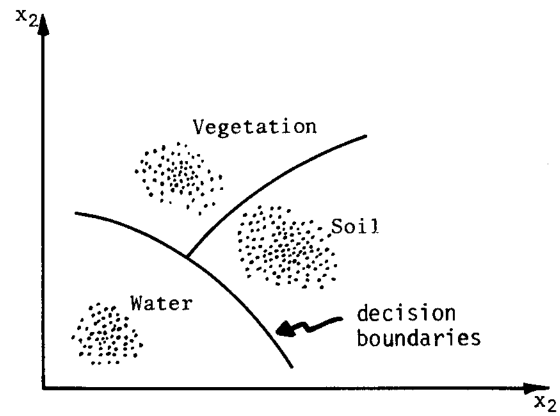


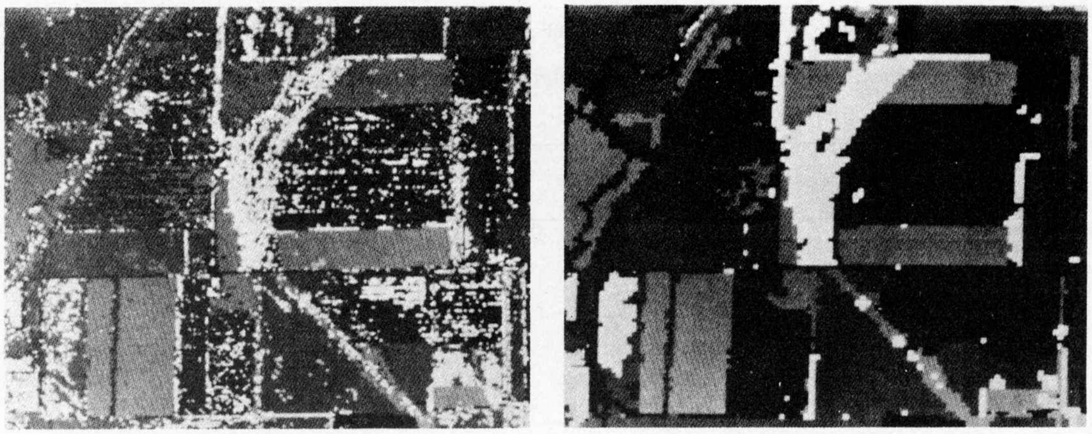
Figure 1 Measurement space.

Clearly, the step of partitioning the measurement space is crucial because the results of this step determine the quality of the ultimate classification, i.e., how accurately the pixels in the scene will be classified. This step is often called "training the classifier".

Although there are numerous ways in which this step may be carried out directly [8], an indirect approach based on statistical decision theory is commonly adopted, motivated by the fact that the regions of the measurement space occupied by the various classes of interest often overlap. Since points in the regions of overlap cannot be classified with certainty, it is desirable to adopt a strategy which performs classifications having maximum probability of being correct. A widely used strategy which accomplishes this is as follows.

Let the spectral measurements for a pixel be the components of a vector X. Assume the pixel is to be classified into one of m classes: $\{\omega_i | i=1,2,\ldots,m\}$. Then X is classified into that class ω_i for which the *posterior probability* $p(\omega_i|X)$ is maximum, i.e., ω_i is the class most probably present given the observed set of measurements.

It is convenient to use the fact that the foregoing classification rule is equivalent to (produces the same results as) a rule which classifies X into the class ω_i maximizing the product $p(X|\omega_i)\ p(\omega_i)$, where $p(X|\omega_i)$ is the probability of observing X given that the class is ω_i and $p(\omega_i)$ is the prior probability that a pixel will belong to class ω_i (the fraction of the observed universe which belongs to class ω_i). The probabilities required to implement this form of the rule are easier to obtain in practice than the posterior probabilities.



(a) (b)

Figure 2(a) An agricultural area classified using pixel classification based upon the maximum likelihood rule. The darkest regions represent wheat and the lightest regions wooded pasture. The various shades of gray indicate forest, hay, other grain crops and non-farming uses. The original data was gathered by an airborne multispectral scanner with twelve bands extending from the visible through the reflective and thermal infrared.

(b) Sample classification of the same area.

When the prior probabilities of the classes are assumed to be equal, the classification is made to maximise just $p(X|\omega_i)$. This is commonly referred to as *maximum likelihood classification*.

Under the additional assumption that the distribution of data in each class is adequately described by a multivariate normal probability density function, i.e., for class ω_i

$$p(X|\omega_i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(X-U_i)^T\Sigma_i^{-1}(X-U_i)\right]$$

where n is the number of measurements and U_i and Σ_i are, respectively, the mean vector and covariance matrix for class ω_i, the decision rule can be reduced to the following specific form (see reference 1, Ch. 3):

Classify the pixel X into the class ω_i for which the following *discriminant function* is maximum:

$$g_i(X) = \log_e p(\omega_i) - \tfrac{1}{2}\log_e|\Sigma_i| - \tfrac{1}{2}(X-U_i)^T\Sigma_i^{-1}(X-U_i)$$

Under this rule, the partitioning of the measurement space is specified by the family of quadratic surfaces defined by

$$g_i(X) - g_j(X) = 0$$
$$i,j = 1,2,\ldots,m; \quad i \neq j$$

These surfaces are completely determined by the set of prior probabilities, mean vectors, and covariance matrices for the classes. Typically all of these quantities are estimated from empirical data.

In outline form, then, multispectral remote sensing image data may be classified, based on the multispectral measurements, by (1) estimating from reference data the parameters (prior probabilities, mean vectors, covariance matrices) characterizing the classes of interest, and (2) classifying each pixel in the scene based on a statistical decision rule and the class parameters. The resulting classification may be used to generate a cover-type map of the area analysed, it may be summarised in tabular form or it may serve as an input to further information extraction processes. A typical cover map produced by such a procedure is shown in fig. 2(a).

Clustering and Unsupervised Classification

Clustering algorithms seek to partition data in the measurement space into subsets or "clusters" such that the within-cluster variability is minimum while between-cluster variability is maximum. This is completely consistent with the presumption noted earlier that an ensemble of spectral measurements characterizes each ground cover type and, as a result, the clustering algorithms, operating without external supervision, achieve partitionings of the measurement space which may be quite meaningful and useful. Each of the classes defined by the partition represents a relatively homogeneous collection of data and each of the classes may be said to be spectrally distinct from the others (assuming that the clustered data consists of spectral measurements).

The results of clustering, sometimes called "unsupervised classification", become most useful only when the resulting clusters are given labels corresponding to ground cover classes of interest. This is accomplished through use of available reference data but, compared to the training-and-classification approach ("supervised classification") described previously, relatively little reference data is required. This is the most important advantage of unsupervised classification. On the other hand, as a consequence of the small amount of information provided to the algorithm for the partitioning process, clustering is not nearly as powerful as the former technique in terms of its ability to achieve the most meaningful and useful partitioning. Clustering methods can only discriminate effectively among distinct clusters in the measurement space; if the regions of the measurement space occupied by classes of interest are likely to overlap substantially (e.g., spectrally similar crop types), unsupervised classification will not be very successful.

Specific clustering methods employed for remote sensing image data analysis are detailed in references 9 and 10.

3.2 Adding Information from the Spatial Domain

The data analysis methods described so far make very little use of the spatial characteristics of the image data. Only the measurements associated with a given pixel are used to classify the pixel. However pixels tend to be organised into "objects" in image data and the shape, size, and texture of the objects as well as the relationships of the objects to each other are also relevant characterizing information that could be used for classification. Extensive research has focused on ways of determining and quantifying spatial features in image data so that computers can make effective use of these features for classification and for other aspects of "image understanding".

Readers interested in the broad range of spatially-oriented image processing methods being explored for remote sensing applications are referred to the extensive recent literature on pattern recognition and image processing, in which these applications are invariably given their due. For present purposes, we shall concern ourselves with two interesting extensions of the statistical classification approach described above, each of which makes use of the spatial organisation of the image data to achieve improved classification performance.

Sample Classification [11,12]

Objects in a scene are often spectrally homogeneous and, if the resolution of the sensor is small compared to the average size of the objects, each object consists of several pixels. This is not to say that all of the pixels comprising the object are spectrally identical but at least any texture evidenced by the object is uniform over the extent of the object. This characteristic of image data can be used to advantage to improve both the accuracy and efficiency of the classification process. Historically this approach was first used in remote sensing to classify agricultural fields and hence became known as "per-field classification". More generally it may be referred to as *object classification* or *sample classification*. In this context, a sample is a collection of observations all assumed to be members of the same population.

Assume for the moment that it is possible to segment a remotely sensed scene into objects or samples (in the sense noted above). Let $\underline{X} = X_1, X_2, \ldots, X_s$ represent a set of s pixels in an object, a sample from a population characterized by one of the class probability functions determined through use of reference data. A *maximum likelihood sample classification strategy* is defined as follows:

$$\text{Assign } \underline{X} \text{ to class } \omega_i \text{ if}$$
$$\log p(\underline{X}|\omega_i) = \max_j \log p(\underline{X}|\omega_j)$$

where $p(\underline{X}|\omega_j)$ is the joint class-conditional probability that s such pixels would occur given that the object belonged to class ω_j.

If the sensor system operates such that adjacent pixels do not cover overlapping ground areas, it may be assumed that the pixel measurements are class-conditionally independent, i.e.,

$$p(\underline{X}|\omega_j) = \prod_{k=1}^{s} p(X_k|\omega_j).$$

This represents an important simplification because the high-dimensional joint conditional density function can be computed readily from the ordinary class-conditional densities. As in the case of the pixel classifier defined in the previous section, the class densities may again be assumed to be multivariate normal, each characterized by a mean vector and covariance matrix.

It can be shown that for a given classification problem, the probability of making an incorrect classification falls off rapidly as the size of the object (number of pixels) increases. In addition, a reduction in computation time is effected when the pixels are classified as objects, provided only that the computation time required to isolate the objects does not negate the time saved in performing the classification. The saving will be realised as long as the resolution of the sensor is small compared to the average size of the objects in the scene. The references describe an object-finding strategy that is especially compatible with sample classification under the assumption of multivariate normal class statistics. Figure 2 shows a comparison of an agricultural area classified using pixel classification and sample classification.

Contextual Classification

Rather than focusing on just the tendency of pixels to constitute uniform objects in a scene, more general "neighbor" relationships can be exploited by contextual classifiers. The context is defined to be a prespecified spatial arrangement of pixels, such as shown in fig. 3, and the predilection of certain classes of pixels to occur "in the company of" other classes is used to improve classification accuracy.

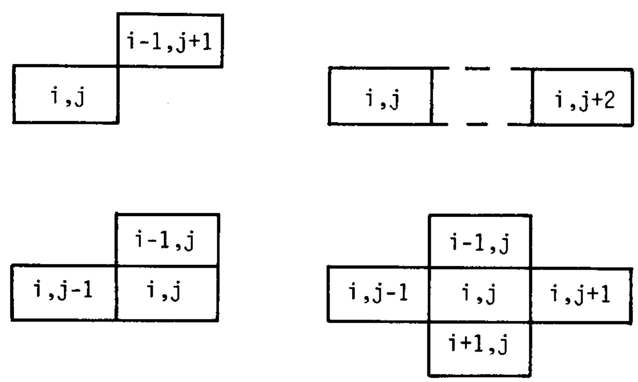


Figure 3 Examples of contextual neighborhood (the pixel labeled "i,j" is to be classified).

One form of contextual classifier is a generalisation of the "maximum posterior probability classifier", the first classifier introduced above for pixel classification. The computations required are somewhat similar to the sample classifier except that they also involve the prior probabilities of observing specified collections of classes in the prespecified spatial arrangement [13].

Another form of contextual classifier [14,15] presumes that all pixels have been initially processed to determine their probabilities of belonging to each of the classes of interest. The resulting array of class probabilities is then iteratively processed, making stepwise adjustments of the probabilities, so that they eventually are consistent with a prespecified spatial model for the scene. The final classification obtained amounts to a "compromise" between the spectral information contained in the original pixels and the spatial model presumed to be appropriate for the scene from which the data originated.

The utility of both of these contextual classifiers lies in the improvement in classification accuracy they can achieve over simpler, non-spatially oriented classifiers. This improvement, it turns out, hinges on the quality of the contextual information made available to the classifiers. The references contain numerous empirical results related to this matter.

3.3 Information from the Temporal Domain

Scene variation as a function of time has long been known to be an important source of information about the earth's surface and environment. Of course, for some applications, the change itself is of fundamental interest; examples of such applications are land-use planning and hydrological studies of snowpack variation. In other cases, the observed changes help to discriminate among ground cover types that may otherwise be spectrally indistinguishable. A notable example is the recognition of wheat amidst other crops [16].

Availability of periodic observations from Landsat has provided the opportunity to develop methods for utilising "the temporal domain". The image processing technology necessary for precisely registering multiple passes over an area has been created [17] and the development of information extraction methods for multitemporal data is being actively pursued. The most direct approach is to create "stacked" data vectors by simply concatenating sets of measurements from successive observations of a given location and to apply the same multivariate analysis methods typically applied to unitemporal data. This raises a number of issues, however. First, it is not yet known with any generality how sensitive the analysis methods are to the inevitable registration errors, however small. The effects are certainly significant in highly variable regions and at object boundaries in the data.

Second, the resulting increase in the dimensionality of the analysis task raises the complexity and the cost of the analysis and may require the prior application of dimensionality reduction methods. Several interesting methods for dimensionality reduction, motivated by this application, have appeared in the literature [18].

Finally, ordinary measurement space features are not necessarily the most information-bearing characteristics of multitemporal data. It has been found, for example, that temporal trajectories of "greenness" and "brightness", which are empirically determined linear combinations of spectral measurements, are especially well-suited for discriminating among crop species [19].

Sequential decision-making processes are also being developed for processing multitemporal data [20].

4. CONCLUSIONS AND A LOOK TO THE FUTURE

Nearly two decades of research, development and application of image processing techniques for earth observational image data have demonstrated the wealth of information contained in such data and have provided a firm foundation of fundamental principles as a basis for future advancements in earth observational information systems. These principles are rooted in the well-established techniques of photo-interpretation and photogrammetry, more recently merged with evolving principles of information theory and the communication sciences.

The future for this technology is bright from two points of view. First, operational utilisation of land observation systems is now well under way. User agencies are learning what such systems can do for them and early expectations are being confirmed that their use will be valuable, extensive and diversified. It can reasonably be assumed that once these users reach the point of repeating activities, as compared to doing them the first time, the results will prove even more significant to them. For example, as city planners draw up the second five-year plan based upon such techniques, as foresters monitor a given area for the second and third time and as crop forecasters survey an area for the second and third growing seasons, they will find they can be more precise and draw conclusions which they did not anticipate being able to draw based on their initial use of the technology.

Beyond the operational outlook for today's remote sensing technology, the potential for new technology and its benefits appears to be very large. Current sensors, while practical, do not by any means capture all of the information available from the orbital vantage point. Current information extraction techniques do not yet make available all of the information in current data systems, let alone future, more advanced ones. Also, the enormous rate of development of the computer hardware and software technologies is rapidly bringing to feasibility algorithmic approaches which, for reasons of complexity and sophistication, could not be considered previously.

4 REFERENCES

1. Swain, P.H. and Davis, S.M., (Eds.), "Remote Sensing-The Quantitative Approach", McGraw-Hill, New York (1978).

2. Shannon, C.E., "A Mathematical Theory of Communication", B.S.T.J., Vol. 27, No. 3, P.I, July 1948, p. 379-423, P.II, October 1948, p. 623-56.

3. Gabor, D., "Theory of Information", J.I.E.E., P.III, Vol. 93, No. 26, November 1946, p. 429-41.

4. Wiersma, D.J. and Landgrebe, D.A., "Analytical Design of Multispectral Sensors", Trans. I.E.E.E., Vol. GE-18, No. 2, April 1980, p. 180-9.

5. Vanderbilt, V.C., "A Model of Plant Canopy Polarization Response", Proc. 1980 Machine Processing of Remotely Sensed Data Symposium, Purdue University, (IEEE Catalog No. 80CH 1533-9 MPRSD).

6. Curran, P.J., "Remote Sensing: The Use of Polarized Visible Light (PVL) to Estimate Surface Soil Moisture", Applied Geography, Vol. 1, 1981, p. 41-53.

7. Hudson, R.D., "Infrared System Engineering", Wiley (1969).

8. Nilsson, N.J., "Learning Machines", McGraw-Hill, New York (1965).

9. Haralick, R.L. and Dinstein, I., "A Spatial Clustering Procedure for Multi-image Data", Trans. I.E.E.E., Vol. CAS-22, No. 5, May 1975, p. 440-9.

10. Hartigan, J.A., "Clustering Algorithms", Wiley, New York (1975).

11. Kettig, R.L. and Landgrebe, D.A., "Classific-

ation of Multispectral Image Data by Extraction and Classification of Homogenous Objects", Trans, I.E.E.E., Vol. GE-14, No. 1, January 1976, p. 19-26.

12. Landgrebe, D.A., "The Development of a Spectral-Spatial Classifier for Earth Observational Data", Pattern Recognition, Vol. 12, No. 3, 1980, p. 165-75.

13. Swain, P.H., Vardeman, S.B. and Tilton, J.C., "Contextual Classification of Multispectral Image Data", LARS Technical Report 011080, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, IN 47907, January 1980.

14. Richards, J.A., Landgrebe, D.A. and Swain, P.H., "Pixel Labeling by Supervised Probabilistic Relaxation", Trans. I.E.E.E., Vol. PAMI-3, No. 2, March 1981, p. 188-191.

15. Richards, J.A., Landgrebe, D.A. and Swain, P.H., "On the Accuracy of Pixel Relaxation Labeling", Trans. I.E.E.E., Vol. SMC-11 April,1981.

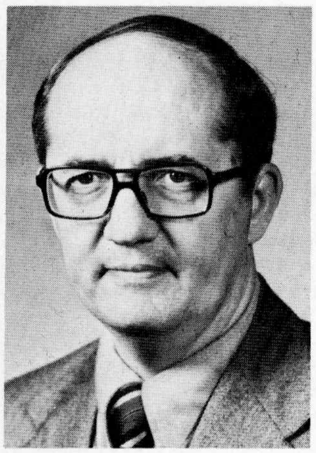
16. MacDonald, R.B. and Hall, F.G., "Global Crop Forecasting", Science, Vol. 208, No. 4445, 16 May 1980, p. 670-9.

17. Pratt, W.K., "Correlation Techniques of Image Registration", Trans. I.E.E.E., Vol. AES-10, No. 3, May 1974, p. 353-8.

18. Decell, H.P. and Guseman, L.F., "Linear Feature Selection with Applications", Pattern Recognition, Vol. 11, No. 1, 1979, p. 55-63.

19. Misra, P.N. and Wheeler, S.G., "Crop Classification with Landsat Multispectral Scanner Data", Pattern Recognition, Vol. 10, No. 1, 1978, p. 1-13.

20. Chin, R., Beaudet, P. and Argentiero, P., "An Automated Approach to the Design of Decision Tree Classifiers", Proc. Fifth International Conference on Pattern Recognition, IEEE Catalog No. 80CH1499-3, December 1980, p. 660-5.



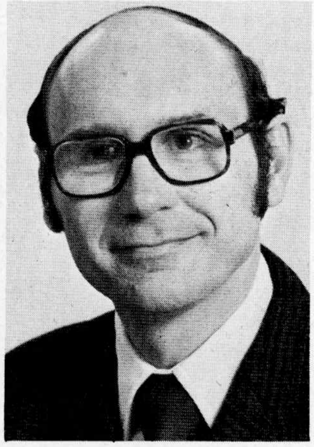
DR. DAVID A. LANDGREBE

Dr. Landgrebe is Professor of Electrical Engineering at Purdue University and Director of Purdue's Laboratory for Applications of Remote Sensing (LARS). Dr. Landgrebe holds the BSEE, MSEE and Ph.D. from Purdue University. He has held positions at the Bell Telephone Laboratories, Interstate Electronics Corporation, and the Douglas Aircraft Company. He joined the Purdue faculty in 1962 and began teaching and research activities in the area of signal representation, systems and data processing. He was named Program Leader for Data Processing Programs at LARS at the time of its organization in 1966. He became its Director in 1969.

Dr. Landgrebe is the author of a number of scientific publications in the field of remote sensing and data representation and analysis. In addition, he has lectured extensively on remote sensing both in the U.S. and abroad. He has participated in a number of National Academy of Sciences and NASA study groups and advisory committees since 1967. In 1975 he chaired the Landsat D Thematic Mapper Working Group which defined the system parameters for this instrument which is to be flown in 1982. He is currently a member of the NASA Space Science Board's Commitee on Data Management and Computation.

He serves as a member of the Editorial Board and Associate Editor of the Journal "Remote Sensing of Environment," and as a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society.

He is a Fellow of the Institute of Electrical and Electronic Engineers, a member of the American Society of Photogrammetry, the American Association for the Advancement of Science, and the American Society for Engineering Education. He is also a member of the Tau Beta Pi Association, Eta Kappa Nu, and the Society of Sigma Xi honoraries. He holds the NASA Exceptional Scientific Achievement Medal for his work in the field of machine analysis methods for remotely sensed earth resources data.



DR. PHILIP H. SWAIN

Philip H. Swain is Associate Professor of Electrical Engineering, Purdue University, and Program Leader for Data Processing and Analysis Research at Purdue's Laboratory for Applications of Remote Sensing (LARS).

Affiliated with LARS since its inception in 1966, Dr. Swain has developed methods and systems for the management and analysis of remote sensing data. He has been employed by the Philco-Ford Corporation and Burroughs Corporation and served as consultant to the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association. His research interests include theoretical and applied pattern recognition, methods of artificial intelligence, and the application of advanced computer architectures to image processing. He is co-editor and contributing author of the textbook *Remote Sensing: The Quantitative Approach* (New York: McGraw-Hill, 1978).

Dr. Swain received the B.S. degree in electrical engineering from Lehigh University in 1963 and the M.S. and Ph.D. degrees from Purdue University in 1964 and 1970, respectively. He is a member of the Pattern Recognition Society, the IEEE Computer Society and Geoscience and Remote Sensing Society, and the American Society of Photogrammetry.