Techniques for Multispectral Image Classification

Mark J. Carlotto

The Analytic Sciences Corporation
One Jacob Way
Reading, MA 01687

Over the last twenty years a variety of pattern recognition techniques for classifying terrain and cultural features using multi-spectral imagery have been developed. The purpose of this paper is to review and assess representative methods from major technique classes categorized according to the kinds of pattern models used (statistical, or heuristic), the types of information used (spectral, textural, spatial, and contextual), the manner in which they are applied to the image (i.e., to pixels or regions), and the manner in which they partition the image into classes (e.g., single step or hierarchical). An assessment of the accuracy, computational efficiency, and reliability is performed and trends in the technology are identified.

1. Introduction

A variety of techniques for classifying multi-spectral images have been developed for applications which include crop monitoring, land-use studies, geologic exploration and mapping. While the feasibility of automated techniques was demonstrated in the mid-1960s, it was not until the launch of the Landsat satellite in 1972 that widespread development and use of multispectral classification techniques began. MacDonald [1] reviews the history of automated remote sensing for agricultural applications, which has been a major driver of the technology. Other applications in urban/suburban land use analysis, water resources assessment, geologic exploration, forest and rangeland monitoring are summarized in Colwell [2]. In general, the development of automatic classification techniques has been motivated by their potential ability to process imagery data at rates and accuracies beyond those possible for photo-interpreters or image analysts.

The development of new classification techniques has also been driven by advances in sensor technology. Table 1 compares the Landsat multispectral scanner (MSS) and thematic mapper (TM) with future sensors such as the French Systeme Probatoire d'Observation de la Terre (SPOT) and NASA's Airborne Imaging Spectrometer (AIS). (References 3 and 4 review these and other planned multispectral sensors.) It is evident from this table that future multispectral classification techniques must be able to handle data having a greater number of spectral bands, higher spatial resolution and shorter repeat times. All of these factors will contribute to an increase in the volume of data that must be processed, thus dictating that future techniques be accurate and computationally efficient to justify their use.

The objective of this review paper is to assess the state-of-the-art in multispectral classifiers by examining current multispectral techniques within the context of the overall classification process. Techniques, both for classification, and to support classification (i.e., pre-processing, feature extraction, training and post-processing) are reviewed. The paper discusses current methodologies and future directions in the technology and addresses issues such as computational efficiency, accuracy, and reliability.

The organization of this paper is as follows. Section 2 provides an overview of the classification process and addresses such activities as scene and sensor pre-processing requirements, signal vs. semantic considerations in feature extraction, and issues which relate to training a classifier. Section 3 begins with a review of statistical classification theory. The limitations in statistical classifiers motivates the subsequent discussion of the use of spatial and contextual information, multi-temporal classification techniques, hierarchical classifiers, and methods for incorporating collateral information sources to improve classification accuracy and efficiency. Section 3 also describes post-processing techniques for extracting and interpreting spatial and structural information in the thematic map. Finally, Section 4 summarizes current trends in applications of multispectral classification techniques, and outlines future directions in the technology.

Table 1 Comparision of Multispectral Sensors

Sensor	Bands	Resolution (m)	Repeat
MSS	4	56x79	18 days
TM	7	30 (120 thermal)	18 days
SPOT	3	20	26 (5 off nadir)
AIS	128	4-10	—

2. Multispectral classification process

Fig. 1 depicts the typical processing flow involved in computing a thematic map from multispectral imagery. Pre-processing involves registering, restoring and normalizing multispectral images. Spectral, textural, and temporal features (or measurements) of the images are computed during feature extraction. Training involves estimating the parameters of an underlying model for each class present in a training data set, and computing decision rules for classification based on the model. The classifier applies these rules to pixels or regions, either in a single step or hierarchical process. The classifier may use spectral information alone or in conjunction with spatial, contextual, temporal, or collateral information. Following classification, analysis of the structure of and relations between groups of pixels in the thematic map may then be performed depending on the application.

The remainder of this section addresses the process up to classification (with classification and post-processing techniques discussed in Section 3). Specific attention is directed to scene and sensor requirements in pre-processing, signal and semantic considerations in feature extraction, and issues related to training a classifier.

2.1. Pre-processing requirements

Pre-processing involves activities such as registering multi-spectral images and collateral data sources (e.g., digital terrain elevation models, previously compiled maps and charts), performing corrections for sensor and atmospheric effects, and restoring data lost or degraded during transmission. The goal of pre-processing is to normalize the imagery so as to allow subsequent processes to access the data as an image of vectors, where the vector elements may contain spectral, spatial and temporal measurements of the image. Although a detailed discussion of image pre-processing techniques is beyond the scope of this paper, its importance in supporting classification cannot be over-emphasized. For example, misregistration between spectral bands, in particular, the Landsat TM thermal band, can cause linear features to be lost, and can result in classification errors along region boundaries [5]. Also, since the resolution of the thermal band is four times less than the other six bands, it must be spatially sharpened [6] prior to being merged with the other bands.

In general, multispectral data is pre-processed either to enhance the imagery for human interpretation or to normalize it for subsequent machine processing. In either case, an important step in pre-processing is image registration since subsequent feature extraction techniques require imagery that is spatially aligned across spectral bands (and across time for multi-temporal classification). It may also be necessary to register the imagery with maps and collateral data sources for use by the classifier. Juday [7] provides an overview of image registration techniques developed for Landsat. In general for flat terrain, simple geometric transformations (e.g., polynomial warping) will suffice. However, in areas having significant terrain relief, terrain elevation models may be required to achieve the required registration accuracy.

Following registration, image restoration techniques are applied to correct for:

- Sensor-dependent effects such as degraded spatial frequency response and noise in the sensor
- Scene-dependent effects such as atmospheric haze and illumination variations within the scene due to shadows and topographic relief.

Sensor-dependent corrections can, in principle, be determined at the outset and applied to all images in a similar fashion. Examples of sensor corrections are spatial frequency sharpening (e.g., thermal band sharpening [6]) and destriping. The latter correction is concerned with the removal of fixed pattern noise introduced by varying gains in the detector elements of the sensor [8]. Thermal band sharpening is motivated by the desire to classify Landsat TM imagery at the full 30 meter resolution. In cases where the resolutions of spectral bands differ (as in the TM), it is clearly unacceptable to degrade the higher resolution bands to achieve registration at the lowest resolution. Thermal band sharpening (Fig. 2) allows the lower resolution thermal band in the TM to be sharpened from 120 to 30 meters based on a local correlation assumption between bands.

Scene-dependent corrections require corresponding methods for accurately estimating the effects to be removed. For example, empirical methods for estimating the bias term in Landsat MSS due to atmospheric haze have been developed by Chavez [9]. One method involves plotting the intensities in bands 4,5, and 6 against band 7 (which is least affected by the haze) within a homogeneous area in the image. A linear regression is performed on each scatter plot to determine the intercept, which is subsequently subtracted from the respective band. In mountainous areas, interactions between the surface, local topography, and the intervening atmosphere must be taken into account. Sjoberg [10] developed a six parameter model of the imaging process where the model parameters are empirically determined from the image and auxillary data. The model is shown is shown to be useful for evaluating the effect of varying the parameters of the model on the estimated albedo map. Ongoing research is concerned with developing models capable of accurately estimating and removing terrain and atmospheric effects.

2.2. Signal vs. semantic considerations in feature extraction

While the original spectral measurements may suffice as features in many classification applications, feature extraction may be necessary:

- For data compression (i.e., to reduce the dimensionality of the measurement vector)
- To provide features or measurements which effectively discriminate between the classes of interest in the image
- For measuring physical properties of the scene
- To provide measures that are invariant to certain types of scene-dependent effects.

The first two reasons are based on signal considerations: to reduce the complexity of the classifier and to optimize its performance (i.e., to minimize the probability of error). The second two reasons are based on semantic considerations: to compute measurements which relate to physical scene properties such as vegetative content or soil moisture, and to compute a set of equivalent features that are quasi-invariant to topographic or environmental effects.

The principal components transformation [11,pp 275-283] provides a set of uncorrelated measurements ordered in terms of their variance. To reduce the dimensionality of the measurement vector, the components which account for most of the variance (generally the first two for the Landsat MSS, and the first four for the TM [12]) are retained. Although the principal components transformation optimizes the structure of feature space in a signal or statistical sense, its interpretation is scene-dependent. For a TM image containing agricultural and urban areas, the first principal component is correlated with vegetation and crops, the second with bare soil areas, and the third with urban and man-made areas. On the other hand, for images containing large bodies of water, the first principal component discriminates land and water, while the second clearly defines cultural features such as buildings and roads [13].

Tasseled cap transformations [14,15] have been used to compute physically-meaningful measures from Landsat MSS and TM imagery. Experience has shown that the data variability in the MSS four-dimensional feature space is largely confined to a single plane (i.e., the first two principal components account for most of the variance) in agricultural regions. Tasseled cap transformations rotate the feature space such that the new coordinate axes coincide with features such as "soil brightness", "greenness" and "wetness".

The tasseled cap transform thus differs from principal components transformation in that the goal is not data compression (i.e., attempting to compress as much signal information into as few dimensions as possible). Instead, it attempts to measure physical properties of the scene in an image-independent fashion. This difference is illustrated in Fig. 3 which compares the two transformations over three different scenes. The tasseled cap transform greenness measure provides a direct measurement of the amount of vegetation present in each scene. On the other hand, the first principal component must be interpreted scene-by-scene. Dave [16] has shown that the tasseled cap transform is affected significantly by viewing geometry and atmospheric composition, and thus must be used with care in hazy and mountainous areas. Jackson et al [17] discuss the effect of atmospheric path radiance and absorption on the tassled cap transform.

Use of features such as spectral band ratios [18] provides still another option during feature extraction. The use of band ratios allows effects such as illumination variations and shadows in the image to be reduced and is important in classifying terrain in mountanous regions. Ratios are also used in geologic mapping to enhance subtle differences between rock types [2, pg 1749].

2.3. Training

Training a classifier involves developing a model for each class of interest over areas in the image where "ground truth" is known. Traditionally, statistical models are employed to represent the classes of interest and may take the form of conditional probability densities or a set of parameters such as the mean vector and covariance matrix for each class. Another approach has been to use discriminant functions. Several issues are important in training:

- Selecting spectrally homogeneous regions that represent distinct classes in the image
- Ensuring that one has a valid statistical sample for each class
- Obtaining reliable estimates of the probability of error for the classifier from the training set
- Determining to what extent the signatures obtained through training are extendable to other scenes separated in space and/or time.

In general, if the classes are not homogeneous and well-separated with respect to one another, statistical classification techniques cannot be expected to perform well.

One way of assisting the image analyst in selecting homogeneous regions is by clustering the data prior to training [19]. By mapping clusters back into the image, homogeneous regions can be identified. Richardson [20] describes an interactive software package developed at ERIM for clustering and grouping pixels into training regions. The package allows a user to supervise the above process until a specified error rate is achieved. A review of the various methods used in estimating the error rate of a classifier may be found in Toussaint [21].

In estimating class statistics or probability models, one must be aware of sample size considerations (i.e., the minimum size of a training region given the number of features used). Foley [22] specifically considers how the size of the training sample biases the training set classification error rate and has shown that for the two-class problem with multi-variate Gaussian densities, the number of samples should be three times greater than the number of features. If fewer samples are used, he shows that the estimated training set error rate will much lower than the true error rate.

The selection of training regions is predicated on the availability of ground truth, and on the analyst's ability to infer ground truth from collateral data sources (maps and charts) or directly from the imagery. In some applications (e.g., in denied areas), ground truth may not be available or it may be outdated. Having to rely on the analyst to retrain the classifier on each scene will not be cost-effective in a production application. Moreover, if the analyst is not familiar enough with the sensor to be able to infer ground truth from the imagery, he may provide incorrect information based on subjective judgements. An important question then is to what extent can the spectral signatures obtained in one scene be used to classify others.

One approach to the problem of extending the usefulness of the training set to other scenes separated in time and/or space (often called signature extension) involves computing multiplicative and additive signal correction factors to map signatures in one data set to those in another [23,24]. The technique is limited to scenes which fall within the same stratum, i.e., the region in space and time which has the same types of materials, similar atmospheric and environmental effects, and is viewed under similar conditions. Within strata, such techniques have been shown to reduce the error rate by up to 50% over that possible without signature extension. Another approach proposed recently [25] is to use prior knowledge of how surface materials ought to appear in a particular scene to predict their spectral signature. This approach is discussed further in Section 3.1.

3. Multispectral classification techniques

This section begins with a brief review of the types of models used and the decision techniques employed in multispectral classifiers. A more complete discussion is provided by Swain [26]. The limitations of pixel classifiers motivates our review of techniques which use spatial, contextual, temporal, and collateral information to improve accuracy and reduce computation time. Advantages of a hierarchical approach to multispectral classification are then discussed. Examples illustrating ways in which spatial and structural information may be derived from the thematic map during post-processing are then provided.

3.1. Class models and decision rules

The simplest form of class model is to represent each class by its mean vector, the average values of its measurements computed from the training set. These measurements may pertain to spectral, spatial, or temporal properties of the image. The corresponding decision rule for classifying the vector $\underline{x}(i,j)$ is:

$$If: \; d[\underline{x}(i,j), \underline{x}_m] < d[\underline{x}(i,j), \underline{x}_n]$$
$$Then: \; y(i,j) = \omega_m$$

for $m = 1,2,..M$, $n \neq m$, where $y(i,j)$ is the output image, $d[\,]$ is the distance measure, and $\underline{x}_m$ is the mean vector for the m th class. Examples of distance (or similarity) measures include Euclidean distance, Mahalanobis distance (used if one wishes to account for class covariance), the angle between vectors (used if the classes cluster radially in the feature space) and Hamming distance (for binary-valued features).

Fig. 4 (a) illustrates a minimum (Euclidean) distance classification of a Landsat MSS image over Saudi Arabia. The thematic map (b) contains four major land cover categories: water, vegetation, coastal areas, and desert. The minimum distance image (c) shows that although all pixels are classified, certain regions (e.g., highly textured areas in the desert and shallow water) are in fact quite different from the training regions. One would thus expect a higher classification error rate in these regions. Although the minimum distance classifier is easy to implement and computationally inexpensive, its disadvantage is that there is no quantitative method for rejecting outliers (i.e., pixels which do not belong to any of the training classes).

Probability models are based on the idea that each type of material in the image gives rise to an associated distribution in the feature space. Probability models characterize the distribution of pixel values within each class in terms of conditional probability densities. For example, the Bayes (minimum probability of error) decision rule [27] is:

$$\textit{If}: P[\omega_m \mid \underline{x}(i,j)] > P[\omega_n \mid \underline{x}(i,j)]$$
$$\textit{Then}: y(i,j) = \omega_m,$$

which chooses the class having the largest *a posteriori* probability where

$$p[\omega_m \mid \underline{x}] = \frac{p(\underline{x} \mid \omega_m)P(\omega_m)}{p(\underline{x})}$$

is the *a posteriori* probability, $p(\underline{x} \mid \omega_m)$ is the conditional probability density, $P(\omega_m)$ is the relative frequency for the *m th* class, and $p(\underline{x})$ is the sum of all M class-conditional densities weighted by their relative frequencies. If the densities are multivariate normal, and the relative frequencies and covariance matrices are (or are assumed to be) equal, the maximum likelihood decision rule is equivalent to choosing the class with the minimum distance.

Probability models based on multivariate normal distributions are the most common used due to the relative ease with which they may be represented and manipulated, and the fact that the probability distributions of large homogeneous regions can often be modelled reasonably well using Gaussian distributions. (Hunt [28] shows that Gaussian behavior can be demonstrated if an image is modeled as consisting of intensity fluctuations about a local mean.) An advantage of using probability models over simple statistics is the ability (in principle) to predict and control the error rate of the classifier. For example, if the a posteriori probability is lower than a threshold, the pixel may not be classified at all. This situation was noted above for the case of the minimum distance classifier. A disadvantage is assuming a model that is not justified (e.g., making the Gaussian assumption). There is also the extra computational cost involved in computing a probability rather than just a distance.

Instead of evaluating a distance or probability, a more direct method of classification is to partition the decision space into disjoint regions at the outset. Often the decision space can be partitioned into regions using linear discriminant functions of the form

$$w_p[\underline{x}] = w_{p0} + \sum_{k=1}^{K} w_{pk} x_k = 0$$

where $\underline{x} = \{x_k\}$ and the $\{w_{pk}\}$ are weighting coefficients. Classification then involves evaluating rules such as

$$\textit{If}: w_p[\underline{x}(i,j)] > 0$$
$$\textit{Then}: y(i,j) = \omega_p.$$

In contrast to maximum likelihood or minimum distance classifiers, the above decision rules are evaluated until only one rule fires. If no rule fires, the sample is not classified. Techniques for computing discriminant functions from training data are discussed in [27].

Decision rules may also be derived from prior knowledge concerning the appearance of surface materials in the image [25]. For example, since water absorbs nearly all of the incident radiation at near and middle infrared (IR) wavelengths, if water is present in the scene, the darkest regions in those bands (e.g., TM bands 4,5, or 7) are probably water. A rule for recognizing water might then be

If: (band-4 < dark-threshold)
Then: water.

As another example, since the spectral signature of vegetation generally peaks in the near IR (since the green leaf absorbs little energy in this region with most of the incident radiation being either reflected or transmitted) a rule for recognizing vegetation might appear as

If: (band-4 > band-3) and (band-4 > band-5)
Then: vegetation.

Although the amount of water in the soil affects its overall reflectivity, the reflectance of soils generally increases in the near IR. Thus a simple rule for recognizing soil-like materials is

If: (band-4 < band-5)
Then: soil-like.

The multispectral image analysis system (MSIAS) [28] is currently being used to develop surface material classifiers based on the latter approach. Classification accuracies comparable to those achievable with a trainable classifier have been obtained in preliminary experiments.

3.2. Use of spatial information and context

Pixel classifiers consider each pixel individually without regard to either the class of neighboring pixels or the value of their measurement vectors. In addition to the computational cost involved in classifying each pixel in the image separately, the minimum achievable error rate is limited by the overlap between classes in feature space.

Landgrebe [29] reviews various methods by which local information can be used to augment spectral measurements. They include:

- Region-growing followed by sample classification
- Updating a pixel's class using contextual information
- Concatenating spectral and textural features.

The classification of multi-spectral data through the extraction and classification of homogeneous regions (ECHO) [30] is a two-step process which "grows" spectrally homogeneous regions, and classifies them on the basis of their sample distributions. ECHO uses a likelihood ratio test to decide if adjacent regions are similar based on their probability densities (assumed to be Gaussian). The technique suffers from a problem which affects most region-growers; namely, not having a sufficient number of samples in the early stages to reliably estimate the probability densities. One solution to this problem is to reduce the dimensionality of the data (i.e., reduce the number of spectral bands) so that smaller sample sizes can be tolerated. The classification error associated with the above technique is shown to be dependent on the annexation threshold. For very small thresholds, few regions form, and the classification error equals that of the pixel classifier since little or no annexation takes place. As the threshold increases, the statistical test becomes less stringent, a greater amount of inhomogeneity is tolerated and larger regions form. Classification accuracy increases as the annexation threshold increases to a point, and then decreases with improvements in accuracy on the order of 3% reported in [30].

Swain [31] and Chittineni [32] are among those who have developed methods for using contextual information in multispectral classification. Swain's method is an extension of an earlier method developed by Welch and Salter [33] for interpreting black-and-white images. Chittineni's method is based on Markov models and involves computing *a posteriori* probabilities for all classes on a pixel-by-pixel basis. (This would be a by-product of running a Bayes or minimum probability of error classifier.)

In training regions transition probabilities, i.e., the probability that a pixel at (i, j) is class ω_m given an adjacent pixel at $(i+i_p, j+j_p)$ is class ω_n, are estimated for all combinations of classes m, n and positions p within a window. These transition probabilities are then used to sequentially update the *a posteriori* probabilities in small neighborhoods. (Relaxation techniques [34] have also been used in a similar fashion.) The size of the neighborhood determines the spatial extent of the update process. After the update process has converged, the class having the largest *a posteriori* probability of occurance is assigned to each pixel. Improvements in classification accuracy of 5% and 7% using 3x3 and 5x5 windows are reported in [32].

The use of texture measures to augment spectral features has also been proposed. Wiersma and Landgrebe [35] evaluated texture measures derived from Haralick's co-occurance matrix [36]. They found that the although certain texture measures performed better in urban areas, their performance was generally no better than the ECHO classifier described above. For most scenes (with the exception of regularly textured urban and residental areas), ECHO was found to perform best at the lowest computational cost. Both context and texture methods are the most expensive to implement, with the cost proportional to the size of the window used. Moreover, the context classifier requires two levels of training: first to determine the class conditional pixel statistics, and second, to determine the transition probabilities. ECHO is relatively inexpensive, involves the same amount of training as pixel classifiers, but requires somewhat more supervision during region growing. All of the above techniques can reduce the error rate by up to 5% in most scenes over that possible with a pixel classifier.

Fig. 5 illustrates the effect of incorporating spatial information into a classification. The thematic map (a) was processed by a "mode-filter" which replaces the center pixel in a sliding window by the pixel class which occurs most frequently within the window. In the resulting image (b), isolated pixels and misclassified pixels between homogeneous regions (caused by classes mixing at the boundary) have been removed.

3.3. Multi-temporal classification

The development of multi-temporal classification techniques has been primarily motivated by the difficulty in discriminating between crop types based on the spectral signature at a single point in time. Multi-temporal techniques can be grouped into three categories: those which involve concatenating two or more multispectral data sets separated in time and analyzing the combined spectral-temporal feature vector, those which involve computing and analysing temporal trends of physical scene properties (e.g., vegetation content, soil brightness), and those which involve relating multi-temporal signatures to crop development models.

A problem in simply combining multiple data sets is that there may be significant differences in the amount of atmospheric haze and other effects which will tend to mask the subtler surface changes between the scenes. A solution to this problem is to perfrom a principal components analysis in order to separate large correlated differences such as haze from local variations (e.g., crop development). Such an approach was studied for monitoring land cover change [37]. As noted earlier, the principal components transformation is scene-dependent and thus must be interpreted on a scene-by-scene basis.

Physically-based transformations such as the tasseled cap provide measures of scene properties which can be related directly to vegetation development and soil properties [38,39]. Crist [38] uses the tasseled cap greenness and soil brightness features to discriminate between corn and soybeans. The greenness and brightness measurements made at about 10 day intervals are smoothed and features such as peak value and time computed. Corn was distinguished from soybeans by a lower and earlier green peak. Greenness and brightness are highly correlated for soybeans. Both peak later and reach higher values for soybeans than for corn.

A simpler type of classifier called the "delta classifier" [40] examines the differences between MSS bands 4 and 5, 5 and 6, and 6 and 7 for the purpose of separating wheat from other surface classes. At least three time samples are required - one each during the pre-emergence, emergence or heading stage, and brown or harvest stage of wheat. A decision rule classifier is used to classify wheat according to its similarity to a typical wheat development signature. Among the advantages of the technique are that it uses fixed decision rules and thus requires a minimal amount of training and supervision, is computationally efficient, and provides estimates of the proportion and spatial distribution of wheat that have been shown to agree closely with the ground truth over a four year period. A disadvantage is that classification results are dependent on the acquisition dates of the MSS data relative to the biological growth stage of wheat. The technique is only applicable in areas which have at least three acquisitions, one during each of the major growth phases.

Added sophistication in crop classification is possible by attempting to relate the measured multi-temporal signature to the growth state signature for the crop [41]. Initial results show the method able to classify wheat with about the same degree of accuracy as the more conventional multi-temporal classifiers described above. It also shows some potential to determine crop maturity without crop condition information in the training set.

3.4. Hierarchical classification

Previous sections have dealt exclusively with single-step classification techniques. This section discusses the advantages of decision tree or hierarchical classifiers which recursively partition multispectral images into subclasses [42].

Surface material classes naturally structure themselves in a hierarchical fashion. For example, it is natural to think of an aerial image as being composed of major surface classes such as water, vegetation, and soil-like materials, of vegetation having sub-classes crops and forests, and so forth. A decision tree structure applied to image data, recursively classifies each pixel, refining its classification upon each trial. The root node of the decision tree is the whole image with intermediate nodes representing increasingly refined surface material subclasses.

In general, decision tree classifiers are useful in domains whose feature spaces have complex decision boundaries since each classifier need only concern itself with the classes present at its node and does not have to be capable of classifying the entire image. The implication here is that the complexity of the classification rules is greatly reduced, and the rules can be developed in a systematic fashion [43]. Classification efficiency is also increased as each pixel is not tested for membership in every class. This is an important consideration given the potentially significant size of multi-spectral data sets. Finally, only the features that are needed for classification at a given node are used so that potential sample size problems which can result in using the full feature vector can be avoided [44].

We have recently implemented a decision tree classifier which uses a cooperative hierarchy of expert system modules (Fig. 6). As shown in the figure, initial modules are responsible for partitioning the image into surface material classes, with subsequent modules designed to identify objects within the image. Fig. 7 illustrates the hierarchical classification process for the Landsat TM image shown in (a). Major classes (b) are water (black), vegetation (grey), and soil-like materials (white). Soil-like materials are further divided into plowed fields (light grey), concrete (white), and other (dark grey) in (c). Vegetation is subclassified into crops (white) and less dense vegetation (grey). (Examples of the kinds of decision rules used within the classifier are contained in Section 3.1.)

3.5. Incorporating collateral data sources

In many circumstances collateral data sources (e.g., maps, charts, elevation data) and prior knowledge concerning the expected classes and their relative proportions may be available. A simple way to incorporate prior knowledge concerning the relative frequency of occurance for each class is as prior probabilities in a maximum likelihood classifier [45]. Relative proportions may be derived from historical data, old maps, and field reports.

Map information may be used to guide the interpretation of aerial images by focusing the attention of the image interpretation (or understanding) system to the parts of the image of interest (e.g., roads, urban areas, crop fields) [46]. Map guided segmentation techniques [47] have also been proposed for extracting specific types of regions (e.g., dark compact areas).

Relaxation techniques have been used to combine elevation, slope, and angle preferences of various tree species to improve classification accuracy [48,49]. In probabilistic relaxation labeling techniques, each pixel is assigned a vector of probabilties, one for each class. The likelihood of a pixel at (i,j) being a member of class ω_n given a neighboring pixel at $(i+i_p, j+j_p)$ is a member of class ω_m for all m, n and p is represented by a matrix of compatability values. In this case, the compatability matrix contains the context distribution which describes the likelihood that certain tree classes are spatially compatable, as well as the likelihood of a class of tree occuring as a function of elevation. By combining elevation information into the classification in this way an increase in classification accuracy from 68% to 81% was obtained.

3.6. Post-processing

The previous sections addressed techniques for computing thematic maps from one or more multispectral images. Depending on the application, post-processing may be required to count the number of pixels in each class (e.g., to estimate the number of acres of a crop) or to extract objects of interest (e.g., cartographic features) from the thematic map. This section describes how spatial and structural information can be computed from the thematic map and used to identify semantically-significant features such as road networks, urban areas and major agricultural regions within the scene.

3.6.1. Extracting spatial and structural information

Up to this point in the classification process we have made use of information that exists at the pixel level. After classification, pixels with the same class may be grouped into regions and the properties of those regions used to identify objects in the image. Pavlidis [50] reviews techniques for labeling connected regions. Useful properties of connected regions such as area, perimeter, compactness, elongatedness (structural properties), orientation, location, and collinear/nearest neighbors (spatial properties) may be computed from the image of labeled connected regions as described in Winston [51]. The resulting symbolic descriptions are not unlike those used in image understanding systems [52,53].

3.6.2. Binary image processing

Prior to extracting spatial and structural information from the thematic map, it may be necessary to remove isolated pixels, to separate compact and thin regions, and to merge adjacent regions. These operations are implemented by binary image operators such as shrinking (replacing the center pixel in a sliding window by a zero if there is a zero within the window) and expanding (replacing the center pixel in a sliding window by a one if there is a one within the window). In the example in Fig. 8, regions having soil-like properties that are bright in the visible (e.g., concrete and silt) are shown in (a). This image is processed with a shrink operator to eliminate small and thin regions, followed by an expand operator to restore the regions which remained after shrinking to their original size. By subtracting the resulting image from the original, the small thin regions which were eliminated by the shrink/expand operation are obtained. Long thin regions and compact regions are shown in (b).

3.6.3. Grouping

Regions in the thematic map may be grouped in a bottom-up fashion, using information derived from the image such as spatial proximity and collinearity, and top-down based on prior knowledge concerning the physical composition and structure of objects. Grouping is a key step in the image understanding process since it organizes the image into perceptually significant objects [54]. Grouping techniques for recognizing cultural features in aerial photographs are discussed in [55]. Although road networks are visible in TM imagery, only partial segments are generally extracted by the classifier due to the limited spatial resolution of the sensor. As a result, some form of line-growing and linking may have to be performed as illustrated in Fig. 9. (In this example, line segments are linked if they are within R segment lengths of each other and if the orientation of the link is within α degrees of the orientation of the segments.)

3.6.4. Object identification

Objects may be composed of one region or more depending on the resolution of the imagery, as well as on the objects themselves. Assuming spatial resolutions of the order of Landsat TM and SPOT, many types of natural and man-made features can be identified on the basis of region properties. In order to identify large agricultural areas in a scene (Fig. 10), candidate regions such as crop and plowed fields (a) are aggregated based on their relative proximity to one another. Aggregates whose area is greater than a minimum selection threshold are shown in (b). More complicated identification rules which take into account the shape of and relations between regions and objects within a scene are discussed in [52]. Nagao and Matsuyama describe an system for interpreting color IR aerial photographs using this approach [53].

4. Summary

This paper reviewed and assessed various methods for classifying multispectral images. The methods were categorized in terms of the kinds of pattern models used (statistical, or heuristic), the types of information used (spectral, temporal, spatial, and contextual), the manner in which they are applied to the image (i.e., to pixels or regions) and the manner in which they partition the image into classes (e.g., single step or hierarchical). An assessment of the accuracy, efficiency and reliability of representative techniques was performed for representative techiques in each category.

Several trends in the technology were evident throughout the course of the review. Due to the difficulty in discriminating between certain crops based on their spectral signature alone, the trend in agricultural applications (e.g., crop classification) is towards the development of multi-temporal techniques to exploit imagery acquired at regular time intervals. To better resolve surface materials spectrally and spatially, the trend in geologic exploration and mapping applications is towards developing new techniques for exploiting imagery acquired by sensors with higher spatial resolutions and more spectral bands. In general, current interest is in developing techniques which effectively combine multiple imagery sources, use spatial, textural, and temporal information derived from the imagery and incorporate collateral information sources to improve classification accuracy.

In the future, technologies such as artificial intelligence and image understanding will play a major role in expanding the capabilities of multispectral classifiers. Expert systems may be used to supervise and monitor the operation of the classifier much like the image analyst does today and image understanding techniques may extend the ability of classifiers to identify objects in, and detect differences between multispectral images.

References

[1] R.B. MacDonald, "A summary of the history of the development of automated remote sensing for agricultural applications," *IEEE Trans. Geoscience and Remote Sensing,* Vol 22, Nov., 1984.

[2] R.N. Colwell (Editor), *Manual of Photogrammetry,* American Society of Photogrammetry, Falls Church, VA., 1983.

[3] R.A. Holmes, "Advanced sensor systems: thematic mapper and beyond," *1982 Machine Processing of Remotely Sensed Data Symposium.*

[4] P.N. Slater, "Survey of multispectral imaging systems for earth observations," *Remote Sensing of Environment,* Vol. 17, pp 85-102, 1985.

[5] R.C. Wrigley, et. al., "Thematic mapper image quality: registration, noise, and resolution," *IEEE Trans. Geoscience and Remote Sensing,* Vol 22, May, 1984.

[6] V.T. Tom, M.J. Carlotto, and D.K. Scholten, "Spatial resolution improvement of TM thermal band data," *Proc. SPIE,* San Diego, CA., Aug. 20-24, 1984.

[7] R.D. Juday, "Inter-image registration," *Proc. NASA Workshop on Registration and Rectification,* JPL Pub. 82-23, June, 1982.

[8] B.K.P. Horn and R.J. Woodham, "Destriping satellite images," *A.I. Lab Memo No. 467,* Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

[9] P.S. Chavez, "Atmospheric, solar, and MTF corrections for ERTS digital imagery," *Proc. Amer. Soc. Photogrammetry,* 1975.

[10] R.W. Sjoberg, "Atmospheric effects in satellite imaging of mountainous terrain," *Tech. Report No. 688,* Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

[11] J.T. Tou and R.C. Gonzalez, *Pattern Recognition Principles,* Addison-Wesley, 1974.

[12] P.E. Anuta, et. al., "Landsat-4 MSS and thematic mapper data quality and information content analysis," *IEEE Trans. Geoscience and Remote Sensing,* Vol 22, May, 1984.

[13] R. Bernstein, et. al., "Analysis and processing of Landsat-4 sensor data using advanced image processing techniques and technologies," *IEEE Trans. Geoscience and Remote Sensing,* Vol 22, May, 1984.

[14] R.J. Kauth and G.S. Thomas, "The tasseled cap - A graphical description of the spectral-temporal development of agricultural crops as seen by Landsat," *Symposium on Machine Porcessing of Remotely Sensed Data,* Purdue Univ., 1976.

[15] E.P. Crist and R.C. Cicone, "A physically-based transformation of thematic mapper data - The TM tasseled cap," *IEEE Trans. Geoscience and Remote Sensing*, Vol 22, May, 1984.

[16] J.V. Dave, "Influence of illumination and viewing geometry and atmospheric composition on the tasseled cap transformation of Landsat MSS data," *Remote Sensing of Environment*, Vol. 11, 1981.

[17] R.D. Jackson, P.N. Slater and P.J. Pinter, "Adjusting the tasseled-cap brightness and greenness factors for atmospheric path radiance and absorption on a pixel by pixel basis," *Int. J. Remote Sensing*, Vol. 4, No. 2, pp. 313-323.

[18] G.W. Wechsung and J.R. Breedlove, "Some techniques for digital processing, display, and interpretation of ratio images in multispectral remote sensing," *Proc. SPIE*, Vol. 119, San Diego, CA., 1977.

[19] G. Nagy and J. Tolaba, "Nonsupervised crop classification through airborne multispectral observations," *IBM Jornal. Res. Develop.*, March, 1972.

[20] W. Richardson, A. Pentland, R. Crane and H. Horwitz, "Number of signatures necessary for accurate classification," *LARS Symposium*, Purdue Univ., 1976.

[21] G.T. Toussaint, "Bibliography on estimation of misclassification," *IEEE Trans. Information Theory*, Vol. 20, July, 1974.

[22] D.H. Foley, "Consideration of sample and feature size," *IEEE Trans. Information Theory*, Vol. 18, Sept., 1972.

[23] R.G. Henderson, "Signature extension using the MASC algorithm," *IEEE Trans. Geoscience and Remote Sensing*, Vol 14, Jan., 1976.

[24] P.F. Lambeck and D.P. Rice, "Signature extension through the application of cluster matching algorithms to determine appropriate signature transformations," *LARS Symposium*, Purdue Univ., 1976.

[25] M.J. Carlotto, V.T. Tom, P.W. Baim, and R.A. Upton, "Knowledge-based multispectral image classification," *Proceedings SPIE*, Vol. 504, 1984.

[26] P.H. Swain, "Fundamentals of pattern recognition in remote sensing," in *Remote Sensing: The Quantitative Approach*, Swain and Davis (eds), McGraw-Hill, 1978.

[27] R.O. Duda and P.E. Hart, "Pattern Classification and Scene Analysis," Wiley-Interscience, 1973.

[28] R.D. Ferrante, M.J. Carlotto, J.M. Pomarede and P.W. Baim, "Multispectral image analysis system," *First International Conference on Artificial Intelligence Applications*, Denver, 1984.

[29] D.A. Landgrebe, "Analysis technology for land remote sensing," *Proc. IEEE*, Vol. 69, No. 5, May, 1981.

[30] R.L. Kettig and D.A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geoscience Electronics*, Vol. 14, No. 1, Jan., 1976.

[31] P.H. Swain, H.J. Seigel and B.W. Smith, "Contextual classification of multispectral remote sensing data using a multiprocessor system", *IEEE Trans. Geoscience Electronics*, Vol. 18, No. 2, April, 1980.

[32] C.B. Chittineni, "Utilization of spectral-spatial information in the classification of imagery data," *Computer Graphics and Image Processing*, Vol. 16, pp. 305-340, 1981.

[33] J.R. Welch and K.G. Salter, "A context algorithm for pattern recognition and image interpretation", *IEEE Trans. SMC*, Vol. 1, No. 1, Jan., 1971.

[34] S. Peleg, "A new probabilistic relaxation scheme," *IEEE Trans. PAMI*, Vol. 2, No. 4, July, 1980.

[35] D.J. Wiersma and D.A. Landgrebe, "The use of spatial characteristics for the improvement of multispectral classification of remotely sensed data," *Machine Processing of Remotely Sensed Data Symposium*, Perdue Univ., 1976.

[36] R.M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," *IEEE Trans. SMC*, Vol. 3, No. 6, Nov., 1973.

[37] G.F. Byrne, P.F. Crapper and K.K. Mayo, "Monitoring land-cover change by principal component analysis of multitemporal Landsat data," *Remote Sensing of Environment*, Vol. 10, pp 175-184, 1980.

[38] E.P. Crist, "Cultural and environmental influences on temporal-spectral development patterns of corn and soybeans," *Machine Processing of Remotely Sensed Data Symposium*, Perdue Univ., 1982.

[39] J.B. Odenweller and K.I. Johnson, "Crop identification using Landsat temporal-spectral profiles," *Machine Processing of Remotely Sensed Data Symposium*, Perdue Univ., 1982.

[40] J.L. Engvall, J.D. Tubbs and Q.A. Holmes, "Pattern recognition of Landsat data based upon temporal trend analysis," *Remote Sensing of Environment*, Vol. 6, pp 303-314, 1977.

[41] C.A. Hlavka, R.M. Haralick, S.M. Carlyle and R. Yokoyama, "The discrimination of winter wheat using a growth-state signature," *Remote Sensing of Environment*, Vol. 9, pp 277-294, 1980.

[42] P.H. Swain and H. Hauska, "The decision tree classifier: design and potential," *IEEE Trans. Geoscience Electronics*, Vol. 15, No. 3, July, 1977.

[43] P. Argentiero, R. Chin and P. Beaudet, "An automated approach to the design of decision tree classifiers," *IEEE Trans. PAMI*, Vol. 4, No. 1, Jan., 1982.

[44] M.J. Muasher and D.A. Landgrebe, "A binary decision tree feature selection technique for a limited training sample size," *Machine Processing of Remotely Sensed Data Symposium*, Perdue Univ., 1982.

[45] A.H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sensing of Environment*, Vol. 10, pp 135-163, 1980.

[46] J.M. Tenenbaum, et. al., "Map-guided interpretation of remotely-sensed imagery," *Pattern Recognition and Image Processing Conference*, Aug. 6-8, 1979, Chicago.

[47] D.M. McKeown, "Map-guided feature extraction from aerial imagery," *Tech. Report CS-84-117*, Dept. Computer Science, Carnegie-Mellon Univ.

[48] R.M. Hoffer, et. al., "Digital processing of Landsat MSS and topographic data to improve capabilities for computerized mapping of forest cover types," *LARS Tech. Report 011579*, Laboratory for Applications of Remote Sensing, Perdue Univ.

[49] J.A. Richards, D.A. Landgrebe and P.H. Swain, "A means for utilizing ancillary information in multispectral classification," *Remote Sensing of Environment*, Vol. 12, pp 463-477, 1982.

[50] T. Pavlidis, *Algorithms for Graphics and Image Processing*," Comp. Sci. Press, 1982.

[51] P.H. Winston and B.K.P. Horn, *LISP*, Addison-Wesley, 1981.

[52] A.R. Hanson and E.M. Riseman, "VISIONS: A computer system for interpreting scenes," in *Computer Vision Systems*, Hansen and Riseman (eds), Academic Press, 1977.

[53] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, 1980.

[54] D. Marr, "Early processing of visual information," *Phil. Trans. Royal Soc. London, Proceedings Biological Sciences*, Vol 275, 1976.

[55] M. Tavakoli, "Toward the recognition of cultural features," *Proc. Image Understanding Workshop*, 1980.

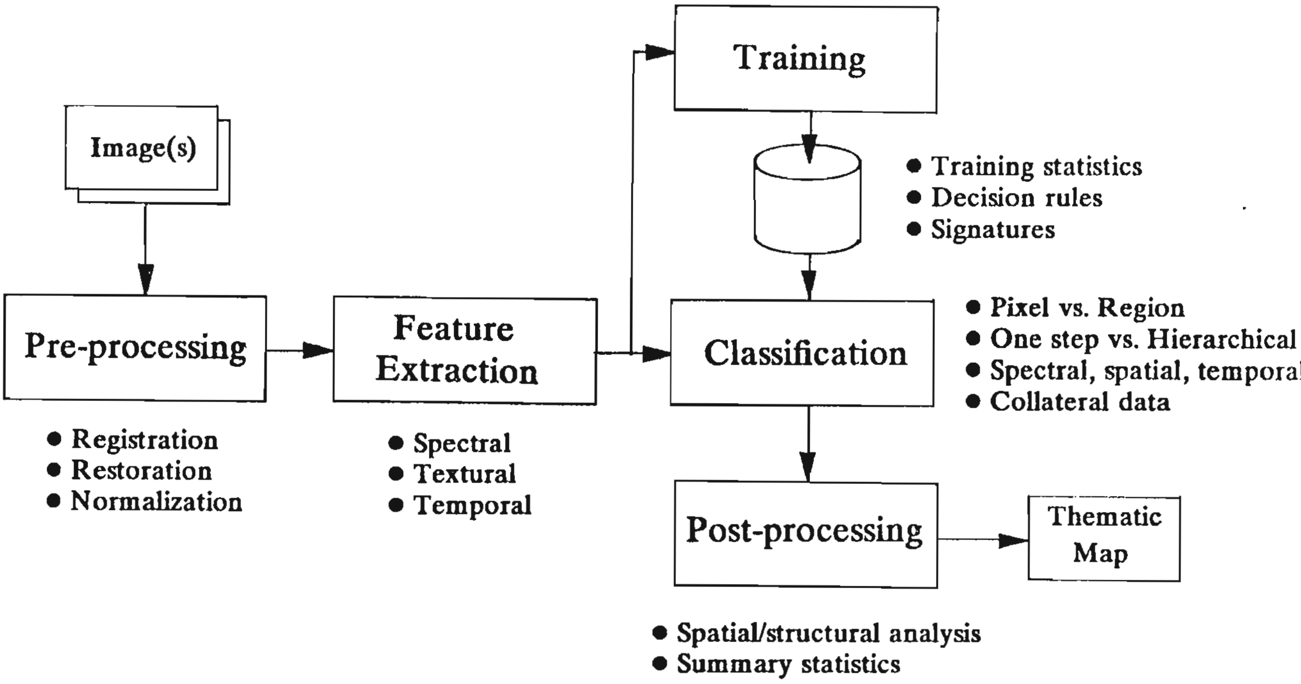


Fig. 1 Overview of multispectral classification process

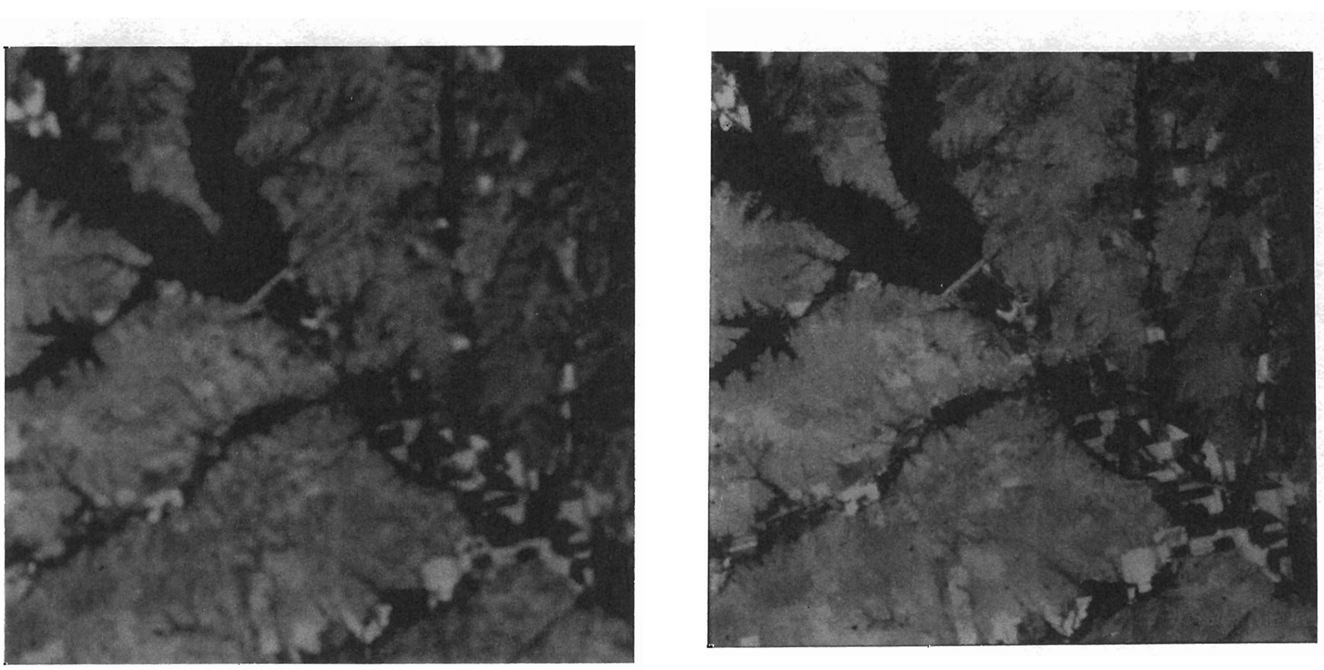


Fig. 2 Example of image pre-processing: thermal band sharpening

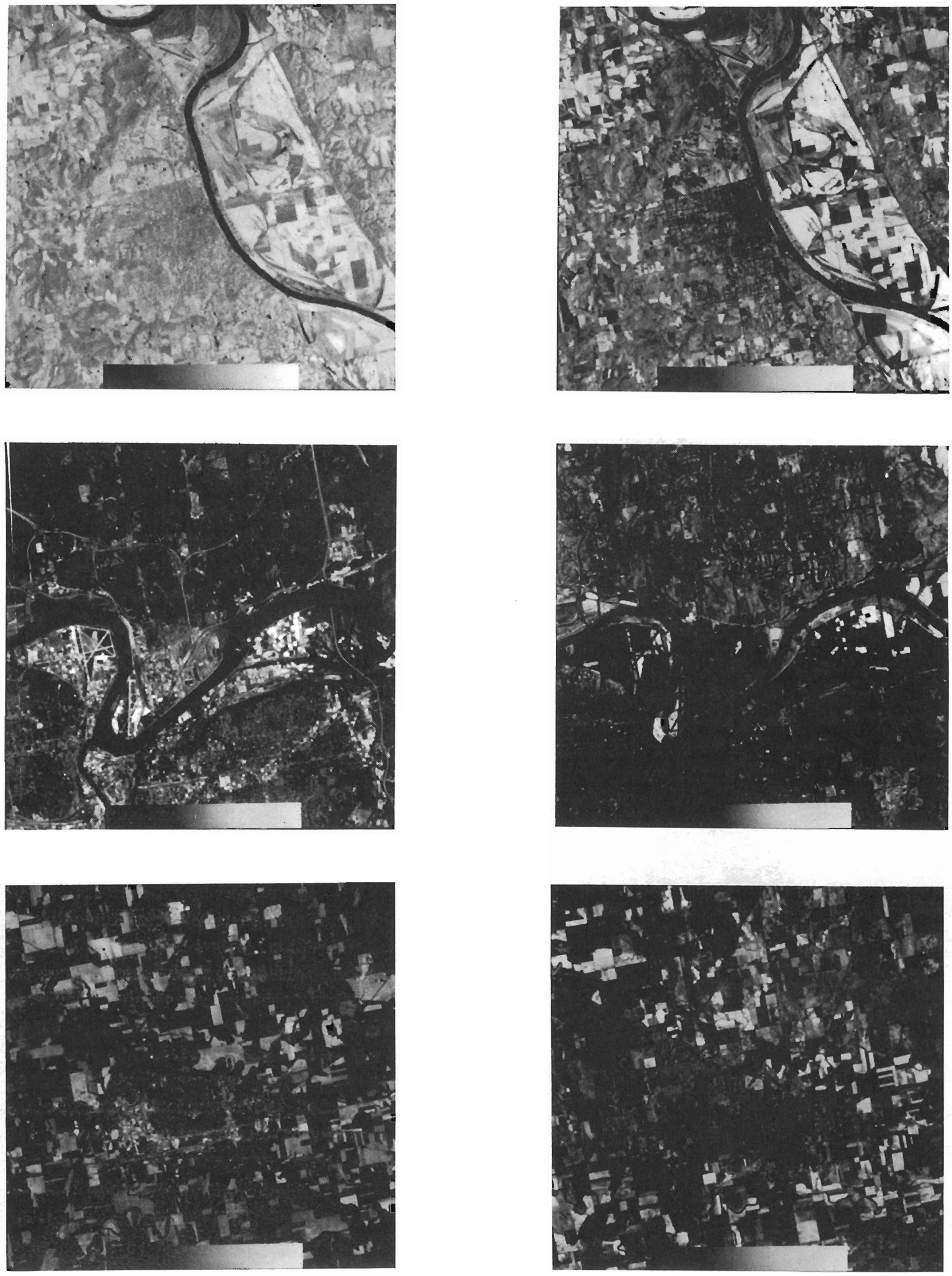
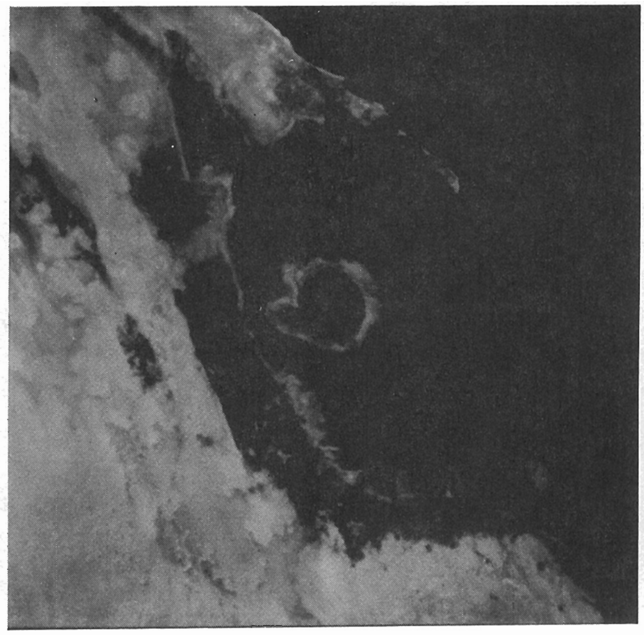
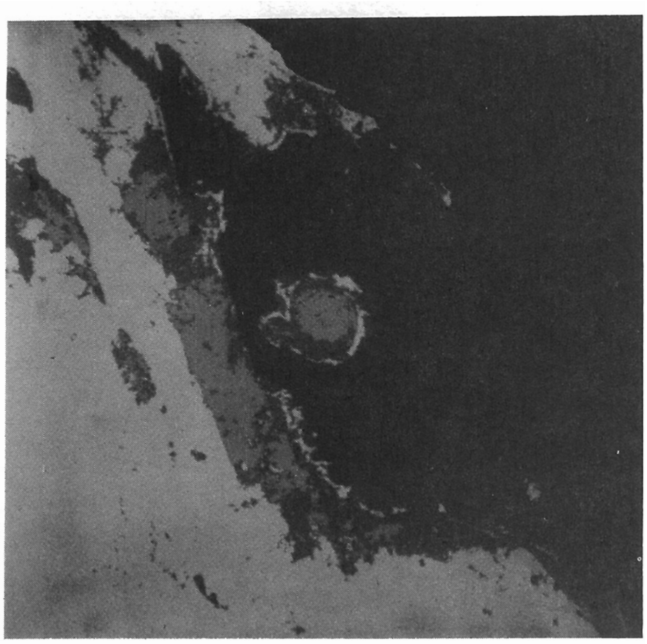


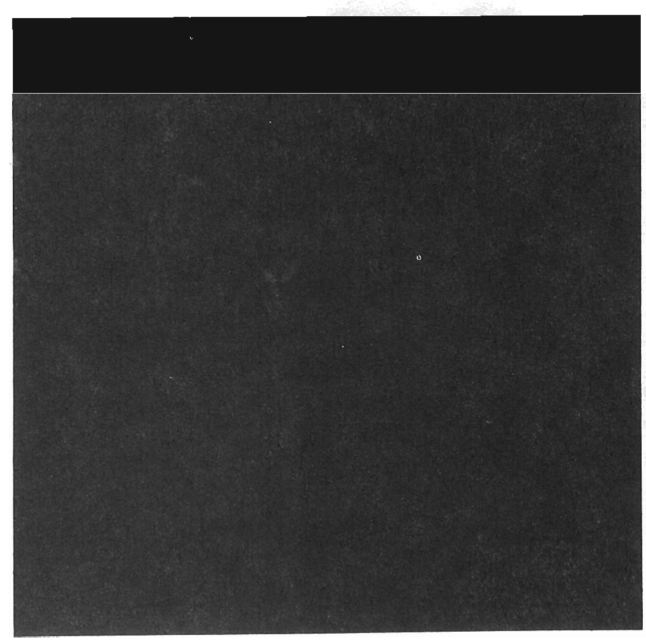
Fig. 3 Comparison of feature extraction techniques in three different scenes: principal components transform (left) and tasseled cap transform (right)



(a) Black and white rendition of MSS bands 4,5, and 6.



(b) Thematic map



(c) Minimum distance image

Fig. 4 Landsat MSS classification example (Saudi Arabia)

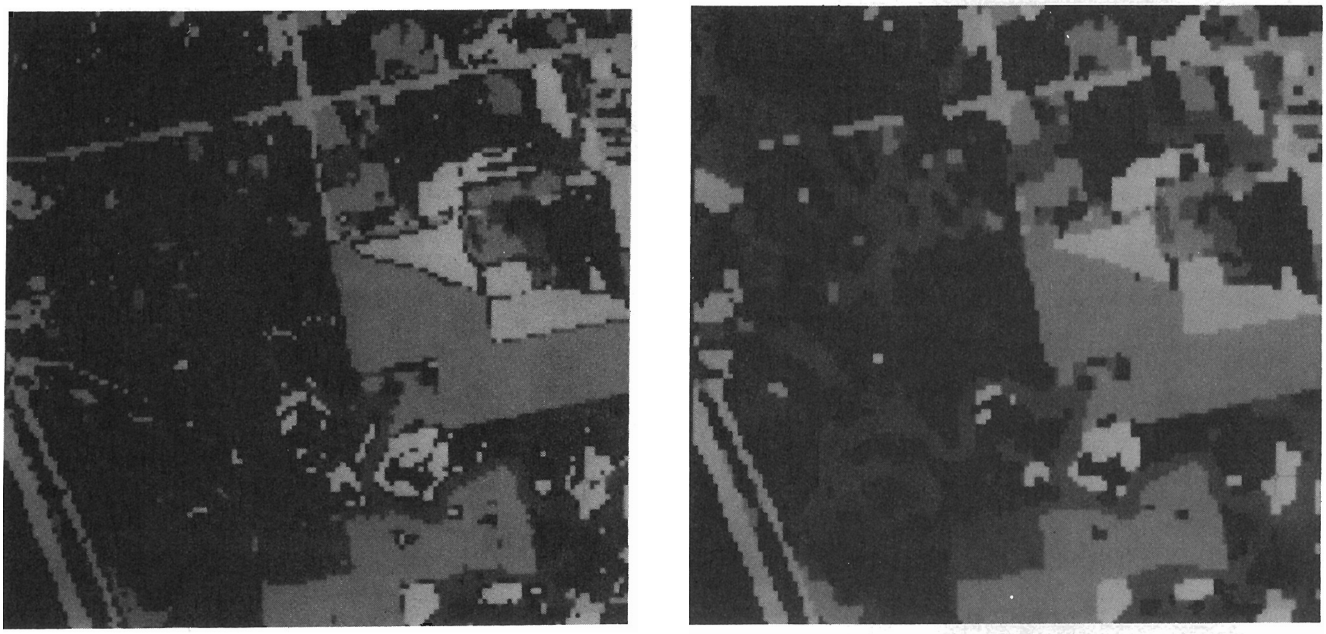


Fig. 5 Effect of using spatial information in classification

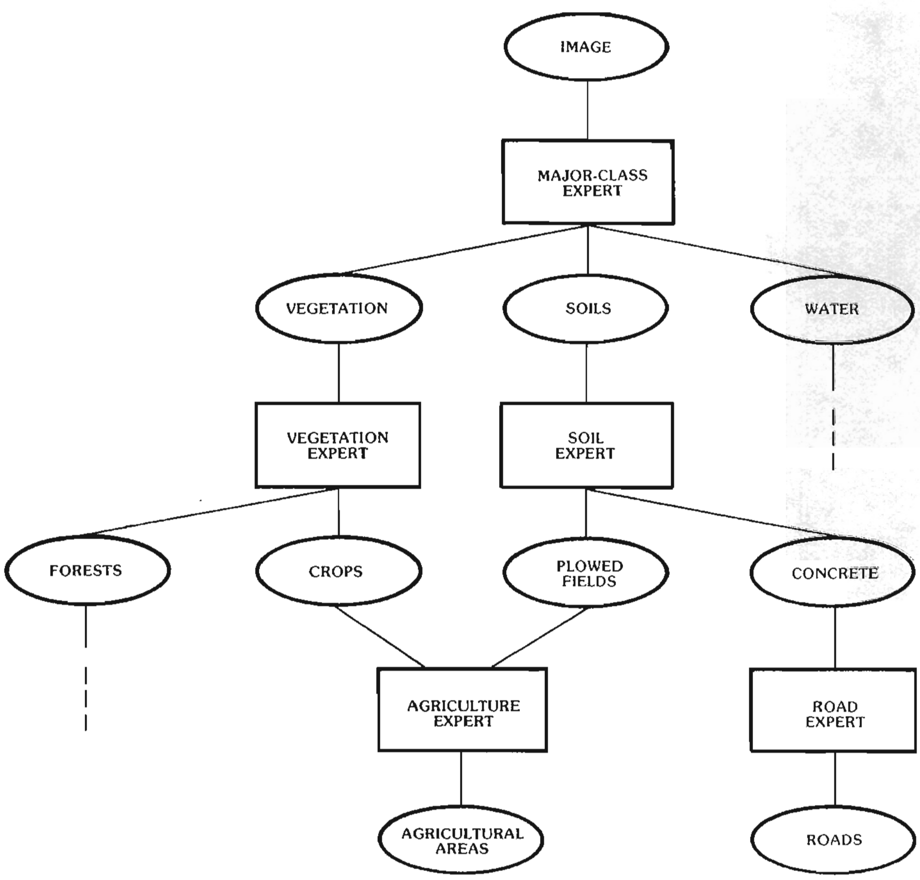
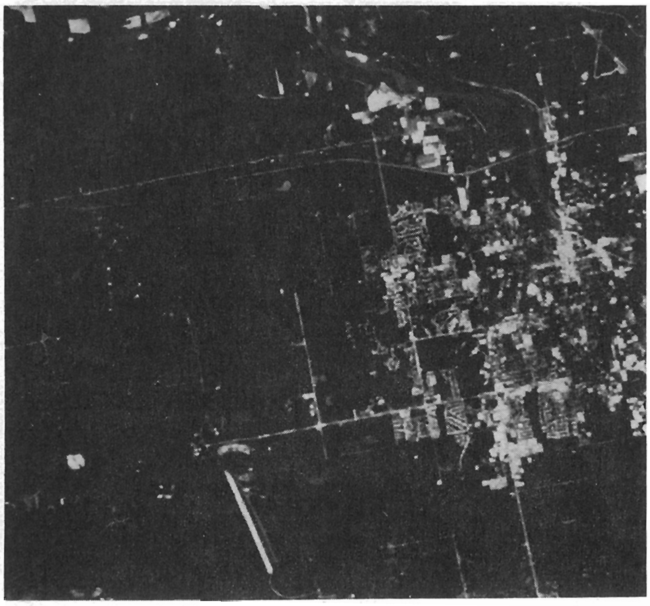


Fig. 6 Hierarchical classification tree



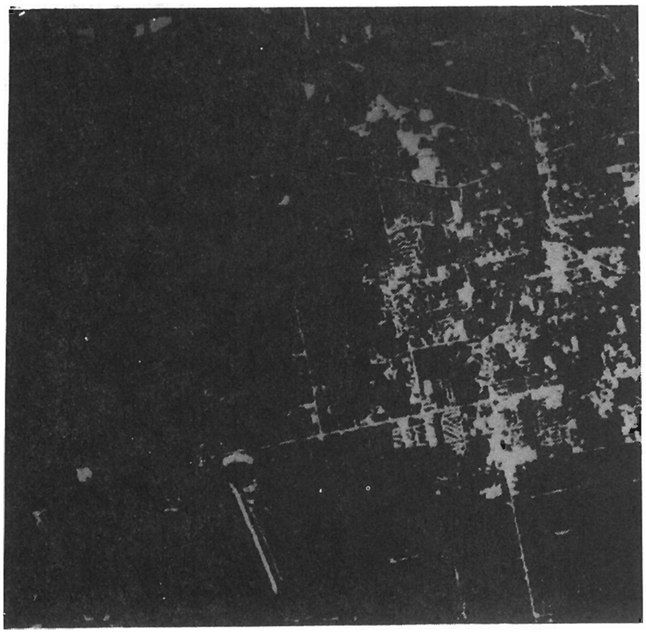
(a) TM image over Lawrence, Kansas

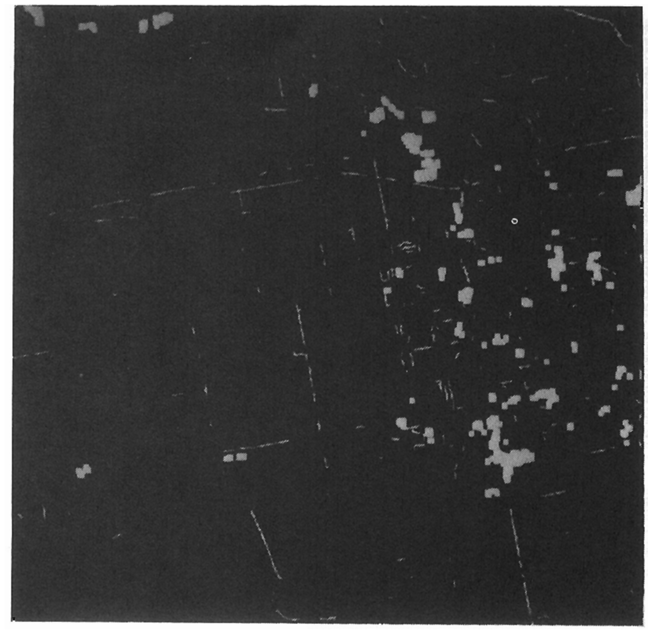
(b) Major classes

(c) Soil-like materials

(d) Vegetation

Fig. 7 Hierarchical classification example





(a) Pixels composed of concrete

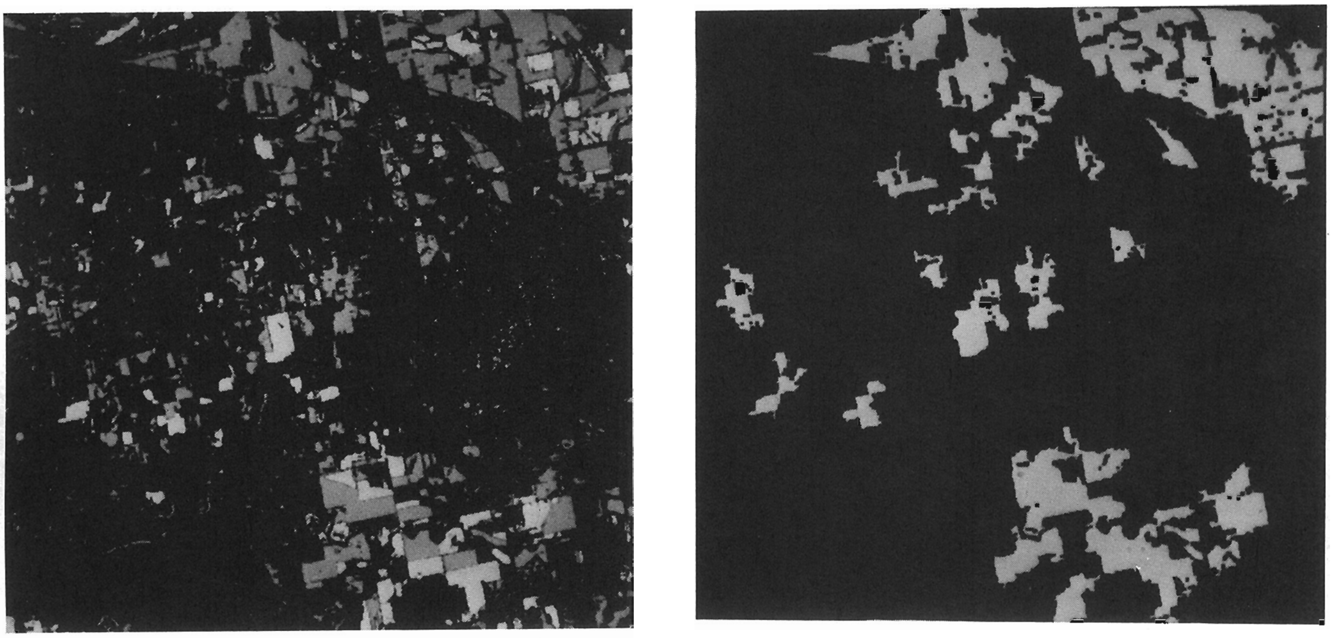
(b) Compact areas and thin segments

Fig. 8 Binary image processing example

Road segments (length > 1 pixel)

Grouped into extended segments ($\alpha < 30$ degrees and R = 2.0)

Fig. 9 Growing extended road segments



(a) Crops and plowed fields

(b) Merged and selected according to aggregate area

Fig. 10 Aggregating regions in the thematic map