2. Automatic Remote Sensor Image Processing

R. M. HARALICK

With 10 Figures

2.1 Remote Sensing

Detecting the nature of an object by external observation, without physical contact with the object, is remote sensing. The advantages of gathering data about objects remotely are that the object is usually not disturbed; objects in inaccessible areas can be examined; and a large amount of information over any spatial area can be obtained.

The earliest and most useful form of remote sensing is photography. Here, photon energy (in the visible or near-visible portion of the spectrum) which is radiating or reflected from objects is collected by a camera (the sensor) and recorded on a light sensitive film emulsion. Aerial multiband and color photography can be used to determine the number of acres of land in different uses, such as rangeland, cropland, forest, urban, swamp, marsh, water, etc. It can help identify rock and soil type, vegetation, and surface water condition [2.1, 2].

The camera, of course, is not the only kind of remote sensor. Other types of remote sensors include the multispectral scanner, the infrared scanner, the scanning radiometer, the gamma ray spectrometer, the radar scatterometer, and the radar imager. Figure 2.1 illustrates the typical scanning sensor. A rotating mirror scans the ground scene in a line by line manner with the forward velocity of the sensing platform causing the line scans to be successively displaced. The mirror directs the received energy to a detector which converts it to a video signal which is then recorded by a film recorder to make an image. Thermal infrared scanner systems can produce imagery in daytime or nightime since their detectors are sensitive to emitted heat and not to light. The near-infrared systems are able to penetrate haze and dust making it possible to get good contrast images not achievable by aerial photography. MALILA [2.3] documented the advantages multispectral scanners have over cameras for image enhancement and discrimination. The NASA Third and Fourth Annual Earth Resources Program Reviews [2.4, 5] and the Michigan Symposia on Remote Sensing of Environment contain numerous papers on the application of multispectral or thermal scanners. See also [2.6, 7].

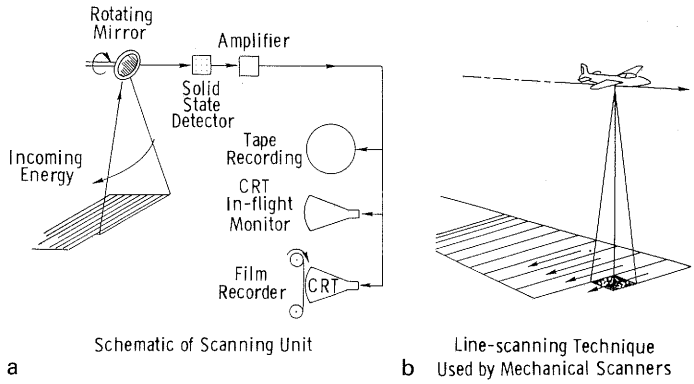

Fig. 2.1. (a) Schematic of scanning unit. (b) Line-scanning technique used by mechanical scanners.

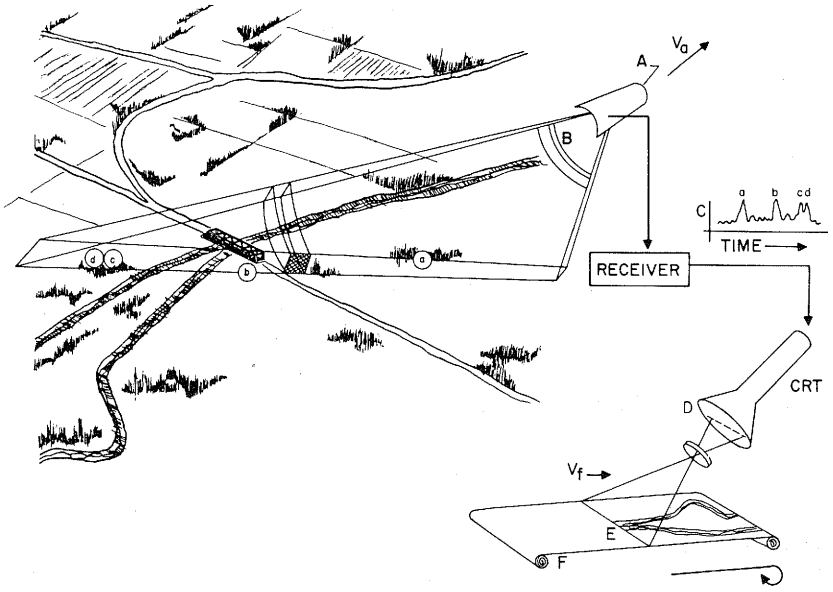

Fig. 2.2. Side-looking radar scanning

The radar imager shown in Fig. 2.2, unlike the others, is an active system. It illuminates the ground scene by transmitting its own microwave radiation and its antenna receives the reflected energy. After appropriate processing and electronic amplification the resulting video signal can be recorded by film recorder to make an image. The radar

Table 2.1. Agriculture/forestry

Application	Type of data Required	Data use
Agriculture	Inventory and distribution	Farm/forest interfaces Boundaries Topographic maps Crop type and density Crop expected yield Livestock census
	Infestation	Disease damage Insect damage Infestation patterns
	Land use	Soil texture Soil moisture and irrigation requirements Soil quality to support vegetation Farm planning
Forestry	Inventory and distribution	Forest texture Boundaries Topographic maps Tree types and count Logging yield and production Location of tree types
	Fire, disease and reclamation	Fire location and damage Pattern and discontinuity Soil moisture and texture Insect and disease damage
Conservation	Land use Grassland vigor	Maps Wildlife management

signal is influenced by ground conductivity and surface roughness. Radar images are particularly good at providing integrated landscape analysis [2.8–8a], as well as vegetation [2.9], geologic [2.10], and soil moisture [2.11] information.

Tables 2.1–5 list the variety of uses to which earth resources remote sensing has been put in the areas of agriculture, forestry, hydrology, geology, geography, and environment monitoring. MERIFIELD et al. [2.12] discussed the potential applications of spacecraft imagery.

The success of manual interpretation or automatic processing of remotely sensed imagery is constrained by the kind of physical character-

Table 2.2. Hydrology

Application	Type of data required	Data use
Water inventory	Water inflow into basins, rivers and streams	River effluents Drainage basin features Reservoir levels Ground water surveys Irrigation routes
Flood control	Excess surface water	Flood location Damage assessment Rainfall monitoring Erosion patterns
Water pollution	Natural and industrial pollution	Color Spectral signature Pollution content Salt content
Water conservation	Evaporation and transpiration	Evapotranspiration
Water resources	Seeps and springs	Temperature variation Water quality
	Glaciology	Frozen water inventory Snow surveys

istics the sensors can detect. The sensors used in the various portions of the electromagnetic spectrum are sensitive to dissimilar energy-matter interactions. They detect different characteristics and, therefore, convey different kinds of informations. In crop identification problems, for example, the molecular absorptions produce color effects in the visible region which convey information about crop type or condition; in the infrared region, the diurnal cycle of thermal response under an insolation load may give information about moisture stresses within crops; in the radar region, the backscattered return is primarily related to surface roughness and dielectric constant, which may in turn be related to crop type, percent cover, and moisture content.

For those readers interested in exploring the remote sensing applications literature further the following guide is offered. A bibliography of the earth resources remote sensing literature has been complied by NASA and is current to 1970. The third and fourth Earth Resources Program Reviews [2.4, 5] (there have been none since 1972) provide descriptions of a large number of investigations funded by NASA. The second and third Earth Resources Technology Satellite (ERTS) Symposia in 1973

Table 2.3. Geology

Application	Type of data required	Data use
Petroleum and minerals detection	Surface and sub-surface patterns	Lithology studies Outcrops Plot magnetic fields Earth folds Drainage patterns Soil compacting and stability Soil density Surface stratification and electrical conductivity
Volcano prediction	Surface feature changes	Temperature variation Lithologic identification Spatial relations
Earthquake prediction	Surface stress and discontinuities	Linear microtemperature anomalies Slope distribution Crust anomalies Soil moisture
Engineering geology	Geothermal power sources	Temperature anomalies Surface gas
	Landslide prediction	Soil moisture Slope distribution Crust anomalies

(NASA GSFC) describe much of the remote sensing done with the 4-channel multispectral scanner in the ERTS satellite.

Since 1962 the Environmental Research Institute of Michigan and the University of Michigan have co-sponsored annual remote sensing symposia with the papers published as Proceedings of the *n*-th International Symposium on Remote Sensing of Environment. In 1972 the University of Tennessee Space Institute also began organizing annual remote sensing symposia. The conference proceedings of these symposia are obtainable from the respective universities and contain unreviewed papers (abstracts only are reviewed). The annual proceedings of the American Society of Photogrammetry have remote sensing related application papers (also unreviewed). In 1969 the journal *Remote Sensing of Environment* was begun and it along with *Photogrammetric Engineering* and *Modern Geology* contain much of the refereed literature.

Table 2.4. Geography

Application	Type of data required	Data use
Transportation	Identify features	Locate terminals, buildings Locate roads, tracks Traffic count
	Locate new facilities	Make maps at scales of 1:25000 to 1:250000 Cultural factors Economic factors
Navigation	Topography	Make maps at scales of 1:100000 to 1:250000
Urban planning	Locate settlements	Boundary and topography
	Type of settlements	Color, texture, contrast
	Distribution of settlements	Pattern of housing density
	Occurrence of recreation areas	Color, texture, shape
	Population distribution	Population count
	Classification of facilities	Industrial planning 1:50000 scale maps Cultural/economic factors Land use intensity Spectral signature Heat budgets

Table 2.5. Environment

Application	Type of data required	Data use
Air quality monitoring	Backscatter pattern analysis Spectroscopy	SO_2 and NO_2 concentration and distribution Forecasting
Water quality monitoring	Oil slicks, Effluent Salt water intrusion	Color tones Spectral signatures
Violation detection	Spatial/temporal source data	Locate land or sea coordinates of pollution source

2.2 The Image Processing Problem

Remote sensor data systems are designed to obtain information about various aspects of an environment by remotely measuring the electromagnetic transmittance, reflective, or emissive properties of the environment. To obtain information about the "pertinent" aspects of the environment, the sensor system must preserve or mirror the "pertinent" structure of the environment onto the data structure. Automatic remote sensor data processing or pattern recognition is concerned with evaluating, inventorying, and identifying those environmental structures which are and which are not preserved through the sensor's eyes.

There are two modes of automatic processing: one mode requires special purpose hardware, uses film format or analog tape input, and operates in near real time; the other mode requires a general purpose computer, uses digital tape input, and operates relatively more slowly. The first mode—a hardware system mode—is capable of limited sorts of processing while the second mode—a software system mode—is capable of rather sophisticated processing. Neither of these modes has reached the stage of production volume pattern recognition of remotely sensed data; progress has, however, brought them out of the experimental stage, into the prototype stage, and as shown by the corn blight experiment, significant amounts of data can be processed [2.13].

FU et al. [2.14] discussed the general pattern recognition approach to the automatic processing of remotely sensed data. HARALICK et al. [2.15] described some pattern recognition techniques for crop discrimination on radar imagery. CENTNER and HIETANEN [2.16] used automatic pattern recognition techniques to distinguish between cultivated land, urban land, wooded land, and water. ANUTA and MACDONALD [2.17] discussed automatic recognition of crop types from multi-band satellite photography. HOFFER et al. [2.18] described how automatic pattern recognition can supplement manual photo-interpretation methods for the recognition of crop types. TURINETTI and MINTZER [2.19] discussed computer analysis of day and night thermal imagery for land-use identification. SWAIN [2.20] treated pattern recognition theory for remote sensing data analysis. LANDGREBE [2.21] discussed the machine processing approach for remotely sensed imagery. Thus, there is no lack of data indicating that automatic processing techniques can be applied to remotely sensed images. The most recent review of digital image processing activities for remotely sensed images was prepared by NAGY [2.22].

To examine the remotely sensed imagery processing problem, we will consider the case where we have obtained a multi-image set by a combination of perhaps radar images of various frequency-polarization

conditions, multispectral scanner images, or multi-band photography. Some of the images may be time-simultaneous and others time-sequential. To be realistic, we will suppose that all sensors which rely on the sun for the illumination source have obtained their pictures at perhaps different solar altitude angles, solar azimuth angles, and meridian angles (see [2.23–25] for documentation of spectral reflectivity variation with angle).

If we assume our problem is an agricultural crop identification problem, we would first like to separate out crops from non-crops (crop detection); then we would like to make an accurate identification of the detected crops and produce a thematic land use map or make estimates of the acreages of the various crops.

To be able to do this rather well-defined interpretation task automatically, we must register or congruence the individual images, calibrate or normalize the grey tones on each image, extract the relevant features required to detect and identify crops, use a decision rule which in some sense makes the best identification possible relative to the extracted features, and finally estimate the decision rule error. Section 2.3 discusses preprocessing procedures; Section 2.4 discrimination, feature extraction, and clustering methods; and Section 2.5 approaches to the quantification of image texture. Sections 2.6 and 2.7 describe some hardware and software aspects of the problem. The glossary of terms used in the remotely sensed image pattern recognition area prepared by HARALICK [2.26] is a general reference for the technical terms used in this paper.

2.3 Preprocessing Problems

2.3.1 Grey Tone Normalization

Each type of remote sensor has its own internal type of calibration or lack of calibration. For most remote sensors, this internal calibration cannot be done accurately or consistently from sensor to sensor and from day to day. Photographic sensors are particularly plagued with calibration problems since little controlled attempt is usually made to compensate for exposure, lens, film, print, developer, and off-axis illumination fall-off differences. PETTINGER [2.27], for example, documented variability in color image quality for photography flown over the Phoenix area. He ascribed the color balance variability to exposure, film (age and storage), and processing differences. The environment compounds the problem by varying atmospheric variables such as haze and cloud cover.

side of the image to the other result in different received radiances even from the same kinds of objects because reflectivity is a function of look angle.

When the spectral bandwidths are small enough, it may be assumed that the product $R_{s_1}\tau$ for one spectral band is close enough to the product $R_{s_2}\tau$ for an adjacent spectral band so that they may be considered to be equal. Then, when the backscatter in both bands is small enough to be negligible, the ratios of the viewed radiances of adjacent bands will be equal to the ratios of the object reflectivities.

$$\left.\begin{aligned} R_{r_1} &= R_{s_1}\tau_1\varrho_1 + b_1, \text{ energy and reflectivity relationship for band 1} \\ R_{r_2} &= R_{s_2}\tau_2\varrho_2 + b_2, \text{ energy and reflectivity relationship for band 2} \end{aligned}\right\} \text{ in general,} \tag{2.2}$$

$$\left.\begin{aligned} R_{r_1} &= R_{s_1}\tau_1\varrho_1 \\ R_{r_2} &= R_{s_2}\tau_2\varrho_2 \end{aligned}\right\} \text{ energy and reflectivity relationship when backscatter is negligible,} \tag{2.3}$$

$$\frac{R_{r_1}}{R_{r_2}} = \frac{R_{s_1}\tau_1}{R_{s_2}\tau_2}\frac{\varrho_1}{\varrho_2} \quad \text{is ratio of received radiances when backscatter is negligible,} \tag{2.4}$$

$$\frac{R_{r_1}}{R_{r_2}} = \frac{\varrho_1}{\varrho_2} \quad \text{when} \quad R_{s_1}\tau_1 = R_{s_2}\tau_2 . \tag{2.5}$$

Thus, for a 12-channel scanner, the original 12-tuple $(x_1, x_2, \ldots, x_{12})$, x_i being the energy from the i-th channel, would be normalized to the 11-tuple $(x_1/x_2, x_2/x_3, \ldots, x_{11}/x_{12})$. Note that with this normalization procedure, one dimension is lost due to the fact that no information is obtained when the twelfth channel is normalized.

It is possible to generalize the normalization procedure to take into account the backscatter if some additional assumptions are made: the adjacent spectral bands are small enough so that the product $R_s\tau$ is equal for each three adjacent bands and the backscatter b is equal for each three adjacent bands. Under this assumption, the ratio $(R_{r_1} - R_{r_2})/(R_{r_2} - R_{r_3})$ equals $(\varrho_1 - \varrho_2)/(\varrho_2 - \varrho_3)$. Thus for a 12-channel scanner the 12-tuple $(x_1, \ldots, x_{12})$ would be normalized to the 12-tuple $[(x_1 - x_2)/(x_2 - x_3), (x_2 - x_3)/(x_3 - x_4), \ldots, (x_{10} - x_{11})/(x_{11} - x_{12})]$. KRIEGLER et al. [2.28] and CRANE [2.29] first reported these normalization techniques. SMEDES et al. [2.30] noted that normalization results in more accurate maps with fewer training areas. NALEPKA and MORGENSTERN [2.31] indicated that percent identification accuracy can improve especially for pixels at the edges of an image by the use of these normalizations.

There are several ways to operate on such uncalibrated imagery to bring it into various normalized forms. We will discuss a few of them which are implementable in both hardware and software data processing systems.

Multi-Image Normalization for Viewing Angle, Atmospheric, and Intensity Variations

Sometimes, due to camera design, the illumination intensity across the film is not uniform. More light energy falls on the center of the film than on the edges (off-axis illumination fall-off). Sometimes, due to incorrect camera exposures or sun and cloud variations, atmospheric haze, or viewing angle variations, parts of frames or whole frames may be lighter or darker than normal. Similar sorts of problems occur with moisture variation in multi-frequency radar imagery and with cloud shadows on multispectral thermal imagery.

When the variations are only in intensity, normalization of intensity can be done. In intensity normalization, the assumption is made that the relevant information concerning crop detection and identification is in hue and not in color intensity; that is, the information is in the relationships of one emulsion or channel to another. The procedure is to take the densities of each emulsion or channel at each resolution cell and divide them by some weighted sum of the emulsion or channel densities at that location. The weights in the weighted sum are chosen so that the resulting sum is proportional to total illumination intensity. In this way the color in each resolution cell on the multi-image can be standardized to the same intensity.

When the variations are due to atmospheric aerosol and haze differences, in addition to look angle differences, the normalization is imperfect and is more complicated. The usual model relating reflectivity to energy received is

$$R_r = R_s \tau \varrho + b\,, \tag{2.1}$$

where R_r is the radiance received at the sensor, R_s is the irradiance of the source, ϱ is the reflectivity of the object, b is the backscatter and sky path radiance, and τ is the atmospheric transmissivity.

In general, all parameters may be functions of spectral waveband, viewing angle, direction, and distance. In processing remotely sensed imagery, the object reflectance ϱ is usually the feature wanted and under ideal conditions the received radiance at the sensor can indeed be proportional to ϱ. The atmospheric aerosol and haze variations make the transmissivity τ and the backscatter b variable from day to day and from spectral band to spectral band. The differences in look angle from one

Single Image Quantization

Consider now a single image in a multi-image set. Perhaps this image had been in the developer too long or it is visibly pre-fogged or it is over-exposed or the film was unusually sensitive. Perhaps the scanner detector lost some of its sensitivity or the A/D converter changed its calibration. Quantization can make single images invariant to these sorts of changes.

In general, the quantization process divides the range of grey tones or grey tone levels into K intervals (λ_1, λ_2), $(\lambda_2, \lambda_3),\ldots,(\lambda_K, \lambda_{K+1})$ determined in some fashion. A new quantized image is generated from the original image. Any resolution cell on the original image having a grey tone lying in the first interval (λ_1, λ_2), is assigned on the new image to the first quantization level; any resolution cell on the original image having a grey tone lying in the second interval (λ_2, λ_3) is assigned on the new image to the second quantization level, and so on. The quantized image, therefore, has only K possible grey tones.

There are two common ways which are used to determine the quantization intervals, depending upon the kind of assumption one is justified in making as to the nature of the uncalibrated changes. In equal interval quantizing, the assumption is that the density change can be described by a simple linear scaling and that the overall darker or lighter appearance can be described by the addition or subtraction of some grey tone. To perform equal interval quantizing, the lightest grey tone λ_1 and the darkest grey tone λ_{K+1} on the original uncalibrated image are determined. The range $\lambda_{K+1}-\lambda_1$ of grey tones is then divided into K equal length intervals, each of length $(\lambda_{K+1}-\lambda_1)/K$. The interval end points are determined by

$$\lambda_i=\lambda_1+(i-1)(\lambda_{K+1}-\lambda_1)/K\,, \qquad i=1, 2,\ldots, K-1\,.$$

In this way, two images which are the same except for a linear scaling and grey tone translation will produce the same equal interval quantized image. Thus, equal interval quantizing is invariant with respect to linear transformations.

Another normalization procedure that achieves the same kind of invariance as discrete equal interval quantizing is to determine the mean and standard deviation of the grey tones on the original image and generate a normalized image by subtracting the mean and dividing by the standard deviation.

In equal probability quantizing, it is assumed that the original image differs from the ideal calibrated image by some monotonic function which can be linear or nonlinear; this assumption implies that any object which is darker than another object on the original uncalibrated image

will also be so for the ideal calibrated image. To perform equal probability quantizing, the range of grey tones is divided into K intervals, each of equal probability. In other words, for any interval $(\lambda_i, \lambda_{i+1})$ of grey tones, the number of resolution cells on the original image having a grey tone in that interval is the same as for any other interval and is in fact equal to $(1/K)$-th of the total number of resolution cells on the image.

If equal probability quantizing were applied to two images, one a calibrated image and the other an uncalibrated image differing from the calibrated image only by a monotonic grey tone transformation, then the two equal probability quantized images produced from them would be identical. Hence, equal probability quantizing is invariant with respect to any monotonic grey tone transformation.

2.3.2 Image Registration, Congruencing, Rectification

In order to create a multi-image set from separate images which might have been taken of the same area but perhaps at different times or with different sensors, each individual image must be aligned or spatially fit with every other image in the set so that there is no geometric distortion and all corresponding points match. This alignment problem occurs for multi-image sets which have time simultaneous image combinations from different sensors (such as radar imagery with multi-band, or multi-band with scanner imagery), or for time sequential imagery from any sensor(s). When the individual images are of the same scale and geometry, the alignment can be done by rotating and translating one image with respect to another until the cross-correlation between the images (or between the derivatives of the images) is maximal. The process of aligning the image by translation and rotation is called registration and devices which do such translation and rotation are often called image correlators.

When the individual images in the multi-image set have different geometry, the alignment process is more difficult. Geometrical irregularity can be caused by 1) a multi-image set containing a combination of images taken by different types of sensors, each introducing its own kind of geometric distortion due to the way the sensor operates, or 2) the same type of sensors having slightly different "look angles" or altitudes, or 3) various sorts of uncompensated platform motion. In this case the alignment process is called congruencing, since it must literally bring points into a point by point correspondence. When the congruencing process has the additional constraint that the image geometry after congruencing must be planimetric, the process is called image rectification.

Often congruencing or recitification for images in photographic form is done by one or a sequence of projective transformations with optical, electro-optical, or electronic equipment. The basic idea of a projective transformation can be easily visualized by thinking of the image being on a flat rubber surface. The projective transformation places a quadrilateral frame on the surface and then pulls or pushes each side of the frame at a different orientation angle to obtain the desired geometry. The optical devices perform the projective transformation on the entire image, while the electro-optical devices are capable of dividing the image into regions and operating on each region with a different transformation. The parameters for the transformation can be determined from supplied ground control points, or by an automatic cross-correlation scheme such as that employed by the BAI Image correlator [2.32]. McEwen [2.32a] evaluated various analog techniques for image registration.

A greater degree of flexibility is afforded by the use of a digital computer to do digital image registration or image congruencing. Projective transformations, affine transformations, and polynomial transformations are all possible. Let R and C be the index sets for the rows and columns of the image and G be the set of grey tones for the image. Correcting for geometric distortion, or spatially fitting one image to another by registering or congruencing, corresponds to constructing from the input image I, I: $R \times C \to G$, an output image J, J: $R \times C \to G$ by some transformation f from the spatial coordinates of the output image to the spatial coordinates of the input image, i.e. $f: R \times C \to R \times C$, so that the congruenced image can be determined by

$$J(r, c) = I(f(r, c)) .$$

This equation says that for each (row, column) coordinates (r, c) in the output image, the grey tone which we put there is $J(r, c)$, and this grey tone is obtained as the grey tone appearing on the input image I at coordinates (r', c'), where $(r', c') = f(r, c)$. For affine transformations [2.33, 34]

$$\begin{aligned} r' &= [a_{11}r + a_{12}c + a_{13}] \\ c' &= [a_{21}r + a_{21}c + a_{23}] . \end{aligned} \tag{2.6}$$

For projective transformations [2.35, 36]

$$\begin{aligned} r' &= \left[\frac{a_{11}r + a_{12}c + a_{13}}{a_{31}r + a_{32}c + a_{33}}\right] \\ c' &= \left[\frac{a_{21}r + a_{22}c + a_{23}}{a_{31}r + a_{32}c + a_{33}}\right] . \end{aligned} \tag{2.7}$$

For K-th order polynomial transformations [2.37]

$$r' = [\sum_{i=0}^{K} \sum_{j=0}^{K-i} a_{ij} r^i c^j]$$
$$c' = [\sum_{i=0}^{K} \sum_{j=0}^{K-i} b_{ij} r^i c^j] . \tag{2.8}$$

(Notation: $[x]$ means the nearest integer to x.)

These kinds of spatial transformations are intended only to account for the low (spatial) frequency, sensor-associated spatial distortions (centering, size, skew, pincushion) and for distortions due to earth's curvature, sensor attitude and altitude deviations. Note that since the pure projective affine or polynomial transformations do not, in general, produce r' and c' as integers, we take whatever values these transformations give for r' and c' and either interpolate on the grey tones surrounding these coordinates, or, as shown in our equations, convert them to the corresponding nearest integers and take the nearest neighbor grey tone. This means that in actual implementation, all the resolution cells of the output image are examined and for each one the corresponding coordinates on the input image are determined. Then in the nearest neighbor interpolation approach, for example, the grey tone on the output image is defined to be the grey tone appearing in the pixel of the input image closest to the determined corresponding coordinates.

Having seen how the transformation is implemented, we must now discuss how the parameters of the transformation are determined. When the transformation is a simple translation

$$r' = r + a$$
$$c' = c + b ,$$

the translation parameters a and b can be determined automatically using cross correlation or distance measures. If I is the image which needs to be translationally registered so that it fits image J, the parameters a and b can be chosen so that

$$\frac{\{\sum_{(r,c)\in R\times C} I(r+a, c+b)J(r, c)\}^2}{\sum_{(r,c)\in R\times C} I^2(r+a, c+b) \sum_{(r,c)\in R\times C} J^2(r, c)}$$
$$\geqq \frac{\{\sum_{(r,c)\in R\times C} I(r+\alpha, c+\beta)J(r, c)\}^2}{\sum_{(r,c)\in R\times C} I^2(r+\alpha, c+\beta) \sum_{(r,c)\in R\times C} J^2(r, c)}$$
for all α, β (cross correlation measure),

or (2.9)

$$\sum_{(r,c)\in R\times C} |I(r+a, c+b) - J(r, c)|^P \leqq \sum_{(r,c)\in R\times C} |I(r+\alpha, c+\beta) - J(r, c)|^P$$
for all α, β (distance measure).

The above summations over rows and columns are actually taken only over those points $(r, c) \in R \times C$ such that $(r+a, c+b) \in R \times C$. ANUTA [2.38] discussed the spatial registration problem from the cross-correlation point of view and implemented the algorithm using the fast Fourier transform technique. BARNEA and SILVERMAN [2.39] described a sequential similarity detection algorithm for translational registration. Essentially it is a distance approach implemented in a fast way by not requiring all terms in the summation to be computed for each translation (α, β). WEBBER [2.40] combined an affine transformation with a sequential similarity detection algorithm to determine all the parameters of the affine transformation.

When the image registration or congruencing problem is more involved than simple translation it is more difficult to determine the spatial transformation parameters automatically. Usually, corresponding ground control points are identified by normal photo interpretation methods. Care is taken so that the ground control points are not concentrated in any one area of the image but to the best possible degree are spread uniformly across the image. Then a least squares fit is performed to determine the parameters. In the K-th order polynomial registration method for $K \geqq 2$, the coefficients of the high-order terms can become strong functions of the location of the matching ground control points. This is undesirable. YAO [2.41] discussed a piecewise rubber sheeting process which is less sensitive to ground control point placement.

MARKARIAN et al. [2.37] developed the polynomial approach to spatial registration and discussed how a point shift algorithm can be used to speed up the calculations when the actual amount of geometric correlation needed is small. RIFFMAN [2.42] compared nearest neighbor, bilinear, and cubic interpolation methods as regards the problem of generating the grey tone on the output image from non-integer spatial coordinates on the input image. As expected, the nearest neighbor interpolation method yields pixel jitter especially for high contrast areas. The bilinear interpolation is free from pixel jitter but reduces resolution. The cubic convolution chosen to approximate a $(\sin x)/x$ kernel has no pixel jitter and does not reduce resolution but requires more computation time.

SZETO [2.35] described a linear interpolation scheme based on projective transformations for the rectification of digitized images. He decomposed the image into quadrilateral subimages and derived a bound on the maximum error between the interpolation scheme and the true projection within each quadrant. When the precomputed error in any quadrilateral subimage exceeds a prespecified limit, the subimage is further divided and the calculation is repeated.

2.4 Image Pattern Recognition

2.4.1 Category and Training Data Selection

Once the image data set has been registered or congruenced and the images appropriately normalized and/or quantized, the most time consuming task of processing may begin. This task involves familiarization with the data and the selection of the categories and appropriate "training data" for them.

The investigator usually has some kinds of land-use categories in mind between which he would like to distinguish. Hopefully, the sensor(s) are sensitive enough to spectral, tone, or texture differences so that the categories can, in fact, be distinguished to a large degree purely on the basis of the images in the multi-image data set. If this is not the case, there is no point to going on and either a redefinition of more distingishable and reasonable categories must be made or different, more sensitive sensors must be used.

Determining the categories to be used is more than a simple matter of naming them. For example, the naming of the category "wheat" in an agricultural remote sensing problem is not enough. Not only are there many kinds of wheat, but the percent of covering of wheat over the ground for wheat of the same maturity may vary from field to field, and the percent of weeds in the wheat field may vary. Does a small-area ground patch the size of a resolution cell consisting of 25% weeds and 75% wheat fall into the wheat category, the weed category, or a mixed category? What percent cover must the wheat have for the category identification to change from bare ground to wheat? How are these variations reflected in the spectral or temporal signatures of these categories?

How can the investigator tell if his categories form a reasonable set? The first step is to examine the data structure by looking at histograms or scattergrams of sampled regions from the different categories. If these histograms or scattergrams show little overlap between data points of different categories, then the categories are resonable. If a pair of categories shows significant overlap for histograms of each channel and/or scattergrams for each pair of channels, then there is a definite problem with distinguishing that pair of categories.

Another way of determining possible categories is by cluster analysis (a topic which we will discuss at greater length in Subsect. 2.4.5) [2.43]. Clustering the training data can determine subclasses having characteristic centroids and typically unimodal distributions over the measurement space. A comparison of the data points in each cluster with their corresponding ground truth characteristics will allow the investigator

to choose category classes whose data points fall in distinct clusters. He thereby assures himself that data points from each category are likely to be discriminated correctly by the decision rule in the pattern identification or classification process. DAVIS and SWAIN [2.44] advocated the clustering approach to category selection.

When the investigator is satisfied with his category choices, he is ready to choose his training data from the multi-image. The training data is given by a specially prepared subset of pixels from the multi-image. The preparation consists of labeling each point in the training data with its true category identification. This identification can come from either ground truth observation or photo interpretation.

Since the training data are to be used to determine a decision rule which will make a category assignment for each pixel in the whole image data set, there are two points to which attention must be paid: 1) the training data for each category must be representative of all data for that category; 2) the training data for each category must come close to fitting the distributional assumptions on which the decision rule is based. For example, point 1) says that if one category is wheat, and there are some dry and wet wheat fields which are distinguishably different on at least one image of the multi-image set, then data points from both kinds of wheat fields must be included in the training data for wheat. Point 2) says that if the histograms or scattergrams of wheat indicate that wheat has a bimodal distribution, one mode corresponding to the wet fields and one mode corresponding to the dry fields, and if the decision rule is to be based on some unimodal distributional assumption (such as normal or Gaussian), then for the purposes of the decision rule determination, the wheat category ought to be split into two subcategories: dry wheat and wet wheat. Then after the decision rule has made its identification assignments for the whole data set, the wet wheat and dry wheat subcategories can be collapsed to the one category, wheat.

The process of defining training set subcategories when using a unimodal distributional assumption is important not only for categories which have obvious subclasses but also for those categories which may not seem to have subclasses. This observation is prompted by the fact that in an agricultural context, multispectral data from a single crop field usually exhibit a unimodal structure, yet the pooled data from several fields of the same species often exhibit a multimodal character, the number of modes being nearly equal to the number of fields. Thus while the unimodal assumption might be appropriate for a single field it may not be appropriate for a group of fields.

Monte Carlo experimental results by WACKER [2.45] using the Gaussian classifier verify that increasing the number of subcategories defined for each category increases the average classifier accuracy over

many fields and in a pronounced way decreases the classifier variability on a field by field basis. In other words, having subcategory training sets when using a Gaussian classifier simultaneously increases identification accuracy over the entire image and more evenly or randomly distributes the identification errors.

A final item to which attention must be paid is the number of samples for each (sub)category in the training set. Parametric classifiers which make the unimodal distributional assumption should have a number of samples per category which at a minimum is between 3 and 10 times the dimensionality of the pattern vector [2.46]. Non-parametric classifiers such as a discrete table look-up rule can require orders of magnitude more samples per category than the parametric classifiers. KANAL and CHANDRASEKARAN [2.47], and FOLEY [2.46] discussed the relationship between sample size and dimensionality.

2.4.2 Decision Rule Determination

We first discuss how the training data determine a distribution-free decision rule, for it is here that we can focus on the essence of decision rules. Then we will describe some decision rules determined from certain kinds of distributional assumptions or approximations. In what follows we denote by "d" the typically 3- or 4-tuple multi-spectral feature vector and by "c" a land use category. The set of all features is D and the set of all categories is C.

To determine a distribution free decision rule, the first step is to estimate the set of conditional probabilities $\{P_d(c) | d \in D, c \in C\}$. $P_d(c)$ denotes the probability of category "c" being the true category identification of a small-area ground patch given that measurements of it generated feature vector "d". The conditional probability of category "c" given feature "d", $P_d(c)$, can be estimated by the proportion of those data points having true category identification "c" in that subset of the training data whose measurements generate feature "d".

Once these conditional probabilities have been computed, common sense should reveal the optimal decision rule. Consider the decision rule's problem when it tries to assign a category identification to a data resolution cell with feature vector or pattern "d". The conditional probabilities $P_d(c_1)$, $P_d(c_2)$,..., $P_d(c_K)$ have been estimated for categories $c_1, c_2, \ldots c_K$. To decide the category identification for a data point with feature "d" is an easy matter. Assign it to any category "c_k" having highest conditional probability; that is, assign it to a category "c_k" if and only if

$$P_d(c_k) \geqq P_d(c_i), \quad i = 1, 2, \ldots, K.$$

Such an optimal decision rule is called a simple Bayes rule.

Sometimes it is more convenient to estimate the set of conditional probabilities $\{P_c(d)|d\in D, c\in C\}$. $P_c(d)$ is the probability of obtaining a feature "d" from a small-area ground patch given that the small-area ground patch has true category identification "c". The conditional probability of feature "d" given category "c", $P_c(d)$, can be estimated by the proportion of training data points having feature "d" in the subset of training data points having true category identification "c". Using these conditional probabilities, the only logical assignment a decision rule can make of a data point with feature "d" is to assign it to any category "c_k" such that

$$P_{c_k}(d)\geqq P_{c_i}(d)\,, \qquad i=1, 2,\ldots, K\,.$$

Such a decision rule is called a maximum likelihood rule.

The relationship between the maximum likelihood rule and the simple Bayes rule is easy to develop. Let us denote by $P(d)$ the proportion of measurements in the training data having feature "d", by $P(c)$ the proportion of measurements in the training data having true category identification "c", and by $P(d, c)$ the proportion of measurements in the training data having feature "d" and true category identification "c". Then by definition of conditional probability

$$P_d(c)=\frac{P(d, c)}{P(d)}\,, \qquad P_c(d)=\frac{P(d, c)}{P(c)}\,. \tag{2.10}$$

By multiplying the inequality $P_d(c_k)\geqq P_d(c_i)$ on both sides by $P(d)$ we obtain

$$P_d(c_k)P(d)\geqq P_d(c_i)P(d)\,. \tag{2.11}$$

But, $P_d(c_k)P(d)=P(d, c_k)=P_{c_k}(d)P(c_k)$. Hence,

$$P_d(c_k)\geqq P_d(c_i) \quad \text{if and only if} \quad P_{c_k}(d)P(c_k)\geqq P_{c_i}(d)P(c_i)\,. \tag{2.12}$$

When the category prior probabilities are all equal, $P(c_j)=1/K$ $(j=1, 2, \ldots, K)$, we obtain $P_d(c_k)\geqq P_d(c_i)$ if and only if $P_{c_k}(d)\geqq P_{c_i}(d)$. Therefore, the simple Bayes rule and maximum likelihood rule are identical if the prior probabilities $P(c)$, $c\in C$, are all equal.

Use of a distribution-free rule in the digital computer is only possible under those circumstances in which it is possible to store the set of conditional probabilities. This storage possibility is strongly conditioned by the number of categories and the number of quantized levels to which

each component of the feature vector is expressed. For example, if there are 10 categories, and if the feature vector has 4 components, and each component is quantized to 5 levels, then a total of $10 \times 5^4 = 6250$ words is needed to store the conditional probabilities. This is a reasonable storage requirement. However, if each component were to be quantized to 10 levels, then a total of $10 \times 10^4 = 10^5$ words of storage are needed to store the conditional probabilities. This is usually an unreasonable storage requirement.

There are two approaches to handling the unreasonable storage problem: The table look-up approach reduces the storage requirement by careful storage and use of only essential information; the parametric approach assumes that the conditional probability $P_c(d)$ can be expressed by a formula having only a few parameters which need to be estimated. The table look-up approach uses more storage than the parametric approach but is much quicker in performing category assignments than the more computationally complex parametric approach. The two approaches are consistent with the observation that memory storage often can be traded for computational complexity.

Table Look-Up Approach

BROONER et al. [2.47a] used a table look-up approach on high altitude multiband photography flown over Imperial Valley, California, to determine crop types. Their approach to the storage problem was to perform an equal probability quantizing from the original 64 digitized grey levels to ten quantized levels for each of the three bands: green, red, and near infrared. Then after the conditional probabilities were empirically estimated, they used a Bayes rule to assign a category to each of the 10^3 possible quantized vectors in the 3-dimensional measurement space. Those vectors which occurred too few times in the training set for any category were deferred assignment. Figure 2.3 illustrates the decision regions associated with such a table look-up discrete Bayes decision rule. Notice how the quantized multispectral measurement vector can be used as an address in the 3-dimensional table to look up the corresponding category assignment.

The rather direct approach employed by BROONER et al. has the disadvantage of requiring a rather small number of quantization levels. Furthermore, it cannot be used with measurement vectors of dimension greater than four; for if the number of quantization levels is about 10, then the curse of dimensionality forces the number of possible quantized vectors to an unreasonably large size. Recognizing the grey level precision restriction forced by the quantizing coarsening effect, EPPLER et al. [2.48] suggested a way to maintain greater quantizing precision by defining a

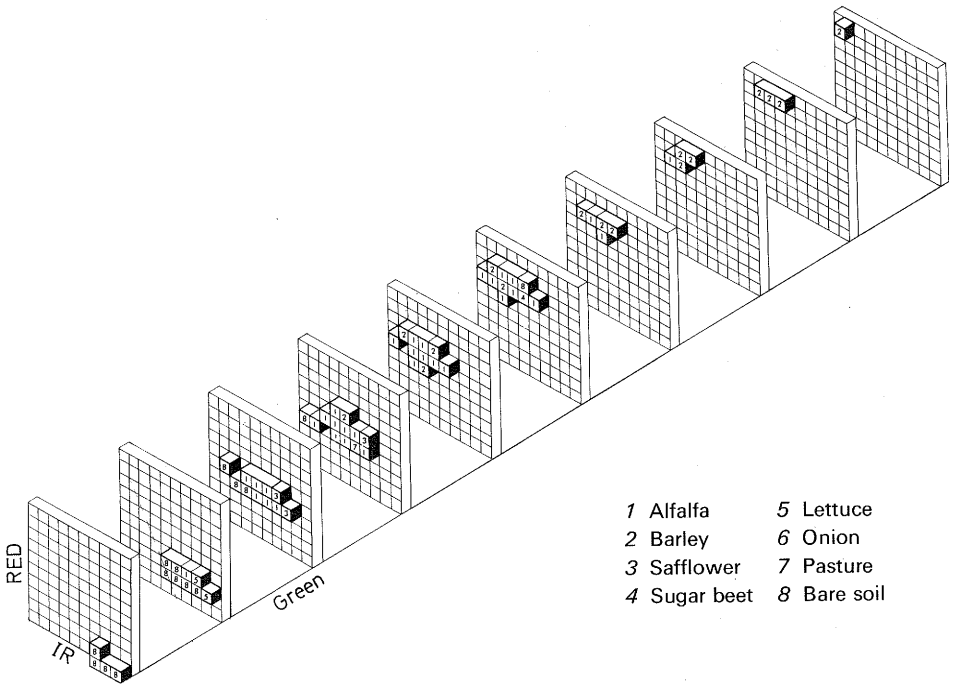

Fig. 2.3. Viewed as an expanded cube, each dimension representing the spectral region of the three multiband images whose density values have been quantized to ten equally spaced levels, this sketch depicts the decision rule boundaries for each land-use category used by the discrete Bayes rule

quantization rule for each category and measurement dimension as follows:

1) fix a category and a measurement dimension component;
2) determine the set of all measurement patterns which would be assigned by the decision rule to the fixed category;
3) examine all the measurement patterns in this set and determine the minimum and maximum grey levels for the fixed measurement component;
4) construct the quantizing rule for the fixed category and measurement dimension pair by dividing the range between the minimum and maximum grey levels into equal spaced quantizing intervals.

This multiple quantizing rule in effect determines for each category a rectangular parallelepiped in measurement space which contains all the measurement patterns assigned to it. Then, as shown in Fig. 2.4, the equal interval quantizing puts a grid over the rectangular parallelepiped.

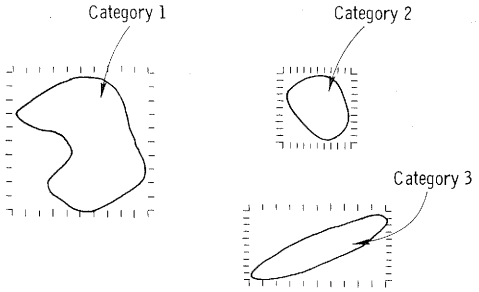

Fig. 2.4. Illustrates how quantizing can be done differently for each category, thereby enabling more accurate classification by the following table look-up rule: 1) quantize the measurement by the quantizing rule for category *1*, 2) use the quantized measurement as an address in a table and test if the entry is 1 or 0, 3) if it is 1 assign the measurement to category *1*; if it is 0, repeat the procedure for category *2*, etc.

Notice how for a fixed number of quantizing levels, the use of multiple quantizing rules in each band allows greater grey level quantizing precision compared to the single quantization rule for each band.

A binary table for each category can be constructed by associating each entry of the table with one corresponding cell in the gridded rectangular parallelepiped. Then we define the entry to be 1 if the decision rule has assigned a majority of the measurement patterns in the corresponding cell to the specified category; otherwise, we define the entry to be 0.

The binary tables are used in the implementation of the multiple quantization rule table look-up in the following way. Order the categories in some meaningful manner such as by prior probability. Quantize the multispectral measurement pattern using the quantization rule for category c_1. Use the quantized pattern as an address to look up the entry in the binary table for category c_1 to determine whether or not the prestored decision rule would assign the pattern to category c_1. If the decision rule makes the assignment to category c_1 the entry is 1 and all is finished. If the decision rule does not make the assignment to category c_1, the entry is 0 and the process repeats in a similar manner with the quantization rule and table for the next category.

One advantage of this form of the table look-up decision rule is the flexibility of being able to use different subsets of bands for each category look-up table and thereby take full advantage of the feature selecting capability to define an optimal subset of bands to discriminate one category from all the others. A disadvantage of this form of rule is the large amount of computational work required to determine the rect-

angular parallelepipeds for each category and the still large amount of memory storage required (about 5000 8-bit bytes per category).

EPPLER [2.49] discussed a modification of the table look-up rule which enables memory storage to be reduced by five times and decision rule assignment time to be decreased by two times. Instead of pre-storing in tables a quantized measurement space image of the decision rule, he suggested a systematic way of storing in tables the boundaries or end-points for each region in measurement space satisfying a regularity condition and having all its measurement pattern assigned to the same category.

Let $D = D_1 \times D_2 \times \ldots \times D_N$ be the measurement space. A subset $R \subseteqq D_1 \times D_2 \times \ldots \times D_N$ is a regular region if and only if there exist constants L_1 and H_1 and functions $L_2, L_3, \ldots, L_N, H_2, H_3, \ldots, H_N$

$$\begin{aligned} &(L_n: D_1 \times D_2 \times \ldots \times D_{n-1} \rightarrow (-\infty, \infty); \\ &H_n: D_1 \times D_2 \times \ldots \times D_{n-1} \rightarrow (-\infty, \infty)) \end{aligned} \tag{2.13}$$

such that

$$\begin{aligned} R = \{(x_1, \ldots, x_N) \in D | & L_1 \leqq x_1 \leqq H_1 \\ & L_2(x_1) \leqq x_2 \leqq H_2(x_1) \\ & \vdots \\ & L_N(x_1, x_2, \ldots, x_{N-1}) \leqq x_N \leqq H_N(x_1, x_2, \ldots, x_{N-1})\} \end{aligned} \tag{2.14}$$

From the definition of a regular region, it is easy to see how this table look-up by boundaries decision rule can be implemented. Let $d = (d_1, \ldots, d_N)$ be the measurement pattern to be assigned a category. To determine if d lies within a regular region R associated with category c we look up the numbers L_1 and H_1 and test to see if d_1 lies between L_1 and H_1. If so, we look up the numbers $L_2(d_1)$ and $H_2(d_1)$ and so on. If all the tests are satisfied, the decision rule can assign measurement pattern d to category c. If one of the tests fails, tests for the regular region corresponding to the next category can be made.

The memory reduction in this kind of table look-up rule is achieved by only storing boundary or end-points of decision regions, and the speed-up is achieved by having one-dimensional tables whose addresses are easier to compute than the three- or four-dimensional tables required by the initial table look-up decision rule. However, the price paid for these advantages is the regularity condition imposed on the decision regions for each category. This regularity condition is stronger than set connectedness but weaker than set convexity.

Another approach to the table look-up rule can be based on ASHBY's [2.50] technique of constraint analysis. ASHBY suggested representing

in an approximate way subsets of Cartesian product sets by their projections on various smaller dimensional spaces. Using this idea for two-dimensional spaces we can formulate the following kind of table look-up rule.

Let $D = D_1 \times D_2 \times \ldots \times D_N$ be the measurement space, C be the set of categories, and $J \subseteq \{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\}$ be an index set for the selected two-dimensional spaces. Let the probability threshold α be given. Let $(i, j) \in J$; for each $(x_1, x_2) \in D_i \times D_j$ define the set $S_{ij}(x_1, x_2)$ of categories having the highest conditional probabilities given (x_1, x_2) by $S_{ij}(x_1, x_2) = \{c \in C | P_{x_1, x_2}(c) \geqq \alpha_{ij}\}$, where α_{ij} is the largest number which satisfies

$$\sum_{c \in S_{ij}(x_1, x_2)} P_{x_1, x_2}(c) \geqq \alpha \,. \tag{2.15}$$

$S_{ij}(x_1, x_2)$ is the set of likely categories given that components i and j of the measurement pattern take the values (x_1, x_2).

The sets S_{ij}, $(i, j) \in J$, can be represented in the computer by tables. In the (i, j)-th table S_{ij} the (x_1, x_2)-th entry contains the set of all categories of sufficiently high conditional probability given the marginal measurements (x_1, x_2) from measurement components i and j, respectively. This set of categories is easily represented by a one-word table entry; a set containing categories c_1, c_7, c_9, and c_{12}, for example, would be represented by a word having bits 1, 7, 9, and 12 on and all other bits off.

The decision region $R(c)$ containing the set of all measurement patterns to be assigned to category c can be defined from the S_{ij} sets by

$$\begin{aligned} R(c) = \{&(d_1, d_2, \ldots, d_N) \in D_1 \times D_2 \times \ldots \times D_N | \\ &\{c\} = \bigcap_{(i,j) \in J} S_{ij}(d_i, d_j)\} \,. \end{aligned} \tag{2.16}$$

This kind of table look-up rule can be implemented by using successive pairs of components (defined by the index set J) of the (quantized) measurement patterns as addresses in the just mentioned two-dimensional tables. The set intersection required by the definition of the decision region $R(c)$ is implemented by taking the Boolean AND of the words obtained from the table look-up for the measurement to be assigned a category. Note that this Boolean operation makes full use of the natural parallel computation capability that the computer has on bits of a word. If the k-th bit is the only bit which remains on in the resulting word, then the measurement pattern is assigned to category c_k. If there is more than one bit on, or no bits are on, then the measurement pattern is deferred assignment (reserved decision). Thus we see that this form of a table look-up rule utilizes a set of "loose" Bayes rules in the lower dimensional projection spaces and intersects the resulting multiple category assign-

ment sets to obtain a category assignment for the measurement pattern in the full measurement space.

Because of the natural effect which the category prior probabilities have on the category assignments produced by a Bayes rule it is possible for a measurement pattern to be the most probable pattern for one category yet be assigned by the Bayes rule to another category having much higher prior probability. This effect will be pronounced in the table look-up rule just described because the elimination of such a category assignment from the set of possible categories by one table look-up will completely eliminate it from consideration because of the Boolean AND or set intersection operation. However, by using an appropriate combination of maximum likelihood and Bayes rules, something can be done about this.

For any pair (i, j) of measurement components, fixed category c, and probability threshold β, we can construct the set $T_{ij}(c)$ of the most probable pairs of measurement values from components i and j arising from category c. The set $T_{ij}(c)$ is defined by

$$T_{ij}(c)=\{(x_1, x_2)\in D_i \times D_j | P_c(x_1, x_2)\geqq \beta_{ij}(c)\}\,, \tag{2.17}$$

where $\beta_{ij}(c)$ is the largest number which satisfies

$$\sum_{(x_1,x_2)\in T_{ij}(c)} P_c(x_1, x_2)\geqq \beta\,.$$

Tables which can be addressed by (quantized) measurement components can be constructed by combining the S_{ij} and T_{ij} sets. Define $Q_{ij}(x_1, x_2)$ by

$$Q_{ij}(x_1, x_2)=\{c\in C | (x_1, x_2)\in T_{ij}(c)\}\cup S_{ij}(x_1, x_2)\,. \tag{2.18}$$

The set $Q_{ij}(x_1, x_2)$ contains all the categories whose respective conditional probabilities given measurement values (x_1, x_2) of components i and j are sufficiently high (a Bayes rule criterion) as well as all those categories whose most probable measurement values for components i and j, respectively, are (x_1, x_2) (a maximum likelihood criterion). A decision region $R(c)$ containing all the (quantized) measurement patterns can then be defined as before using the Q_{ij} sets

$$\begin{aligned} R(c)=\{(d_1, d_2, \ldots, d_N)\in D_1 \times D_2 \ldots \times D_N | \\ \{c\}=\bigcap_{(i,j)\in J} Q_{ij}(d_i, d_j)\}\,. \end{aligned} \tag{2.19}$$

A majority vote version of this kind of table look-up rule can be defined by assigning a measurement to the category most frequently

selected in the lower dimensional spaces

$$R(c)=\{(d_1, d_2, \ldots, d_N)\in D_1\times D_2\times\ldots\times D_N| \\ \#\{(i,j)\in J|c\in Q_{ij}(d_i, d_j)\}\geqq \#\{(i,j)\in J|c'\in Q_{ij}(d_i, d_j) \\ \text{for every}\quad c'\in C-\{c\}\}\}\,. \tag{2.20}$$

Multivariate Normal Assumption

The other approach to easing the storage requirement of the decision rule entails making assumptions about the form the conditional probability $P_c(d)$ takes. By expressing $P_c(d)$ as a formula with only a few parameters which have to be estimated, storage is saved and computational complexity is increased. The most frequent distributional assumption made is the N-dimensional multivariate normal or Gaussian one. Here, the conditional probability is a unimodal function and represented as a formula of the form

$$P_c(d)=\frac{1}{(2\pi)^{N/2}|\Sigma_c|^{1/2}}\,e^{-(x-\mu_c)'\Sigma_c^{-1}a(x-\mu_c)/2}\,, \tag{2.21}$$

where μ_c is the mean vector of category c, and Σ_c is the covariance matrix for category c. The maximum likelihood rule then takes the form: a small-area ground patch having feature "d" is assigned to any category "c_k" where

$$\frac{1}{(2\pi)^{N/2}|\Sigma_{c_k}|^{1/2}}\,e^{-(x-\mu_{c_k})'\Sigma_{c_k}^{-1}(x-\mu_{c_k})/2} \\ \geqq \frac{1}{(2\pi)^{N/2}|\Sigma_{c_i}|^{1/2}}\,e^{-(x-\mu_{c_i})'\Sigma_{c_i}^{-1}(x-\mu_{c_i})/2}\,. \qquad i=1,2,\ldots,K\,. \tag{2.22}$$

The simple Bayes rule takes the form: a small-area ground patch having feature "d" is assigned to any category "c_k" where

$$\frac{P(c_k)}{(2\pi)^{N/2}|\Sigma_{c_k}|^{1/2}}\,e^{-(x-\mu_{c_k})'\Sigma_{c_k}^{-1}(x-\mu_{c_k})/2} \\ \geqq \frac{P(c_i)}{(2\pi)^{N/2}|\Sigma_{c_i}|^{1/2}}\,e^{-(x-\mu_{c_i})'\Sigma_{c_i}^{-1}(x-\mu_{c_i})/2}\,. \qquad i=1,2,\ldots,K\,. \tag{2.23}$$

Taking logarithms on both sides of the inequality we can see that the Bayes rule under a multivariate normal assumption takes the form of a quadratic computation. When the covariance matrices are all equal,

there is considerable simplification in the maximum likelihood rule which then takes a linear form: a small-area ground patch having feature "d" is assigned to any category "c_k" where

$$a'_{ki}d - t_{ki} \geqq 0, \quad \text{for each} \quad i = 1, 2, \ldots, K, \tag{2.24}$$

where

$$a_{ki} = \Sigma^{-1}(\mu_{c_k} - \mu_{c_i})$$

and

$$t_{ki} = (1/2)(\mu_{c_k} - \mu_{c_i})'\Sigma^{-1}(\mu_{c_k} + \mu_{c_i}).$$

The functions $f_{ki}(d) = a_{ki}d - t_{ki}$, $i = 1, 2, \ldots, K$, $k = 1, 2, \ldots, K$ are called linear discriminant functions. More detailed discussions of linear and quadratic decision rules may be found in [2.51] and in the review article by NAGY [2.52]. The book by SEBESTYEN [2.53] is also a good general reference. See ANDERSON and BAHADUR [2.54] for a discussion of the optimal linear decision rule under the multivariate normal assumption for unequal covariance matrices.

2.4.3 Feature Selection

Multi-image data sets, whether stored in analog form on video tape or on film, or stored in digital form on digital magnetic tape, typically contain on the order of 10^6 distinctly resolvable data points with each data point being an N-tuple, $N = 3$ for color film and up to $N = 12$ or more for a multispectral scanner. Most processing algorithms treat each data point independently and this implies that there are on the order of 10^6 separate decisions which the decision rule must make. In order to enable the decision rule to do its job quickly and simply, it is prudent to give the decision rule the least amount of simple non-redundant information which enables it to do its job well. The process of selecting what information to present to the decision rule is called feature selection.

Feature selection algorithms for multi-image data sets are usually based on the fact that the images in the set contain highly redundant information. This is easily understood if we consider the fact that images in neighboring spectral bands taken at the same time or images taken within short time intervals (a few days or weeks) in the same spectral band are highly correlated. Feature selection can then be thought of as choosing that set of, say, three or four images of the multi-image set which minimize the errors which the decision rule makes.

Let I be an index set for the different bands of a multi-image. Let n, the number of bands we wish the decision rule to work with, be given.

Then for each subset S of I indexing exactly n bands of the multi-image, application of a decision rule will produce a probability of error $P_e(S)$. $P_e(S)$ is the probability that the decision rule will make an incorrect assignment on some data point when allowed to use only those bands indexed by S. Clearly, the best set of bands to use is that set S which minimizes the probability of error; that is,

$$P_e(S) \leqq P_e(S') \quad \text{for all} \quad S' \subseteq I \quad \text{satisfying} \quad \# S' = n .$$

It often is the case that the error probabilities are difficult to compute. This is especially so when the domains of the parametric probability density functions are large dimensional spaces. Thus measures have been sought which are easier to compute for parametric functions and which are closely related to probability of error.

If we let $P(i, j; S)$ be the joint probability that a data point from category j will be assigned by the decision rule to category i when the decision rule is only allowed to use the n bands indexed by the set S, then the probability of error $P_e(S)$ can be written as the sum of all the above joint probabilities for which $i \neq j$:

$$P_e(S) = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} P_e(i, j; S) . \tag{2.25}$$

If we consider the probability of error to be the probability of error incurred by a Bayes decision rule, then dropping the notational dependence on the index set S we may express $P_e(i, j)$ by

$$P_e(i, j) = \sum_{g \in A_i} P(g, c_j) , \tag{2.26}$$

where $P(g, c_j)$ is the joint probability of a pattern being in class c_j and having value g and

$$A_i = \{g \in G | P(g, c_i) \geqq P(g, c_n), n = 1, 2, \ldots, K\} . \tag{2.27}$$

Noticing that the set B_{ij} defined by

$$B_{ij} = \{g \in G | P(g, c_i) \geqq P(g, c_j)\} \tag{2.28}$$

contains the set of points A_i we must have

$$P_e(i, j) = \sum_{g \in A_i} P(g, c_j) \leqq \sum_{g \in B_{ij}} P(g, c_j) . \tag{2.29}$$

Hence, $P_e(i,j)+P_e(j,i) \leqq \sum_{g\in B_{ij}} P(g,c_j)+\sum_{g\in B_{ji}} P(g,c_i)=\sum_{g\in G}\min\{P(g,c_j), P(g,c_i)\}$. Since $\min\{a,b\}\leqq\sqrt{ab}$, we have

$$P_e(i,j)+P_e(j,i)\leqq \sum_{g\in G}\sqrt{P(g,c_i)P(g,c_j)} \leqq \sqrt{P(c_i)P(c_j)}\sum_{g\in G}\sqrt{P_{c_i}(g)P_{c_j}(g)}\,. \tag{2.30}$$

The number

$$\varrho_{ij}=\sum_{g\in G}\sqrt{P_{c_i}(g)P_{c_j}(g)} \tag{2.31}$$

is called the pairwise Bhattacharyya coefficient for categories c_i and c_j and provides an upper bound on the error probability [2.55]

$$P_e(S)\leqq \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sqrt{P(c_i)P(c_j)}\varrho_{ij}(S)\leqq \tfrac{1}{2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\varrho_{ij}(S)\,. \tag{2.32}$$

Using this inequality, one way of choosing the feature index set S having n elements is to compute the upper bound over all possible index sets having n elements and choose the one having the smallest upper bound. When the number of bands is large, this exhaustive search is rather expensive in terms of the number of possibilities it must account for. A suboptimal procedure is to construct S iteratively. For the first element of S, choose that feature which alone yields the lowest upper bound on error probability. For the next element of S, take that feature which when used with the previously chosen features yields the lowest upper bound on error probability, and so on [2.56].

For categories having normal distributions, ϱ_{ij} can be computed in terms of the mean and covariance matrices

$$-\ln\varrho_{ij}=\tfrac{1}{8}(\mu_i-\mu_j)'\Sigma^{-1}(\mu_i-\mu_j) + \frac{1}{2}\ln\frac{|\Sigma|}{|\Sigma_i|\,|\Sigma_j|}, \tag{2.33}$$

where $\Sigma=\frac{1}{2}(\Sigma_i+\Sigma_j)$.

The Battacharyya coefficient is not the only measure of category probability distribution separability employed to find bounds on error probabilities. The use of divergence to measure separability between distributions was first suggested by KULLBACK [2.57]. MARILL and GREEN [2.58] proposed it for use in recognition systems, and KAILATH [2.59] suggested it could be used in signal processing applications and showed its close relationship to probability of error in pattern discrimination problems. The divergence between two conditional probability

functions for categories c_i and c_j is the number d_{ij} defined by

$$d_{ij} = \sum_{g \in G} [P_{c_i}(g) - P_{c_j}(g)] \ln \frac{P_i(g)}{P_j(g)}. \tag{2.34}$$

For categories having normal distributions, the divergence d_{ij} can be computed in terms of the mean and covariance matrices:

$$\begin{aligned} d_{ij} = {} & \tfrac{1}{2} \operatorname{trace}(\Sigma_i - \Sigma_j)(\Sigma_j^{-1} - \Sigma_i^{-1}) \\ & + \tfrac{1}{2}(\mu_i - \mu_j)'(\Sigma_i^{-1} + \Sigma_j^{-1})(\mu_i - \mu_j). \end{aligned} \tag{2.35}$$

Under the normal assumption, the Bhattacharyya coefficient and divergence are related by an inequality

$$\varrho_{ij} \geqq e^{-d_{ij}/8}, \quad \text{with equality when} \quad \Sigma_i = \Sigma_j,$$

so that

$$\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \varrho_{ij} \geqq \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} e^{-d_{ij}/8}. \tag{2.36}$$

Since $P_e(S) \leqq \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \varrho_{ij}(S)$,

the inequalities go the wrong way to let the divergence establish a bound on the error probability. Nevertheless, the criteria of maximizing the minimum of the pairwise divergence or maximizing the average divergence d_{ave} have been proposed [2.60–62]

$$d_{ave} = \frac{2}{K(K-1)} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} d_{ij}. \tag{2.37}$$

Comparison between the use of the Bhattacharyya coefficient and the average divergence for band or feature selection reveals the superiority of the Bhattacharyya coefficient [2.63]. SWAIN and KING [2.63a] have observed that minimizing the average of the transformed divergence d_T defined by

$$d_T = \frac{2}{K(K-1)} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} e^{-d_{ij}/8} \tag{2.38}$$

is a better feature selection criterion than minimizing the average divergence. The exponential form is, of course, motivated by the form of the inequality relating divergence and the Bhattacharyya coefficient.

FU et al. [2.14] discussed a general separability measure which relates more closely than divergence does to the probability of error

achievable by a minimax linear decision rule under a multivariate normal assumption. They define the separability measure S_{ij} between the i-th and j-th classes by

$$S_{ij}=\frac{b'_{ij}(\mu_i-\mu_j)}{(b'_{ij}\Sigma_i b_{ij})^{1/2}+(b'_{ij}\Sigma_j b_{ij})^{1/2}}, \tag{2.39}$$

where the vector b_{ij}, the difference between the i-th and j-th linear discriminant functions, takes the form

$$b_{ij}=(\lambda_{ij}\Sigma_i+(1-\lambda_{ij})\Sigma_j)^{-1}(\mu_i-\mu_j) \tag{2.40}$$

and λ_{ij} is the Lagrange multiplier satisfying

$$b'_{ij}(\lambda_{ij}^2\Sigma_i-(1-\lambda_{ij})^2\Sigma_j)b_{ij}=0, \quad 0\leqq\lambda_{ij}\leqq 1. \tag{2.41}$$

Then

$$P_e(S)\leqq\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}P_e(i,j)\leqq\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}P(c_i)P(c_j)\int_{S_{ij}}^{\infty}e^{-x^2/2}dx \tag{2.42}$$

so that the separability measure can be used as a bound on the error probability achievable by a minimax linear decision rule and feature selection can be done by choosing features which minimize

$$\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}P(c_i)P(c_j)\int_{S_{ij}}^{\infty}e^{-x^2/2}dx. \tag{2.43}$$

A different form, slightly easier to compute than the error bound indicated above, is the average separability which Fu et al. [2.14] suggested maximizing

$$\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}P(c_i)P(c_j)S_{ij}. \tag{2.44}$$

Other distance or distance-like separability measures suggested for feature selection have been:

1) The Jeffreys-Matusita distance defined by [2.64, 65]

$$\{\sum_{g\in G}[\sqrt{P_{c_i}(g)}-\sqrt{P_{c_j}(g)}]^2\}^{1/2}. \tag{2.45}$$

2) The Kolmogorov variational distance defined by [2.59]

$$\tfrac{1}{2}\sum_{g\in G}|P(g,c_i)-P(g,c_j)|. \tag{2.46}$$

3) The Kullback-Liebler number defined by [2.66]

$$\sum_{g\in G}\left[\ln\frac{P_{c_i}(g)}{P_{c_j}(g)}\right]P_{c_i}(g)\,. \tag{2.47}$$

4) The Mahalanobis distance defined by [2.67]

$$(\mu_i-\mu_j)'\sum{}^{-1}(\mu_i-\mu_j)\,. \tag{2.48}$$

5) The Euclidean distance defined by

$$\{\sum_{g\in G}[P_{c_i}(g)-P_{c_j}(g)]^2\}^{1/2}\,. \tag{2.49]}$$

The Euclidean distance has been related to an upper bound on probability of error [2.68] which is looser than the upper bound obtained by the Bhattacharyya coefficient. The Kolmogorov variational distance is intimately related to probability of error [2.59]

$$P_e(i,j)=\tfrac{1}{2}(P(c_j)+P(c_i))-\tfrac{1}{2}\sum_{g\in G}|P(g,c_i)-P(g,c_j)|\,. \tag{2.50}$$

In this subsection we have used the criterion that the best features are those which minimize the error probability. Clearly this is an optimal criterion for selecting which bands to use and which ones not to use. A different kind of feature selection problem can be posed by asking what are the combinations of the multispectral bands that can best represent, in a mean square error sense, the information contained in all the bands. In this kind of problem a function is sought which represents each measurement pattern in a compressed form in a smaller dimensional space in such a way that if the original measurement pattern were to be reconstructed from the compressed measurement, the average squared error between the reconstruction and the original would be minimized. The transformation which achieves this minimum turns out to be linear and is obtained as the $K\times N$ matrix whose rows are the K eigenvectors with corresponding largest eigenvalues of the second moment matrix of the entire data set. The representation is called the principal components representation ([2.69]). VIGLIONE [2.70] discussed a feature selection process which uses normality assumptions to determine optimum quadratic discriminant functions in the subspace, and then an interactive process based on minimizing an exponential loss function using sample pattern separation to select, evaluate, and weight the quadratic discriminant features and form a piecewise quadratic discriminant boundary in the original signal space.

Principal component feature selection has the advantage that it does not depend on the choice of categories, and although examples can be constructed illustrating how minimum mean square error may not preserve distinctions between categories, READY et al. [2.71] provided evidence that this tends not to happen with remote sensing spectral data. They indicated that 3 or 4 principal component features may be used from 12-channel multispectral data with essentially no decrease in the probability of correct category identification. HARALICK and DINSTEIN [2.72] used principal components as a feature selection method for their iterative clustering procedure applied to remote sensing multispectral scanner imagery. READY and WINTZ [2.73] indicated that spectral feature selection by principal components results in feature patterns with a better signal to noise ratio than in the original image bands. TAYLOR [2.74] described how the principal components representation can be used to enhance false color combinations of multiband imagery.

2.4.4 Estimating Decision Rule Error

Once a decision rule has been determined, it is important to estimate the percentage of errors which the decision rule makes. To do this another subset of data from the multi-image must be chosen and each point in this data set must be labeled with its true category identification. This data set is called the "prediction data", and for the error estimates to be accurate, the prediction data set must be 1) as representative as the training data set was required to be, and 2) independent of the training data.

For any given category "c" there are two types of errors which the decision rule can make: it can assign a small-area ground patch whose true category identification is "c" to some category $c' \neq c$, an error of misidentification or misdetection for category "c"; or, it can assign a small-area ground patch whose true category identification is "c'", $c' \neq c$, to category "c", an error of false detection for category "c". Both these kinds of errors are easily estimated by applying the decision rule to the prediction data and comparing the true category identification of each point with the category identification assigned by the decision rule.

Let us denote by P_{ij} the proportion of points in the prediction data having true category identification "c_i" and decision rule assigned identification "c_j". The total probability of correct identification can be estimated by

$$\sum_{i=1}^{K} P_{ii} \Big/ \sum_{j=1}^{K} \sum_{i=1}^{K} P_{ij} . \tag{2.51}$$

Given that a point has true category identification "c_i", the probability that it will be misidentified is

$$\sum_{\substack{j=1 \\ j \neq i}}^{K} P_{ij} \Big/ \sum_{j=1}^{K} P_{ij} . \tag{2.52}$$

Given that a point has true category identification "c_k", $c_k \neq c_i$, the probability that it will be falsely identified as category "c_i" can be estimated by

$$\sum_{\substack{j=1 \\ j \neq i}}^{K} P_{ji} \Big/ [1 - \sum_{j=1}^{K} P_{ij}] . \tag{2.53}$$

Sometimes a more easily interpreted false identification error probability can be computed as $\sum_{\substack{j=1 \\ j \neq i}}^{K} P_{ji} \Big/ \sum_{j=1}^{K} P_{ji}$ which is the proportion of those small-area ground patches falsely identified as category "c_i" out of the set of all small-area ground patches identified correctly or incorrectly as category "c_i".

2.4.5 Clustering[1]

We introduce our discussion of clustering by contrasting it to the decision rule algorithms used in pattern discrimination. With pattern discrimination techniques, a training set of data is gathered for which the correct category identification of each distinct entity in the data is known. Then estimates are made of the required category conditional probability distributions and a decision rule is determined from them. The decision rule can then be employed to identify any other data set gathered under similar conditions. With clustering techniques there is no training data set or decision rule. Rather, natural data structures are determined. Distinct structures are then interpreted as corresponding to distinct objects or environmental processes.

The advantage of the discrimination techniques is that the scientist is able to decide the types of environmental categories among which he wishes to distinguish. The decision rule then determines, as well as possible, to which environmental category an arbitrary datum belongs. The disadvantage of the discrimination techniques is that they are sensitive to mis-calibrations which cannot be normalized out. Any difference between the sensor calibrations or state of environment for the training data and the new data will cause error.

[1] See also [2.75].

The advantage of the clustering techniques is that they are not sensitive to calibration errors which shift or stretch measurement space in minor ways. Two small-area patches of corn growing in the same field will be detected as being similar because they have similar grey tones associated with them. The disadvantage of the clustering techniques is that they are not able to name the distinct environment structures they determine.

A wide variety of iterative procedures can be used to cluster remotely sensed image data and there is no general theory of clustering [2.76–79]. Each scheme involves grouping together of similar data. Perhaps the most popular clustering technique for remotely sensed data is the *K*-means or ISODATA technique [2.80–83]. Here each data point is put into that cluster for which the squared distance between it and the cluster mean is least. Then new cluster means are computed and the whole procedure repeated. On each repetition the total squared distance between the data points and the cluster means is guaranteed not to increase. There are numerous variants concerned with what to do when a cluster is too big, too small, etc.

Many clustering procedures deal with measurements in ways that do not consider the natural order or arrangement of the measurements. Indeed, for many problems, the order in which the measurements are taken or the spatial distribution of the units under consideration are irrelevant to the clustering process. This is not the case for remotely sensed image data. When the resolution of the image data is properly selected, each object of interest yields many measurements. Therefore, resolution cells in the same neighborhood are likely to belong to the same object. We call procedures that take into account these kinds of spatial relations between neighboring resolution cells spatial clustering procedures.

Some of the spatial clustering ideas being explored by investigators in the remotely sensed data area closely parallel those of investigators in the artificial intelligence community, which has been an active user of spatial information from scene data. Much artificial intelligence work has gone into the definition of homogeneous region and into edge detection. BRICE and FENNEMA [2.84] described a procedure for partitioning an image into a large set of primitive regions each of which is typically a connected component having the same grey tone. Then a merging algorithm can be applied to group together those regions having similar tone. Gradient and derivative algorithms can be used to find edges [2.85–88]. So we see that scene analysis and spatial clustering can have much in common.

HARALICK and KELLY [2.89] discussed a spatial clustering procedure designed to take into account the spatial dependencies which nearby

pixels have in remotely sensed imagery. Spatial clusters are grown around those resolution cells having grey tone n-tuples of high relative frequency in their own local neighborhoods. The spatial growing proceeds by taking in resolution cells spatially neighboring those resolution cells already in the cluster and having grey tone n-tuples which neighbor in measurement space some of the grey tone n-tuples of the resolution cells already in the cluster.

NAGY et al. [2.90] based a spatially oriented clustering procedure on a chaining idea proposed by BONNER [2.47]. In the first step, points are assembled into row strips. This is based upon the assumption that "spatially adjacent grey tone n-tuples tend to belong to the same type of ground cover except at field boundaries". Examining the resolution cells along the scan lines, similar resolution cells are assigned to strips. Each strip is terminated only when the addition of the grey tone n-tuple of the next resolution cell would increase the internal scatter of the strip above a given threshold. At that point, the formed strip is assigned to a cluster (or designated to start a new cluster), and the formation of a new strip begun. The assignment of a strip to a cluster is done by a measurement space comparison of the strip to the cluster centers. The search for a cluster is done in decreasing order of cluster size in order to save computation time and to eliminate small groups of abnormal grey tone n-tuples.

JAYROE [2.91] introduced a three step spatial clustering procedure for multi-images. Two gradient images are obtained by computing the measurement space Euclidean distance between nearest neighbors in the horizontal and vertical directions. A boundary map is then prepared by thresholding of gradient images. In the second step, clusters are formed by scanning the boundary map with a fixed size square of resolution cells. When the square hits a region in which there are no boundary cells, that region is assigned to cluster 1. The square is then moved farther, and if no boundary cells are encountered, all the resolution cells within the square are assigned to cluster 1. The scanning continues until all possible cells are assigned to that cluster. Next, the square is moved until it hits a new region with no boundary cells, and the process is repeated. In the third step, clusters are merged according to their spectral measurements.

ROBERTSON et al. [2.92] describe a procedure for successively partitioning a multi-image into rectangular blocks. However, the final clustered images produced by the algorithm yield images which suffer from excessive blockiness.

HARALICK and DINSTEIN [2.72], like JAYROE [2.91], discussed a spatial clustering procedure based on gradient images. In the first step the gradient image is computed, thresholded, and cleaned, thereby

creating a binary image showing homogeneous areas. In the second step, the maximally connected spatial regions of the cleaned image are computed. In the third step a measurement space clustering procedure is applied to the distinctly labelled connected regions to determine spatial regions of similar spectral character.

2.5 Image Texture

Spatial environments can be understood as being spatial distributions of various area-extensive objects having characteristic size and reflective or emissive qualities. The spatial organization and relationships of the objects appear as spatial distributions of grey tone on imagery taken of the environment. We call the pattern of spatial distributions of grey tone, texture.

Figure 2.5, taken from [2.93], illustrates how texture relates to geomorphology. It shows plains, low hills, high hills, and mountains in the Panama and Columbia area, as seen by the Westinghouse AN/APQ97 *K*-band radar imaging system. The plains have an apparent relief of 0–50 m, the hills 50–350 m, and the mountains more than 350 m. The low hills have little dissection and are generally smooth convex surfaces, whereas the high hills are highly dissected and have prominent ridge crests.

The mountain texture is distinguishable from the hill texture on the basis of the extent of radar shadowing (black tonal areas). The mountains have shadowing over more than half the area and the hills have shadowing over less than half the area. The hills can be ranked from low to high on the basis of the abruptness of tonal change from terrain front slope to terrain back slope.

Figure 2.6, taken from [2.94], illustrates how texture relates to geology. It shows igneous and sedimentary rocks in Panama, as seen by the same radar system. Figure 2.6i, k, l show a fine-textured drainage pattern which is indicative of non-resistant fine-grained sedimentary rocks. The coarser texture of Figure 2.6h, left half, is indicative of coarser-grained sediments. A massive texture with rugged and peaked divides (Fig. 2.6a–e) is indicative of igneous rocks. When erosion has nearly base-leveled an area, the texture takes on the hummocky appearance of Fig. 2.6g.

Figure 2.7, taken from [2.95], illustrates how texture relates to land use categories. Here, there are six land use categories as they appear on panchromatic aerial photography. Notice how the texture of the wooded area is coarser and more definite than that of the scrub area. The swamp

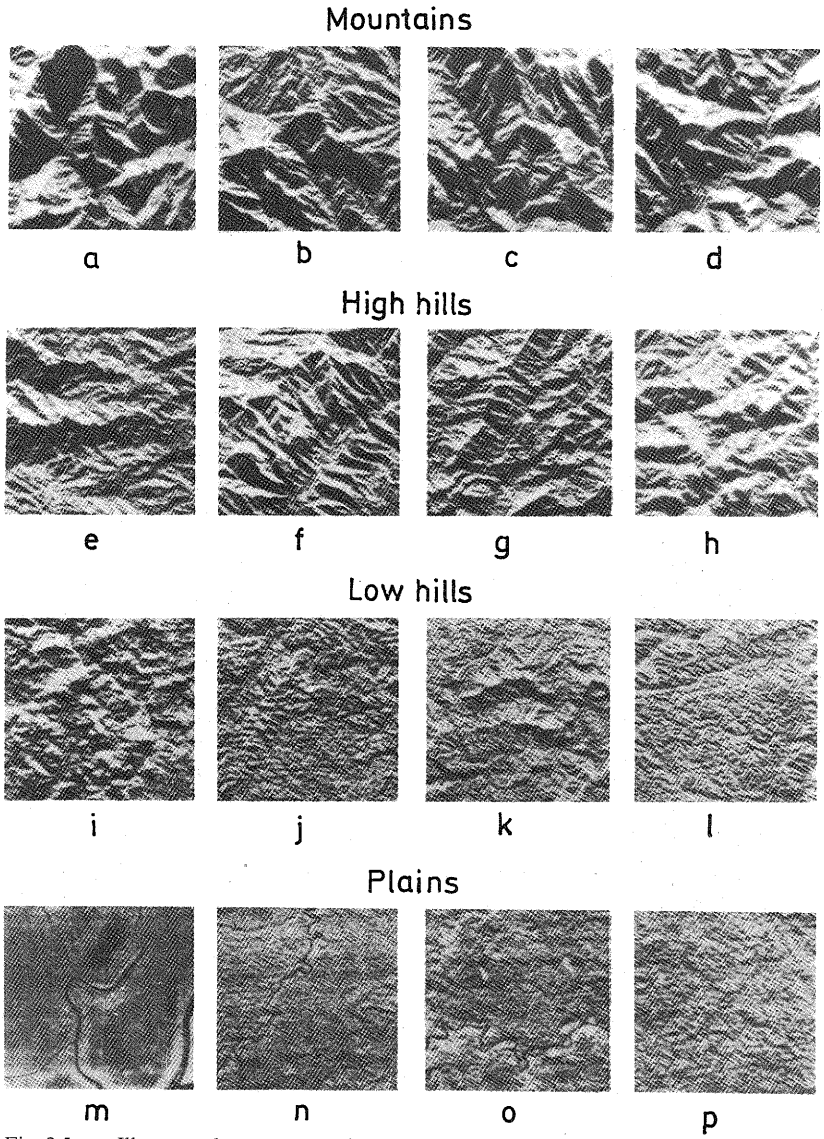

Fig. 2.5a–p. Illustrates how texture relates to geomorphology. (Taken from [2.93])

and marsh generate finer textures than the wood or scrub areas. The swamp texture is finer than and shows more gradual grey tone change than the marsh texture.

Figure 2.8 is taken in the Pisgah Crater area and shows some examples where the same type of terrain generates a variety of textures within the

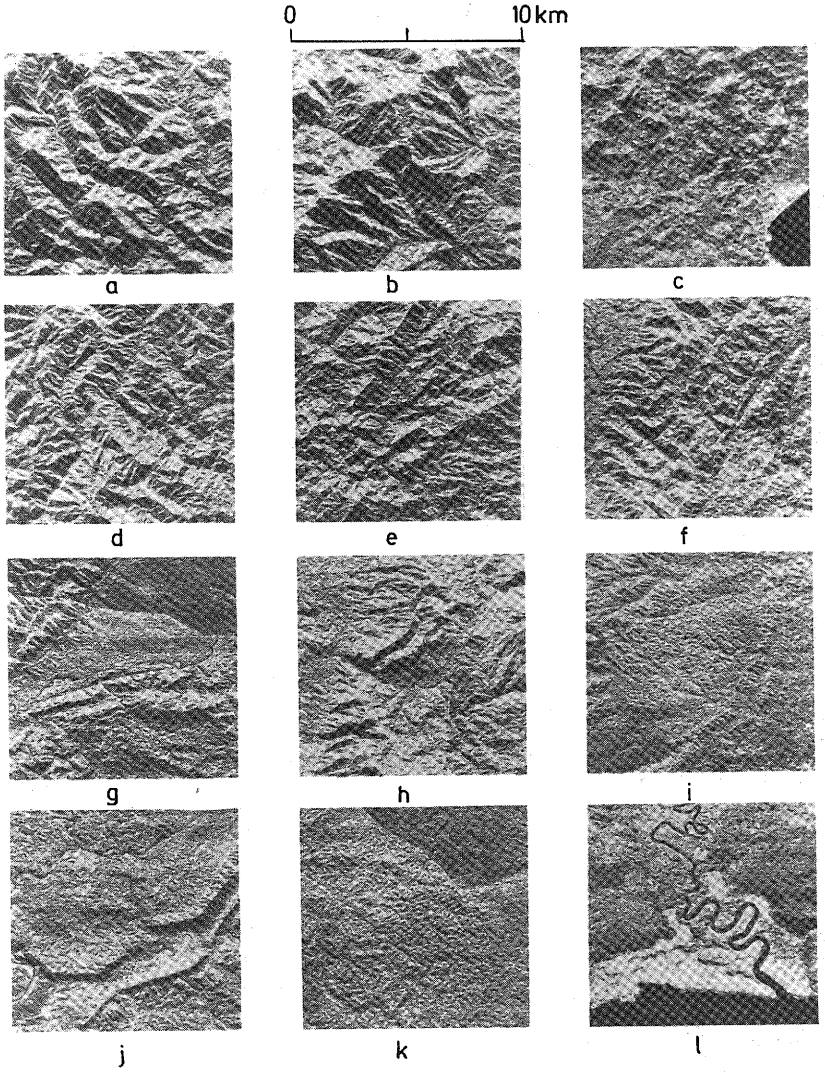

Fig. 2.6a–l. Illustrates how texture relates to geology. (Taken from [2.94])

same texture family. Here, the texture changes are due to the way the vegetation increases in size and disperses.

There have been six basic approaches to the measurement and characterization of image texture: autocorrelation functions [2.96], optical transforms [2.97], digital transforms [2.98–100], edgeness [2.88],

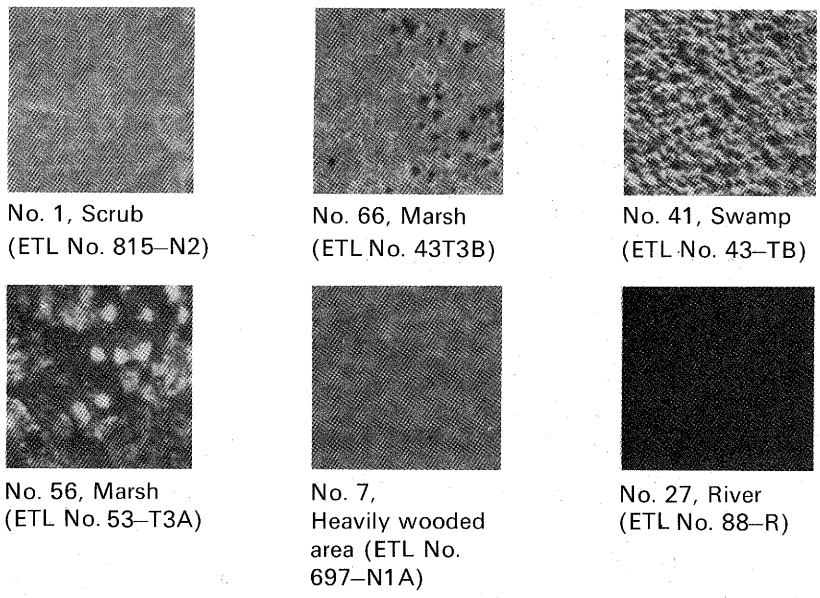

Fig. 2.7. Illustrates how texture relates to land use categories. (Taken from [2.95])

structural elements [2.101, 102], and spatial grey tone co-occurrence probabilities [2.103, 104]. The first three of these approaches are related in that they all measure spatial frequency directly or indirectly. Spatial frequency is related to texture because fine textures are rich in high spatial frequencies while coarse textures are rich in low spatial frequencies.

An alternative to viewing texture as spatial frequency distribution is to view texture as amount of edge per unit area. Coarse textures have a small number of edges per unit area, while fine textures have a high number of edges per unit area.

The structural element approach uses a matching procedure to detect the spatial regularity of shapes called structural elements in a binary image. When the structural elements themselves are single resolution cells, the information provided by this approach is the autocorrelation function of the binary image. By using larger and more complex shapes, a more generalized autocorrelation can be computed.

The grey tone co-occurrence approach characterizes texture by the spatial distribution of its grey tone. Coarse textures are those for which the distribution changes only slightly with distance and fine textures are those for which the distribution changes rapidly with distance.

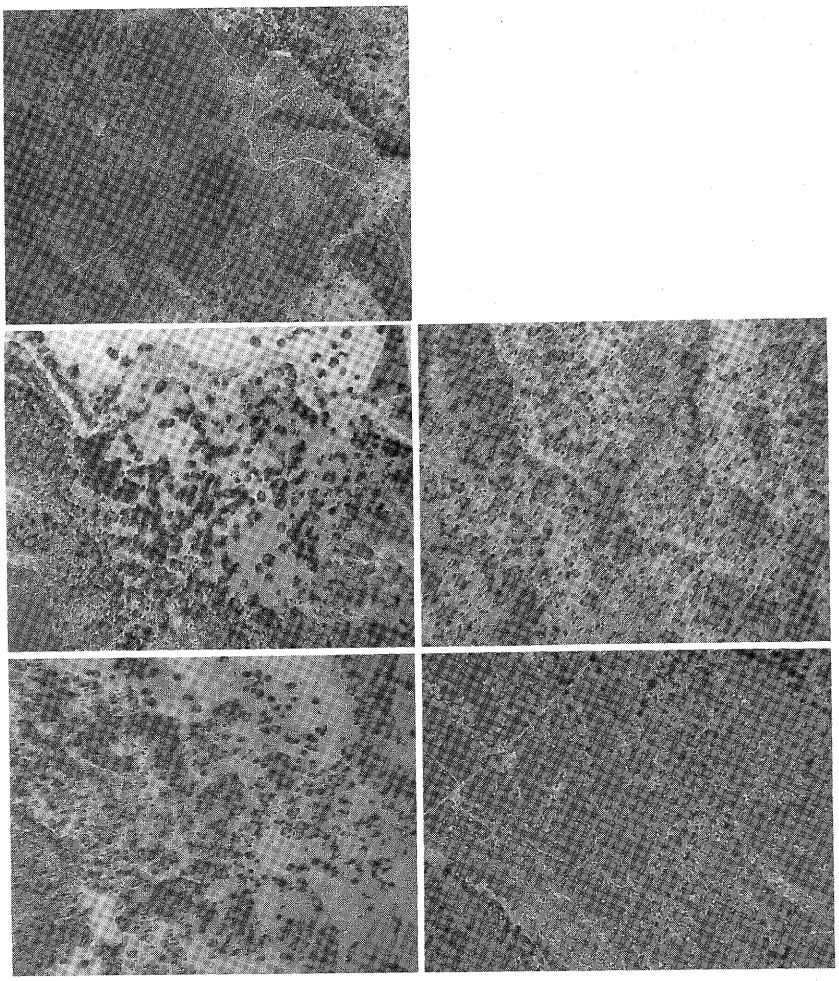

Fig. 2.8. Shows some examples where the same type of terrain generates a variety of textures within the same texture family

2.5.1 Optical Processing Methods and Texture

O'NEILL's [2.105] article on spatial filtering introduced the engineering community to the fact that optical systems can perform filtering of the kind used in communication systems. In the case of the optical systems, however, the filtering is two-dimensional. The basis for the filtering capability of optical systems lies in the fact that the light amplitude

distributions at the front and back focal planes of a lens are Fourier transforms of one another. The light distribution produced by the lens is more commonly known as the Fraunhofer diffraction pattern. Thus, optical methods facilitate two-dimensional frequency analysis of images.

The paper by CUTRONA et al. [2.106] provided a good review of optical processing methods for the interested reader. More recent books by GOODMAN [2.107], SHULMAN [2.108], PRESTON [2.109] and HUANG [2.110] comprehensively survey the area.

In this section, we describe the experiments done by LENDARIS and STANLEY [2.97], EGBERT et al. [2.111], and SWANLUND [2.112] using optical processing methods for aerial or satellite imagery. LENDARIS and STANLEY illuminated small circular sections of low altitude aerial photography and used the Fraunhofer diffraction pattern to derive features for identifying the sections. The circular sections represented circular areas on the ground 750 feet in diameter. The major category distinction they were interested in making was man-made versus non-man-made. They further subdivided the man-made category into roads, road intersections, buildings, and orchards.

The pattern vectors they used from the diffraction pattern consisted of 40 components. Twenty components were averages of the energy in rings on the diffraction pattern, and 20 were averages of the energy in 9° wedges on the diffraction pattern. They obtained over 90% identification accuracy.

EGBERT et al. [2.111] used an optical processing system to examine the texture on ERTS imagery of Kansas. They used circular areas corresponding to a ground diameter of about 35 km and looked at the diffraction patterns for the areas when they were snow covered and when they were not snow covered. They used a diffraction pattern sampling unit having 32 sector wedges and 32 annular rings to sample and measure the diffraction patterns. (See [2.113] for a description of the sampling unit and its use in coarse diffraction pattern analysis.) They were able to interpret the resulting angular orientation graphs in terms of dominant drainage patterns and roads, but were not able to interpret the spatial frequency graphs, which all seem to have had the same character: the higher the spatial frequency, the less the energy in that frequency band.

SWANLUND [2.112] has done work using optical processing of aerial images to identify species of trees. Using imagery obtained from Itasca State Park in northern Minnesota, photo-interpreters identified five (mixture) species of trees on the basis of texture: Upland hardwoods, Jack pine overstory/Aspen understory, Aspen overstory/Upland hardwoods understory, Red pine overstory/Aspen understory, and Aspen. They achieved classification accuracy of over 90%.

2.5.2 Texture and Edges

The autocorrelation function, the optical transforms, and the fast digital transforms (FFT and FHT) basically all relate texture to spatial frequency. ROSENFELD and THURSTON [2.88] conceived of texture not in terms of spatial frequency but in terms of edgeness per unit area. An edge passing through a resolution cell is detected by comparing the values for local properties obtained in pairs of nonoverlapping neighborhoods bordering the resolution cell. To detect microedges, small neighborhoods must be used. To detect macroedges, large neighborhoods must be used.

The local property which ROSENFELD and THURSTON suggested was the quick ROBERTS gradient (the sum of the absolute values of the differences between diagonally opposite neighboring pixels). Thus, a measure of texture for any subimage is obtained by computing the ROBERTS gradient image for the subimage and from it determining the average value of the gradient in the subimage. TRIENDL [2.114] used the Laplacian instead of the ROBERTS gradient.

SUTTON and HALL [2.115] extended the ROSENFELD and THURSTON idea by making the gradient a function of the distance between the pixels. Thus, for every distance d and subimage I defined over a neighborhood N of resolution cells, they compute

$$g(d) = \sum_{(i,j)\in N} \{|I(i,j) - I(i+d,j)| + |I(i,j) - I(i-d,j)| + |I(i,j) - I(i,j+d)| + |I(i,j) - I(i,j-d)|\} . \tag{2.54}$$

The graph of $g(d)$ is like the graph of minus the autocorrelation function translated vertically. SUTTON and HALL [2.115] applied this textural measure in a pulmonary disease identification experiment using radiographic imagery and obtained identification accuracy in the 80 percentile range for discriminating between normal and abnormal lungs when using a 128×128 subimage.

2.5.3 Digital Transform Methods and Texture

In the digital transform method of texture analysis, the digital image is typically divided into a set of non-overlapping small square subimages. Suppose the size of the subimage is $n \times n$ resolution cells; then the n^2 grey tones in the subimage can be thought of as the components of an n^2-dimensional vector. In the transform technique, each of these vectors is re-expressed in a new coordinate system. The Fourier transform uses the sine-cosine basis set, the Hadamard transform uses the Walsh function basis set, etc. The point of the transformation is that the basis

vectors of the new coordinate system have an interpretation that relates to spatial frequency (sequency), and since frequency (sequency) is a close relative of texture, we see that such transformations can be useful.

GRAMENOPOULOS [2.98] used a transform technique using the sine-cosine basis vectors (and implemented with the FFT algorithm) on ERTS imagery to investigate the power of texture and spatial pattern for terrain type recognition. He used subimages of 32 by 32 resolution cells and found that on Phoenix, Arizona, ERTS image 1049-17324-5, spatial frequencies larger than 3.5 cycles/km and smaller than 5.9 cycles/km contain most of the information needed to discriminate between terrain types. The terrain classes were: clouds, water, desert, farms, mountains, urban, riverbed, and cloud shadows. He achieved an overall identification accuracy of 87%.

HORNUNG and SMITH [2.99] have done work similar to that of GRAMENOPOULOS, but with aerial multispectral scanner imagery instead of ERTS imagery. MAURER [2.115a] used Fourier series analysis on color aerial film to obtain textural features to help determine crop types.

KIRVIDA and JOHNSON [2.100] compared the fast Fourier, Hadamard, and slant transforms for textural features on ERTS imagery over Minnesota. They used 8×8 subimages and five categories: Hardwoods, Conifers, Open, Water, City. Using only spectral information, they obtained 74% correct identification accuracy. When they added textural information obtained from the slant transform, they increased their identification accuracy on the training set to 99%. They found little difference between the different transform methods, although performance using 4×4 subimages was poorer than that using 8×8 subimages.

2.5.4 Spatial Grey Tone Dependence: Co-Occurrence

One aspect of texture is concerned with the spatial distribution and spatial dependence among the grey tones in a local area. BIXBY et al. [2.116] used restricted first and second order Markov meshes. DARLING and JOSEPH [2.117] used statistics obtained from the nearest neighbor grey tone transition matrix to measure this dependence for satellite images of clouds and were able to identify cloud types on the basis of their texture. READ and JAYARAMAMURTHY [2.118] divided an image into all possible (overlapping) subimages of reasonably small and fixed size and counted the frequency for all the distinct grey tone patterns. This is one step more general than DARLING, but requires too much memory if the grey tones can take on very many values. HARALICK and co-workers [2.104, 119, 120] suggested an approach which is a compromise between the two: to measure the spatial dependence of grey tones in a co-occur-

rence matrix for each fixed distance and/or angular spatial relationship, and use statistics of this matrix as measures of image texture. JULESZ [2.121] provided evidence that the human visual perception system discriminates texture on the basis of co-occurrence statistics.

The co-occurrence matrix $P=(p_{ij})$ has its (i, j)-th entry p_{ij} defined as the number of times grey tone i and grey tone j occur in resolution cells of a subimage having a specified spatial relation, such as being distance 1 neighbors. The textural features for the subimage are obtainable from the co-occurrence matrix by measures such as

$$\sum_i \sum_j p_{ij}^2$$

and

$$\sum_i \sum_j \frac{p_{ij}}{1+|i-j|}.$$

HARALICK et al. [2.103, 104] listed 14 different such measures.

Using statistics derived from the co-occurrence matrix, HARALICK performed a number of identification experiments. On a set of aerial imagery of eight terrain classes (old residential, new residential, lake, swamp, marsh, urban, railroad yard, scrub and wooded), he obtained 82% correct identification with 64×64 subimages. On an ERTS Monterey Bay, California, image, he obtained 84% correct identification using 64×64 subimages and both spectral and textural features on seven terrain classes: coastal forest, woodlands, annual grasslands, urban areas, large irrigated fields, small irrigated fields, and water. On a set of sandstone photomicrographs, he obtained 89% correct identification on five sandstone classes: Dexter-L, Dexter-H, St. Peter, Upper Muddy, Gaskel. The wide variety of images on which it has been found that grey tone co-occurrence carries much of the texture information is probably indicative of the power and generality of this approach.

2.5.5 A Textural Transform

Each of the approaches described above for the quantification of textural features has the common property that the textural features were computed for subimages of sizes such as 8×8, 16×16, 32×32, and 64×64 resolution cells. To determine textural features for one pixel we would center a subimage on the specified resolution cell and compute the textural features for the subimage. If we had to determine textural features for each pixel in an image, this would require a lot of computation work and would significantly increase the size of our data set. Thus the

usual approach has been to divide the image into mutually exclusive subimages and compute textural features on the selected subimages. Unfortunately, this procedure produces textural features at a coarser resolution than the original image.

In this section we use grey tone co-occurrence textural matrices to define a textural transform, and show how by only doubling or tripling the computation time required to determine the matrix it is possible to produce a resolution-preserving textural transform in which each pixel in the transformed image has textural information about its own neighborhood derived from local and global grey tone co-occurrences in the image. This kind of textural transform belongs to the class of image dependent non-linear spatial filters.

Let $Z_r \times Z_c$ be the set of resolution cells of an image I (in row-column coordinates). Let G be the set of grey tones that can appear on image I. Then $I: Z_r \times Z_c \rightarrow G$. Let R be a binary relation on $Z_r \times Z_c$ pairing together all those resolution cells that are in the desired spatial relationship. The co-occurrence matrix P, $P: G \times G \rightarrow [0, 1]$, for image I and binary relation R is defined by

$$P(i, j) = \frac{\#\{((a, b), (c, d)) \in R | I(a, b) = i \text{ and } I(c, d) = j\}}{\#R} \tag{2.55}$$

The textural transform J, $J: Z_r \times Z_c \rightarrow (-\infty, \infty)$, of image I relative to function f, is defined by

$$J(y, x) = \frac{1}{\#R(y, x)} \sum_{(a,b) \in R(y,x)} f[P(I(y, x), I(a, b))] . \tag{2.56}$$

Assuming f to be the identity function, the meaning of $J(y, x)$ is as follows. The set $R(y, x)$ is the set of all those resolution cells in $Z_r \times Z_c$, that are in the desired spatial relation to resolution cell (y, x). For any resolution cell $(a, b) \in R(y, x)$, $P(I(y, x), I(a, b))$ is the relative frequency by which the grey tone $I(y, x)$, appearing at resolution cell (y, x), and the grey tone $I(a, b)$, appearing at resolution cell (a, b), co-occur together in the desired spatial relation on the entire image. The sum

$$\sum_{(a,b) \in R(y,x)} P(I(y, x), I(a, b))$$

is just the sum of the relative frequencies of grey tone co-occurrence over all resolution cells that are in the specified relation to resolution cell (y, x). The factor $[\#R(y, x)]^{-1}$, the reciprocal of the number of resolution cells in the desired spatial relation to (y, x) is just a normalizing factor. Figure 2.9 illustrates the 0.82 to 0.88 micrometer band of some multi-

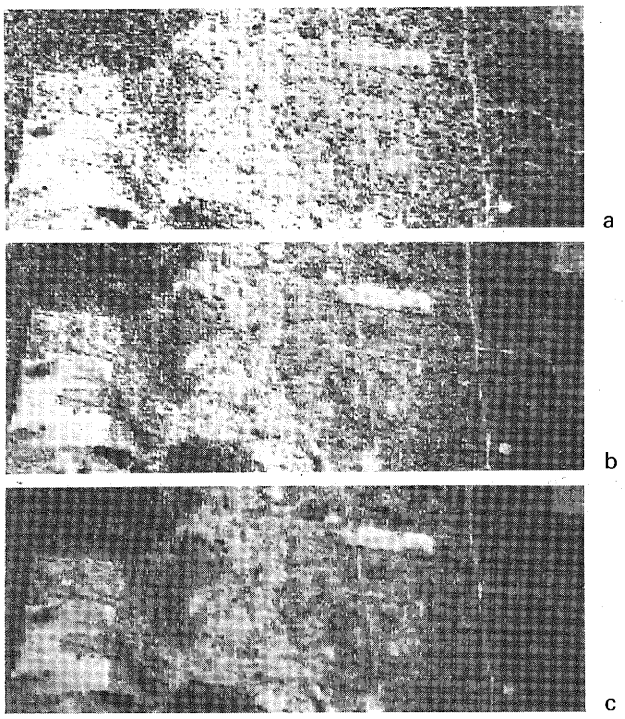

Fig. 2.9a–c. Illustrates the 0.82 to 0.88 μm band of some multispectral scanner imagery taken at 10000 feet over the Sam Houston National Forest, March 21, 1973. The image is shown in its original form, and after a 2×2 and 3×3 rectangular convolution. (a) 1×1 original of band 9 in edit No. 9, SAMH3 S19; (b) 2×2 convolution of band 9 in edit No. 9, SAMH3 S29; (c) 3×3 convolution of band 9 in edit No. 9, SAMH3 S39

spectral scanner imagery taken at 10000 feet over the Sam Houston National Forest, March 21, 1973. The image is shown in its original form, and after a 2×2 and 3×3 rectangular convolution. Figure 2.10 shows the respective textural transforms of the three images of Fig. 2.9 where the spatial relation R consists of all pairs of 8-neighboring resolution cells in $Z_r \times Z_c$. HARALICK [2.122] indicates that classification accuracy improves when both spectral and textural transform features are used on a Skylab S-192 image.

ZUCKER et al. [2.123] discussed a different kind of textural transform based on spot detectors. Applying a spot detector to an image having two different textures, the spot size matching the coarseness of one of the textures, and then averaging the resultant image, they show that it is easy to segment the image into its two textures.

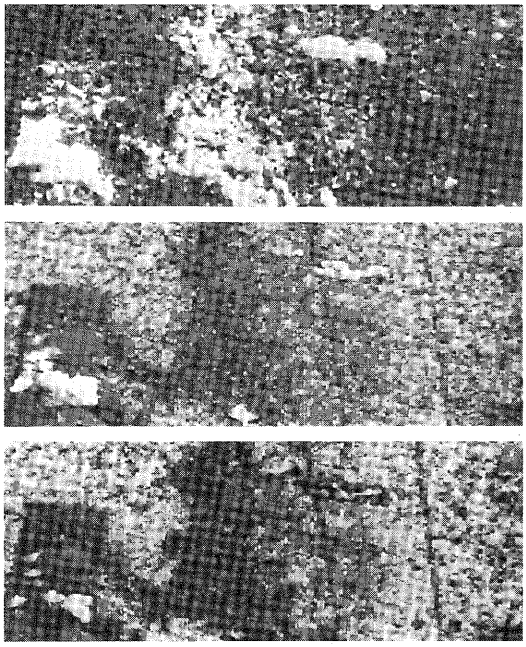

Fig. 2.10a–c. Shows the respective textural transforms of the three images of Fig. 2.9. (a) 1×1 before and 1×1 after $T \times T$, SAMH3 Q19; (b) 2×2 before and 1×1 after $T \times T$, SAMH3 Q29; (c) 3×3 before and 1×1 after $T \times T$, SAMH3 Q39

2.6 Near Real Time Hardware Processing

The ease with which remote sensors can inundate the investigator with data strongly indicates the virtue of a processing system which can process the data quickly. Hardware processing, although limited to performing rather simple basic functions, has the great advantage of near-real-time operation. In this section we survey the typical kinds of operations easily done in a near-real-time hardware system.

Basically, there are two types of image format inputs that a hardware system can have: film or print, and video tape. Systems which have black and white film or print inputs have the advantage of being usable with the most common form of image format data. However, they have the associated problem of registration. The analog tape input usually comes from a multispectral scanner sensor so that the individual images (channels) are already registered on the tape.

Hardware system outputs usually include forms of black and white and color displays. The black and white display can be used for examining

the images, one by one, or for showing simple two-category discriminations. The color display may be used to show a processed false color enhanced image or a color map indicating multi-category discriminations.

Those hardware systems having registration problems usually input the images using independently controllable flying spot scanners or vidicons. The operator manually registers the images by adjusting the rotation, translation, and scale controls of the input devices while quickly flickering between the displayed images. Images which are not in registration show on the flickered display a displeasing interference movement which indicates in what way the images have to be rotated, translated or scaled to be put in registration.

There are at least eight basic processing operations which a near-real time hardware device can perform either singly or in various sequences depending on the required processing function. They are:

1) level slicing or thresholding to produce binary images,
2) Boolean operations on binary images to produce new binary images,
3) differentiation for boundary detection,
4) linearly combining images for enhancement or calculation of discriminant functions,
5) quadratically combining images for enhancement or calculation of discriminant functions,
6) determining which of a set of signals is the maximum for category assignments,
7) analog to digital conversion or quantizing,
8) table look-ups for category assignments.

A level slicer or thresholder produces a binary {0, 1} output which is 1 whenever the video signal it is operating on is between two adjustable thresholds, and 0 at all other times. Two or more level slicers may operate simultaneously on different signals and their binary outputs may be ANDed together, thereby producing category decision boundaries which are rectangular parallelepipeds. However, should the images be linearly combined in two or more linear combiners and then level sliced and ANDed, the category decision boundaries would take on the more general form of parallelepipeds; should they be quadratically combined in two or more quadratic combiners, then level sliced and ANDed together, the category decision boundaries would be piecewise quadratic surfaces. In general, the combiner/level slicer combination implements a decision rule for a two-category problem.

If an image is differentiated and displayed, the boundaries can be enhanced. If it is first differentiated and then thresholded, its boundaries can be detected. If two or more images are differentiated, and then the thresholded differentiated outputs are ORed together, all the boundaries

on the set of images can be detected. In this case, the differentiator/thresholder combination implements a two category decision rule: boundary, no boundary.

The linear or quadratic combiners may be used by themselves to produce a false color image enhancement. For example, if the images are combined with one combiner driving each primary color in a color display, false color image enhancement results.

The linear or quadratic combiners may be used in conjunction with the maximum selector to implement a multi-category decision rule. In this rule, the combiners act as discriminant functions and the maximum selector determines which combiner has the largest signal. The decision rule then assigns the category associated with the combiner having the largest signal being processed.

The papers by MARSHALL and KRIEGLER [2.124] and ANDERSON et al. [2.54] describe the hardware processing systems at the Universities of Michigan and Kansas, respectively.

2.7 Software Processing

Although software processing of image data on a general purpose digital computer is more costly and definitely slower than near-real time, it has the advantage of great versatility. In fact, software can be written to implement any well defined processing task of finite length. LILLESTRAND and HOYT [2.125] reviewed the nature of the rapidly evolving digital systems for image processing. They characterize the system by the rate at which digitized imagery can be processed, by the complexity of the processing algorithms, and by the type of output produced. In this section we discuss the different formats in which multi-images are stored and the necessity of checking the A/D conversion process, and describe some of the basic library programs which any software processing system must have.

There are relatively few published descriptions of software systems. JOSEPH et al. [2.126] discussed a pattern recognition system for reconnaissance applications. In a continuing series of papers, JOSEPH and co-workers [2.127–129] described a system for the interactive preprocessing of image data. They discussed both the computer hardware configuration and the required image processing functions. FRIEDEN [2.129a] discussed the VICAR system developed at the Jet Propulsion Laboratory. BEBB et al. [2.130] presented an overview of image processing systems. HOFFER [2.131] gave a brief user's description of LARSYAA, a software system designed to analyze multispectral scanner data at Purdue University. JARVIS [2.132] discussed interactive image

processing and pattern recognition systems based on minicomputers. GAMBINO and CROMBIE [2.133] described the DIMES software system (Digital Image Manipulation and Enhancement System). TURINETTI and HOFFMANN [2.134] explained how OLPARS (On-Line Pattern Analysis and Recognition System) can be used to process remotely sensed image data. An Electromagnetic Sensing Laboratory report [2.135] describes the IDIM system (Interactive Digital Image Manipulation).

2.7.1 Multi-Image Formats

Since image processing systems must handle such large amounts of data, they are usually input-output bound. This makes the format in which the multi-image data are stored quite important. There are two basic kinds of formats for storing multi-image data. In what we shall call photo format, all the grey tones from the first image are stored as integers in a row by row logical record format on the first file. This file is followed by all the grey tone integers from the second image, also stored in a row by row logical record format, and so on. In what we shall call the corresponding point form, the grey tone integer from the first image's first resolution cell, first line, is followed by the grey tone integer from the second image's first resolution cell, first line, ..., followed by the grey tone integer from the last image's first resolution cell, first line, ..., followed by the grey tone integer from the first image's last resolution cell, first line, followed by the grey tone integer from the second image's last resolution cell, first line, and so on. Each such multi-image line of integers is stored as a logical record.

A format perhaps more versatile than the photo or corresponding point format is one which stores line 1 of image 1 followed by line 1 of image 2, ..., followed by line 1 of image N, and so on. Noticing that logical records are subimages, we see that the line format can be generalized to a subimage format where each logical record is a subimage of say K_1 rows by K_2 columns.

Editing, congruencing, registering, digital printer displaying and texture analysis are most quickly done with data in photo form, while most other operations such as feature extraction, clustering or pattern identification are done more conveniently in corresponding point format.

To conserve magnetic tape or disc storage space, speed up I/O time, and make more effective use of buffer space, the grey tone integers are often packed two, three or four integers per computer word. The packing is usually done by special machine language FORTRAN callable subroutines which convert a line or subimage of compacted integers at a time.

2.7.2 Checking A/D Conversion of Image Data

The first problem any image processing software system has to handle is the checking of the digital image conversion process. The digital tapes must be checked to verify that the A/D conversion was done successfully. Preliminary checking can be done by dumping the first few records on the tape; however, this is by no means a complete check. The image display program can make a complete check by outputting the tape in picture format on the digital printer, creating the grey tones by overprinting. If the number of resolution cells on the image is large enough to make the printing of the digital image awkward, or display of it impossible, a program may be utilized which reduces the image size by averaging blocks of $N \times N$ resolution cells or by selecting every N-th row and every N-th column.

Examination of the digital picture output should indicate what kind of editing needs to be done on the sides and top and bottom of the image, as well as indicating skewing and A/D conversion distortion. (Skewing can occur because it may be impossible to start digitizing each line of the image in exactly the same place. A/D conversion distortion can occur when jitter or noise external to the A/D conversion makes the conversion go awry.) If necessary, a deskewing program may have to be used to remove skew, and a special smoothing-replacement program may have to be used to operate on those resolution cells which were improperly converted. Histograms are helpful in identifying and isolating failure of individual A/D conversions.

2.7.3 User Oriented Commands and Library Programs

The heart of any image processing software system is its utility library programs, written in modular form, which perform each kind of special processing task. For ease of writing the library programs, transferring them to various computers, or understanding them, the programs are usually written in a universally available high level language such as FORTRAN, rather than in machine assembly language. Each library program or set of programs implements a specific user oriented command which can be input in a free format form. The library typically contains programs for image displaying, editing, scaling, registering, congruencing, mosaicking, selecting, quantizing, filtering, clustering, feature extraction, texture analysis, and pattern recognition.

The display commands can display an image either on a video CRT screen [2.136] or by using the digital printer. The printer can overprint characters to obtain shades of grey and use different color ribbons to obtain color map displays.

The editing commands can print out the grey tone values for any subimage, replace all occurrences of one grey tone value by another, change the grey tone in any one resolution cell from one value to another, and locate the next occurrence of any specified grey tone. The image scaling commands can either decrease the size of the image by sampling grey tones at every K_1-th pixel vertically and K_2-th pixel horizontally, or increase the size of the image by duplicating each pixel K_1 times vertically and K_2 times horizontally.

The registering command can translate any image vertically or horizontally so that it is aligned with another image; then it can combine the aligned images into a multi-image. The congruencing command can change the geometry of any image to the geometry of another image and then combine the images into a multi-image by the registering operation. The mosaicking command can paste together images in a side by side manner. The selecting or subimage command can construct a new image consisting of any specified size subimage from any set of bands of a multi-image.

The quantizing command can quantize or contrast stretch an image by an equal interval or equal probability quantizing procedure. The filtering command performs a spatial filtering of the image by a specified point spread function.

The clustering, feature extracting, texture, and pattern recognition operations are much more complex than the utility operations just described. However, one processing operation universally included in this set has the capability to compute histograms and scattergrams of any portion of any image or for any ground truth category of an image. Histograms or scattergrams are usually displayed on the line printer, or more conveniently on a CRT with alphanumeric as well as line generating abilities.

Ideally, all user oriented programs should accept image data in a standard format and accept data parameters and control information in a free-format manner. System routines should provide for error processing and dynamic storage allocation.

2.8 Conclusion

We have briefly discussed aspects of the automatic processing of remotely sensed imagery. We have indicated the importance of various normalizing, feature extraction, decision rule determination and clustering algorithms. We have discussed processing from a hardware and software perspective. Hopefully, we have provided enough insight about the vocabulary and

concepts of the automatic data processing world to enable a scientist to read and understand the remote sensing papers in his field which deal with automatic data processing.

References

2.1 R. N. COLWELL: American Scientist **49**, 9–36 (1961)
2.2 R. N. COLWELL: "Uses and Limitations of Multispectral Remote Sensing", *Proc. 4th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, (1966)
2.3 W. A. MALILA: Photogrammetric Engineering **34**, 566–575 (1968)
2.4 *NASA MSC 3rd Annual Earth Resources Program Review*, Vols. I–III, NASA MSC-03742, Manned Spacecraft Center, Tex. (1970)
2.5 *NASA MSC 4th Annual Earth Resources Program Review*, Vols. I–V, NASA MSC-05937, Manned Spacecraft Center, Tex. (1972)
2.6 J. D. LENT, G. A. THORLEY: Remote Sensing of Environment **1**, 31–45 (1969)
2.7 C. M. HAY: Photogrammetric Engineering **40**, 1283–1293 (1974)
2.8 N. R. NUNNALLY: Remote Sensing of Environment **1**, 1–7 (1969)
2.8a J. D. BARR, R. D. MILES: Photogrammetric Engineering **36**, 1155–1165 (1970)
2.9 S. A. MORAIN, D. S. SIMONETT: Photogrammetric Engineering **33**, 730–740 (1967)
2.10 H. C. MACDONALD, P. A. BRENNAN, L. F. DELLWIG: IEEE Trans. GE-**5**, 72–78 (1967)
2.11 F. T. ULABY: IEEE Trans. AP-**22**, 257–265 (1974)
2.12 P. M. MERIFIELD, J. CRONIN, L. L. FOSHEE, S. J. GAWARECKI, T. J. NEAL, R. E. STEVENSON, R. O. STONE, R. S. WILLIAMS, JR.: Photogrammetric Engineering **35**, 654–668 (1969)
2.13 T. PHILLIPS: "Corn Blight Data Processing Analysis and Interpretation", *4th Annual Earth Resources Program Review*, NASA/MSC (1972)
2.14 K. S. FU, D. A. LANDGREBE, T. L. PHILLIPS: Proc. IEEE **57**, 639–653 (1969)
2.15 R. M. HARALICK, F. CASPALL, D. S. SIMONETT: Remote Sensing of Environment **1**, 131–142 (1969)
2.16 R. M. CENTNER, E. D. HIETANEN: Photogrammetric Engineering **37**, 177–186 (1971)
2.17 P. E. ANUTA, R. B. MACDONALD: Remote Sensing of Environment **2**, 53–67 (1971)
2.18 R. M. HOFFER, P. E. ANUTA, T. L. PHILLIPS: Photogrammetric Engineering **37**, 989–1001 (1972)
2.19 J. D. TURINETTI, O. W. MINTZER: Photogrammetric Engineering **39**, 501–505 (1973)
2.20 P. H. SWAIN: "Pattern Recognition: A Basis of Remote Sensing Data Analysis", LARS Information Note 111572. The Laboratory for Applications of Remote Sensing, Purdue Univ., Lafayette, Ind. (1973)
2.21 D. A. LANDGREBE: "Machine Processing for Remotely Acquired Data", LARS Information Note 031573. The Laboratory for Applications of Remote Sensing, Purdue Univ., Lafayette, Ind. (1973)
2.22 G. NAGY: Proc. IEEE **60**, 1177–1199 (1972)
2.23 D. EGBERT, F. T. ULABY: 'Effect of Angular Variation on Terrain and Spectral Reflectivity", *Proc. of 17th Symp. of the AGARD Electromagnetic Wave Propagation Panel on Propagation Limitations in Remote Sensing* (1971)
2.24 D. STEINER, H. HAEFNER: Photogrammetric Engineering **31**, 269–280 (1965)
2.25 R. M. HOFFER, R. A. HOLMES, J. R. SHAY: "Vegetative Soil and Photographic Factors Affecting Tone in Agricultural Remote Multispectral Sensing", *Proc. 4th Intern. Symp. on Remote Sensing of the Environment*. Univ. of Mich., Ann Arbor, Mich. (1966)
2.26 R. M. HARALICK: Pattern Recognition **5**, 391–403 (1973)

2.27 L. R. PETTINGER: "Analysis of Earth Resources on Sequential High Altitude Multiband Photography", Special Report, Forestry and Conservation, Univ. of Calif. (1969)
2.28 F. J. KRIEGLER, W. A. MALILA, R. F. NALEPKA, J. RICHARDSON: "Preprocessing Transformations and Their Effects on Multispectral Recognition", *Proc. 6th Intern. Symp. on Remote Sensing of the Environment*, Univ. of Mich., Ann Arbor, Mich. (1969) pp. 97–131
2.29 R. B. CRANE: "Preprocessing Techniques to Reduce Atmospheric and Sensor Variability in Multispectral Scanner Data", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1971) pp. 1345–1350
2.30 H. W. SMEDES, M. M. SPENCER, F. J. THOMSON: "Preprocessing of Multispectral Data and Simulation of Earth Resources Technology Satellite Data Channels to Make Computer Terrain Maps of a Yellowstone National Park Test Site", *3rd Ann. Earth Resources Program Review*, NASA MSC-03742 (1970) pp. 10–1 to 10–25
2.31 R. F. NALEPKA, J. P. MORGENSTERN: "Signature Extension Techniques Applied to Multispectral Scanner Data", *Proc. 8th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan (1972) pp. 881–893
2.32 J. S. CRABTREE, J. O. MCLAURIN: Photogrammetric Engineering **36**, 70–76 (1970)
2.32a R. B. MCEWEN: "An Evaluation of Analog Techniques for Image Registration", U.S. Geological Survey Professional Paper 700-D (1970) pp. D305–D311
2.33 F. G. PEET: "Affine Transformations from Aerial Photos to Computer Compatible Tapes", *3rd Earth Resources Technology Satellite Symp.*, NASA SP-351, Goddard Space Flight Center (1973) pp. 1719–1724
2.34 R. A. EMMERT, C. D. MCGILLEM: "Multitemporal Geometric Distortion Correction Utilizing the Affine Transformation", *Machine Processing of Remotely Sensed Data*, Purdue Univ. (1973)
2.35 M. SZETO: "An Approximation for Projective Transformation for Satellite Imagery", Report RC-3320, IBM T.J. Watson Research Center, Yorktown Heights, New York (1971)
2.36 R. O. DUDA, P. E. HART: *Pattern Classification and Scene Analysis* (John Wiley & Sons, New York 1973) pp. 405–424
2.37 H. MARKARIAN, R. BERNSTEIN, D. G. FERNEYHOUGH, L. E. GREGG, F. S. SHARP: "Implementation of Digital Techniques for Correcting High Resolution Images", Paper No. 71-326, AIAA (1971) pp. 285–304
2.38 P. F. ANUTA: IEEE Trans. GE-**8**, 353–368 (1970)
2.39 D. I. BARNEA, H. F. SILVERMAN: IEEE Trans. C-**21**, 179–186 (1972)
2.40 D. F. WEBBER: "Techniques for Image Registration", *Machine Processing of Remotely Sensed Data*, Purdue Univ. (1973)
2.41 S. S. YAO: "A Method for Digital Image Registration Using a Mathematical Programming Technique", *Machine Processing of Remotely Sensed Data*, Purdue Univ. (1973)
2.42 S. S. RIFFMAN: "Digital Rectification of ERTS Multispectral Imagery", *Symp. on Significant Results Obtained from the Earth Resources Technology Satellite*, NASA SP-327, Goddard Space Flight Center (1973) pp. 1131–1142
S. S. RIFFMAN: "Evaluation of Digitally Corrected ERTS Imagery", *Amer. Society of Programmetry Symp. Proc. on the Management and Utilization of Remote Sensing Data*, Sioux Falls, SD (1973) pp. 206–221
2.43 E. DIDAY, J.-C. SIMON: In *Pattern and Speech Recognition*, ed. by K. S. FU (Springer, Berlin, Heidelberg, New York 1975) Chapter 3
2.44 B. J. DAVIS, P. H. SWAIN: "An Automated and Repeatable Data Analysis Procedure for Remote Sensing Application", *Proc. 9th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1974) pp. 771–774

2.45 A. G. WACKER: "The Effect of Subclass Numbers on Maximum Likelihood Gaussian Classification", *Proc. 8th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1972) pp. 851–859
2.46 D. FOLEY: IEEE Trans. IT-**18**, 618–626 (1972)
2.47 L. KANAL, B. CHANDRASEKARAN: Pattern Recognition **3**, 225–234 (1971)
2.47a W. G. BROONER, R. M. HARALICK, I. DINSTEIN: "Spectral Parameters Affecting Automated Image Interpretation Using Bayesian Probability Techniques", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1971) pp. 1929–1949
2.48 W. G. EPPLER, C. A. HEMKE, R. H. EVANS: "Table Look-up Approach to Pattern Recognition", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1971) pp. 1415–1425
2.49 W. G. EPPLER: "An Improved Version of the Table Look-up Algorithm for Pattern Recognition", *Proc. 9th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1974) pp. 793–812
2.50 R. W. ASHBY: "Constraint Analysis of Many-Dimensional Relations", *Yearbook of Soc. General Systems Research*, Vol. 9, 99–105 (1964)
2.51 P. W. COOPER: Cybernetica **5**, 215–238 (1962)
2.52 G. NAGY: Proc. IEEE **56**, 836–862 (1968)
2.53 G. S. SEBESTYEN: *Decision-Making Processes in Pattern Recognition* (The MacMillan Co., New York 1962)
2.54 P. N. ANDERSON, G. W. DALKE, R. M. HARALICK, G. L. KELLY, R. K. MOORE: "Electronic Multi-Image Analog-Digital Processor and Color Display", *1972 IEEE Convention Digest* (1972)
2.55 D. G. LAINIOTIS: IEEE Trans. IT-**15**, 730–731 (1969)
2.56 W. A. WHITNEY: IEEE Trans. C-**20**, 1100–1103 (1971)
2.57 S. KULLBACK: *Information Theory and Statistics* (John Wiley & Sons, New York 1959)
2.58 T. MARILL, D. M. GREEN: IEEE Trans. IT-**19**, 11–17 (1963)
2.59 T. KAILATH: IEEE Trans. CT-**15**, 52–60 (1967)
2.60 N. J. NILSSON: *Learning Machines* (McGraw-Hill, New York 1965)
2.61 T. L. GRETTENBERG: IEEE Trans. IT-**19**, 11–17 (1963)
2.62 K. S. FU, C. H. CHEN: "Sequential Decision, Pattern Recognition and Machine Learning", Techn. Rept. TR-EE 65-6, School of Electrical Engineering, Purdue Univ., Lafayette, Ind. (1965)
2.63 P. H. SWAIN, T. V. ROBERTSON, A. G. WACKER: "Comparison of the Divergence and B-distance in Feature Selection". LARS Information Note 020871, The Laboratory for Applications of Remote Sensing, Purdue Univ., Lafayette, Ind. (1971)
2.63a P. H. SWAIN, R. C. KING: "Two Effective Feature Selection Criteria for Multispectral Remote Sensing", *1st Intern. Joint Conf. on Pattern Recognition*, Washington, D.C. (1973)
2.64 H. JEFFREYS: Proc. of the Roy. Soc. A **186**, 454–461 (1946)
2.65 K. MATUSITA: Ann. Inst. Stat. Math. (Tokyo) **3**, 17–35 (1951)
2.66 S. KULLBACK, R. A. LIEBLER: Ann. Math. Stat. **22**, 79–86 (1951)
2.67 P. C. MAHALANOBIS: Proc. Nat. Inst. of Science (India) **27**, 49–55 (1936)
2.68 C. C. BABU: IEEE Trans. IT-**18**, 428 (1972)
2.69 H. HOTELLING: J. Educ. Psych. **24**, 417–441 (1933)
2.70 S. S. VIGLIONE: "Application of Pattern Recognition Technology", *in Adaptive Learning Systems*, ed. by J. M. MENDEL and K. S. FU (Academic Press, New York 1970)
2.71 P. J. READY, P. A. WINTZ, S. J. WHITSITT, D. A. LANDGREBE: "Effects of Data Compression and Random Noise on Multispectral Data", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1971) pp. 1321–1342

2.72 R. M. HARALICK, I. DINSTEIN: IEEE Trans. CAS-**22**, 440–450 (1975)
2.73 P. J. READY, P. A. WINTZ: IEEE Trans. COM-**21**, 1123–1131 (1973)
2.74 M. M. TAYLOR: "Principal Components Color Display of ERTS Imagery", *3rd Earth Resources Technology Satellite-1 Symp.*, NASA SP-351, Goddard Space Flight Center (1973) pp. 1877–1887
2.75 E. DIDAY, J. C. SIMON: Clustering Analysis, in: *Communication and Cybernetics*, Vol. 10, ed. by K. S. FU (Springer, Berlin, Heidelberg, New York 1975)
2.76 R. M. HARALICK: "Adaptive Pattern Recognition of Agriculture in Western Kansas by Using a Predictive Model in Construction of Similarity Sets", *Proc. 5th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1967)
2.77 E. M. DARLING, J. G. RAUDSEPS: Pattern Recognition **2**, 313–335 (1970)
2.78 R. M. HARALICK, I. DINSTEIN: IEEE Trans. SMC-**1**, 275–289 (1971)
2.79 L. BORRIELLO, F. CAPOZZA: "A Clustering Algorithm for Unsupervised Crop Classification", *Proc. 9th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1974) pp. 181–188
2.80 G. H. BALL, D. J. HALL: "ISODATA, A Novel Method of Data Analysis and Pattern Classification", Techn. Rept., Stanford Research Institute, Menlo Park, Calif. (1965)
2.81 A. G. WACKER, A. LANDGREBE: "Boundaries in Multi-Spectral Imagery by Clustering", *Proc. 1970 IEEE Symp. on Adaptive Processes (9th) Decision and Control*, Univ. of Texas, Austin, Tex. (1970)
2.82 H. W. SMEDES, H. J. LINNERUD, L. B. WOOLOVER, M. Y. SU, R. R. JAYROE: "Mapping of Terrain by Computer Clustering Techniques Using Multi-spectral Scanner Data and Using Color Aerial Film", *NASA 4th Ann. Earth Resources Progress Review*, MSC-05937 **3**, Houston, Tex. (1972) pp. 61–1 to 61–30
2.83 M. Y. SU, R. E. CUMMINGS: "An Unsupervised Classification Technique for Multispectral Remote Sensing Data", *Proc. 8th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1972) pp. 861–879
2.84 C. R. BRICE, C. L. FENNEMA: Artificial Intelligence **1**, 205–226 (1970)
2.85 L. G. ROBERTS: "Machine Perception of Three-Dimensional Solids", in *Optical and Electro-Optical Processing of Information* (MIT Press, Cambridge, Mass. 1965) pp. 159–197
2.86 J. PREWITT: "Object Enhancement and Extraction", in *Picture Processing and Psychopictorics*, ed. by B. S. LIPKIN and A. ROSENFELD (Academic Press, New York 1970) pp. 75–149
2.87 M. H. HUECKEL: J. Asso. Computing Machinery **18**, 113–125 (1971)
2.88 A. ROSENFELD, M. THURSTON: IEEE Trans. C-**20**, 562–569 (1971)
2.89 R. M. HARALICK, G. L. KELLY: Proc. IEEE **57**, 654–665 (1969)
2.90 G. NAGY, G. SHELTON, J. TOLABA: "Procedural Questions in Signature Analysis", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Mich., Ann Arbor, Mich. (1971) pp. 1387–1401
2.91 R. R. JAYROE: "Unsupervised Spatial Clustering with Spectral Discrimination", NASA Techn. Note TN D-7312, George C. Marshall Space Flight-Center, Ala. (1973)
2.92 T. V. ROBERTSON, K. S. FU, P. H. SWAIN: Multispectral Image Partitioning", LARS Information Note 171373, Purdue Univ., Lafayette, Ind. (1973)
2.93 A. J. LEWIS: "Geomorphic Evaluation of Radar Imagery of Southeastern Panama and Northwestern Columbia", CRES Techn. Report 133–18, Univ. of Kansas Center for Research, Inc., Lawrence, Kan. (1970)
2.94 H. C. MACDONALD: "Geologic Evaluation of Radar Imagery from Darien Province, Panama", CRES Techn. Report 133–6, Univ. of Kansas Center for Research, Inc., Lawrence, Kan. (1970)

2.95 R.M.HARALICK, D.E.ANDERSON: "Texture-Tone Study with Application to Digitized Imagery", CRES Techn. Report 182-2, Univ. of Kansas Center for Research, Inc., Lawrence, Kan. (1971)

2.96 H.KAIZER: "A Quantification of Textures on Aerial Photographs", Techn. Note 121, Boston Univ. Research Laboratories (1955) AD 69484

2.97 G.G.LENDARIS, G.L.STANLEY: Proc. IEEE **58**, 198–216 (1970)

2.98 N.GRAMENOPOULOS: "Terrain Type Recognition Using ERTS-1 MSS Images", *Symp. on Significant Results Obtained from the Earth Resources Technology Satellite*, NASA SP-327 (1973) pp. 1229–1241

2.99 R.J.HORNUNG, J.A.SMITH: "Application of Fourier Analysis to Multispectral/Spatial Recognition", *Management and Utilization of Remote Sensing Data ASP Symp.*, Sioux Falls, South Dakota (1973)

2.100 L.KIRVIDA, G.JOHNSON: "Automatic Interpretation of Earth Resources Technology Satellite Data for Forest Management", *Symp. on Signification Results Obtained from the Earth Resources Technology Satellite*, NASA SP-327 (1973) pp. 1076–1082

2.101 G.MATHERON: *Elements Pour Une Theorie des Milieux Poreux* (Masson, Paris 1967)

2.102 J.SERRA, G.VERCHERY: Film Science and Technology **6**, 141–158 (1973)

2.103 R.M.HARALICK, K.SHANUMGAM: *2nd Symp. on Significant Results Obtained from Earth Resources Technology Satellite – 1*, Goddard Space Flight Center (1973) pp. 1219–1228

2.104 R.M.HARALICK, K.SHANUMGAM, I.DINSTEIN: IEEE Trans. SMC-**3**, 610–621 (1973)

2.105 E.O'NEILL: IRE Trans. IT-**5**, 56–65 (1956)

2.106 L.J.CUTRONA, E.N.LEITH, C.J.PALERMO, L.J.PORCELLO: IRE Trans. IT-**6**, 386–400 (1960)

2.107 J.W.GOODMAN: *Introduction to Fourier Optics* (McGraw-Hill, New York 1968)

2.108 A.R.SHULMAN: *Optical Data Processing* (John Wiley & Sons, Inc., New York 1970)

2.109 K.PRESTON: *Coherent Optical Computers* (McGraw-Hill, New York 1972)

2.110 T.S.HUANG (Ed.): *Topics in Applied Physics*, Vol. 6: Picture Processing and Digital Filtering (Springer, Berlin, Heidelberg, New York 1975)

2.111 D.EGBERT, J.MCCAULEY, J.MCNAUGHTON: "Ground Pattern Analysis in the Great Plains", *Semi-annual Earth Resources Technology Satellite A Investigation Report*, Remote Sensing Laboratory, Univ. of Kansas, Lawrence, Kan. (1973)

2.112 G.SWANLUND: "Honeywell's Automatic Tree Species Classifier", Report 9D-G-24, Honeywell Systems and Research Division (1969)

2.113 N.JENSEN: Photogrammetric Engineering **39**, 1321–1328 (1973)

2.114 E.E.TRIENDL: "Automatic Terrain Mapping by Texture Recognition", *Proc. 8th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1972) pp. 771–776

2.115 R.SUTTON, E.HALL: IEEE Trans. C-**21**, 667–676 (1972)

2.115a H.MAURER: "Texture Analysis with Fourier Series", *Proc. 9th Intern. Symp. on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Mich. (1974) pp. 1411–1420

2.116 R.BIXBY, G.ELERDING, V.FISH, J.HAWKINS, R.LOEWE: "Natural Image Computer Final Report, Vol. 1, C-4035, Philco-Ford Corp. Aeronutronic Division, Newport Beach, Calif. (1967)

2.117 E.M.DARLING, R.D.JOSEPH: IEEE Trans. SSC-**4**, 38–47 (1968)

2.118 J.S.READ, S.N.JAYARAMAMURTHY: IEEE Trans. C-**21**, 803–812 (1972)

2.119 R.M.HARALICK: "A Texture-Context Feature Extraction Algorithm for Remotely Sensed Imagery", *Proc 1971 IEEE Decision and Control Conference*, Gainesville, Fla. (1971) pp. 650–657

2.120 R. M. HARALICK, K. SHANUMGAM, I. DINSTEIN: "On Some Quickly Computable Features for Texture", *Proc 1972 Symp. on Computer Image Processing and Recognition*, Vol. 2, Univ. of Missouri (1972) pp. 12-2-1 to 12-2-10
2.121 B. JULESZ: Scientific American **232,** 34–43 (1975)
2.122 R. M. HARALICK: "A Resolution Preserving Textural Transform for Images", *Computer Graphics Pattern Recognition and Data Structure Conference*, IEEE Computer Society and ACM Group on Computer Graphics, Beverly Hills, Calif. (1975)
2.123 S. W. ZUCKER, A. ROSENFELD, L. S. DAVIS: "Picture Segmentation by Texture Discrimination", Computer Science Techn. Rept. 356, Univ. of Maryland, College Park (1975)
2.124 R. E. MARSHALL, F. J. KRIEGLER: "An Operational Multispectral Survey System", *Proc. 7th Intern. Symp. on Remote Sensing of Environment*, Univ. of Michigan, Ann Arbor, Mich. (1971) pp. 2169–2192
2.125 R. L. LILLESTRAND, R. R. HOYT: Photogrammetric Engineering **40**, 1201–1218 (1974)
2.126 R. D. JOSEPH, R. G. RUNGE, S. S. VIGLIONE: "A Compact Programmable Pattern Recognition System", MDAC Paper WD 1409, McDonnell Douglas Astronautics Co., Newport Beach, Calif. (1970) [presented at *The Symp. on Applications of Reconnaisance Technology to Monitoring and Planning Environmental Change*, Rome Air Development Center, Rome, New York (1970)]
2.127 R. D. JOSEPH, S. S. VIGLIONE: "Preprocessing for Interactive Imagery Analysis", *Two-Dimensional Digital Signal Processing Conf.*, Univ. of Missouri, Columbia, Missouri (1971)
2.128 R. D. JOSEPH, E. E. NELSON: "Orbital Information Systems—Image Processing and Analysis", Report MDC G2593, McDonnell Douglas Co., Huntington Beach, Calif. (1972)
2.129 S. S. VIGLIONE: "Digital Imagery Processing and Analysis", *Symp. on Operational Remote Sensing*, American Society of Photogrammetry, Houston, Tex. (1972) (McDonnell Douglas Report WD 1845, Huntington, Calif.)
2.129a H. FRIEDEN: "Image Processing System VICAR Guide to System Use", Report 71–135, Jet Propulsion Laboratory, Pasadena, Calif. (1971)
2.130 J. BEBB, W. D. STROMBERG, R. E. NIMENSKY: "An Overview of Image Processing", TM-5068/001/00, System Development Corp., Santa Monica, Calif. (1973) AD 757 119
2.131 R. M. HOFFER: "ADP of Multispectral Scanner Data for Land Use Mapping", LARS Information Note 080372, The Laboratory for Applications of Remote Sensing, Purdue Univ., Lafayette, Ind. (1973)
2.132 R. A. JARVIS: Computer **7** (10), 49–59 (1974)
2.133 L. A. GAMBINO, M. A. CROMBIE: Photogrammetric Engineering **40**, 1295–1302 (1974)
2.134 J. D. TURINETTI, R. J. HOFFMAN: Photogrammetric Engineering **40**, 1323–1330 (1974)
2.135 ESL: "PECOS II Users Manual", ESL-TM-503, Electromagnetic Systems Laboratory, Inc., Sunnyvale, Calif. (1974)
2.136 N. H. KREITZER, W. J. FITZGERALD: IEEE Trans. C-**22**, 128–134 (1973)