SECTION 38

DATA PROCESSING AT THE UNIVERSITY OF KANSAS

by

Robert M. Haralick
Center for Research, Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas

The University of Kansas has taken a two-fold approach to the data processing of remotely sensed imagery. Our approach has been based upon the need to have a special purpose hardware facility for the near-real time processing of multi-image data and the need to have a general purpose digital computer facility for the more sophisticated non-real time processing. Our near-real time facility is called IDECS (Image Discrimination Enhancement Combination System) and our non-real time facility is called KANDIDATS (Kansas Digital Image Data System). These facilities have been funded from both NASA and DOD sources.

THE NEED FOR A DUAL APPROACH

During the next decade there is a large amount of research yet to be done on data processing methods in order to bring the maturity of data processing up to the maturity of sensor technology. Yet, while this research is being done, many remotely sensed data sets from both aircraft and satellite platforms will have to be processed. Important constraints are, therefore, imposed on the remote sensing data processing center. It must have a flexible enough computational facility to implement and evaluate new ideas so that basic long range research on effective processing algorithms can be done; and for those data sets which now have to be processed quickly, it must have near real-time equipment to process and display those image sets economically.

For the near real-time equipment, it is extremely important that the man-machine interface be as convenient as possible for the interpreter since all image data is ultimately utilized by a human interpreter in some form. A color display is very useful for presenting image data to an interpreter since the human eye can distinguish differences in colors more readily than differences in grey levels. IDECS is one of the first systems to utilize a color display for presenting remote sensing data and it is described in the next section.

A HARDWARE SYSTEM FOR MULTI-IMAGE PROCESSING: IDECS

The IDECS (Image Discrimination, Enhancement, and Combination System) is an analog-digital near real-time image processing system and has been in continual development at the University of Kansas Center for Research, Inc., since 1964. The IDECS is a unique facility for performing a wide variety of enhancements, measurements, and category discriminations on single and multiple images. Currently, the input images must be in photographic form, but their source may be aerial and

space photography, airborne radar, infrared, multi-spectral scanners, medical and industrial X-rays, or maps. The primary IDECS output is on a color display unit; however, other outputs include a black-and-white monitor, area measurements on a counter, and a pseudo three-dimensional display.

A photograph of the IDECS is shown in Figure 1, and a block diagram of the total system is shown in Figure 2. The input to the IDECS consists of three flying-spot scanners suitable for inputting image transparencies from 3 x 4 inches to 35 mm format, a vidicon camera utilized for map or photographic inputs, and a congruencing unit which can rotate, translate, and scale images. The image scanners have the following three modes of operation:

(1) a continuous scan where the horizontal and vertical deflections for the CRT are driven by ramps which are synchronized with the display units,

(2) a staircase or dot scan where deflections are determined by the output of two digital-to-analog converters driven by two binary counters which are synchronized with the display unit, or

(3) a PDP-15/20 computer controlled slow-scan, where the scanning dot is moved horizontally and vertically to specific location.

Modes (1) and (2) are used when real time processing is desired, and the program controlled scan is used to gather information from specific areas of the film for training purposes and also to obtain a more accurate analog-to-digital conversion when necessary. Images to be scanned are positioned above the raster with an enlarging lens placed between. A condenser lens focuses the light transmitted through the image onto the cathode of a video photomultiplier tube. A reference photomultiplier tube and an automatic gain control (AGC) loop to modulate the cathode of the CRT are used to assure uniformity of light level over the entire area of the image being scanned. The reference photomultiplier tube is placed beside the enlarging lens and senses the light output from the raster at full value. The desired signal is used as an input into the AGC amplifier which is an error amplifier with a voltage reference. The resultant error signal is level shifted and is used as the signal to modulate the cathode of the CRT.

The operator can congruence multiple images by adjusting the position and size for horizontal and vertical position and rotating one image with respect to another image as the two images are electronically flickered. When the images are aligned, the displeasing interference pattern which occurs for misaligned images disappears.

Once the images are congruenced, they may be processed and enhanced by IDECS. There are a variety of processing and enhancing functions available. One function is a linear combiner unit which performs a linear transformation on a multi-image set in 1/30 of a second. This unit may be used for coordinate rotation if the coefficients of the linear transformation are selected appropriately. Another function is level selection. A level selector produces a binary output for an image

if and only if the input video signal is between two adjustable thresholds. The level selectors can be operator controlled or computer controlled. In addition, two more level selectors can be logically combined producing an output if and only if each one has an output, thus implementing a MIN-MAX decision rule.

The automatic classifier is a unit (not computer controlled) which is used to select and display all points on a video image whose levels fall within the same range of level as those detected in a small rectangular training area of the image. The position and size of this training area is selected using a small joy stick to determine position, and control pots are used to adjust vertical and horizontal size. To accomplish this operation, a rectangular training area is first defined on the image. Then peak detectors sample and hold the positive and negative signal peaks within the range of the training area. The remaining portion of the video signal is then compared with the peak levels of the training area; whenever the video signal falls within the training voltage range, a digital output is produced for processing or display.

Other functions include a unit to measure the area of any displayed category or grey level, a variable time constant differentiation unit to enhance edges and a pseudo three-dimensional display unit which permits one to view the three-dimensional surface generated by the grey tone density of an image. Soon to be implemented is a near real-time (1/30 of a second) table lookup pattern classifier which assigns categories on the basis of the digitized levels of two video signals and stored parameters for a Bayes decision rule.

A recently acquired PDP-15/20 computer is being interfaced to the system so that the IDECS can be program controlled and have a wider capability in performing image enhancements and category identifications. In effect, the PDP-15/20 will perform the task of generating a decision rule from data gathered by the IDECS, and the IDECS implements the resulting rule (in near real-time) on data derived by scanning the images. In general, five steps will be required in performing category identifications for images:

(1) the images must be congruenced,

(2) training data must be obtained by the computer from the images by directing the IDECS to scan appropriate areas,

(3) from the training data, the PDP-15/20 is programmed to determine the parameters for the chosen decision rule,

(4) the calculated parameters are used to set control voltages in the analog processing subsystems in the IDECS, and

(5) the specified category identifications are made and displayed by IDECS.

The PDP-15/20 will control the image processing steps of the IDECS by issuing commands to the IDECS central processing unit (CPU). The CPU in turn directs data flow and processing in the IDECS. The system configuration can be digitally selected utilizing a twenty by twenty video configuration matrix.

A digital disc memory having twenty-four channels and containing 24,000,000 bits of storage is also utilized in the system. There is an interface between the disc and computer capable of transferring at the rate of 18,000,000 bits per second. Also, there is an interface between the disc and the color display unit of the IDECS with the capability of up to three six-bit digital-to-analog converters or any number of binary outputs between one and twenty-four. The color selector is a unit such that any of the twenty-four channels of the disc can be assigned any of ten fixed colors and any of ten textures. The textures are merely series of horizontal and vertical lines superimposed on the displayed information. Figure 3 illustrates one image from a radar image pair and the IDECS processed image.

A SOFTWARE SYSTEM FOR DIGITAL IMAGE PROCESSING: KANDIDATS

KANDIDATS (Kansas Digital Image Data System), currently being developed, is a software package consisting of a monitor and a set of multi-image processing programs designed to run on a GE-635 computer. The multi-image processing programs are all written in FORTRAN IV and allow for image editing, registering, congruencing, quantizing, clustering, feature extraction, image size and/or dimensionality reduction, image texture analysis, image pattern recognition. It has a variety of decision rules, data display capability with scatterograms and histograms, grey-tone image display with overprinting or digital image color map display. The KANDIDATS monitor is a GMAP assembly language program designed to integrate the multi-image processing programs by handling all bookkeeping type and I/O operations and to minimize the cost of processing image data by speeding up I/O time and overlapping I/O time with execute time. Figure 4 illustrates a block diagram of the basic KANDIDATS organization.

The KANDIDATS monitor inputs in free-format all instructions required by the image processing program, supervises the execution of the programs, provides error processing, and dynamic storage allocation and tape input and output for the programs. The monitor has been written so that during a single activity of KANDIDATS many processing programs may be sequentially executed using many different data sets. The monitor does this by treating each program as a separate task and by allocating and releasing data tapes as necessary.

Once remotely sensed data is converted to digital type format, it is necessary to check the digitized tape to see of the conversion was made 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 KANDIDATS 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 image has so many resolution cells as to make the digital picture printing awkward, a program may be utilized which reduces the image size by averaging blocks of N x N resolutions or by selecting every N^{th} row and every N^{th} column.

Examination of this picture output will indicate what kind of editing will have to be done on the sides and top and bottom of the image as well as indicate 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 internal or external to the A/D conversion makes the conversion go awry.) If necessary, a KANDIDATS deskewing program may have to be used to remove skew and a special smoothing-replacement program may have to be used for those resolution cells which were improperly converted.

When multi-image data is being processed, it is often necessary to align the individual images to the same place. To do this KANDIDATS employs a registering program. When different sensors or the same sensors with different look directions are involved it may be necessary to bring the images to the same geometry. In this case a congruencing program must be used.

When the geometries on the images to be congruenced are quite different, the congruencing job may be quite hard. However, where only minor geometric distortions are involved, congruencing may be done by a KANDIDATS program which treats the image as a rubber surface and expands or contracts it to best match up a set of given corresponding points.

There are two formats by which multi-image data may be stored on taps by KANDIDATS. In the photo format all the grey-tones from the first image are stored on a matrix followed by the grey-tones from the second image and so on. In the corresponding point format the grey-tone from image one resolution cell (1,1) is followed by the grey-tone from image two resolution cell (1,1) and so on. Editing and congruencing imagery from different sensors is usually done with data in photo form as is image display and texture analysis. Most of the other programs work most easily with the data in corresponding point format. KANDIDATS has programs which convert multi-image data from one format to the other.

After initial editing and congruencing it is convenient to obtain an intuitive idea of what is happening in the data. To help with this, programs are available which pick out specified regions on the image and display the data points in scatterogram or histogram fashion. The scatterograms or histograms may be indexed by ground truth categories when the ground truth is available. The axes of the scatterograms may be combinations of pairs of the different sensor signals or the axes of a rotated coordinate system. Rotation can be accomplished from principal component analysis or from linear discriminant functions and there are programs available for these operations. Either of these operations will allow a significant reduction of dimensionality and, therefore, allow a reduction in storage and display of data, especially in 12 or 24 channel multi-spectral scanner data.

Before pattern discrimination or clustering is done, a feature extraction is performed which selects the relevant variables or which combines the original variables in some optimum way. Sometimes as part of the feature extraction process quantizing is done to normalize the data as well as to reduce the memory required for storage of the data. KANDIDATS has available programs which do equal interval, equal probability, minimum variance, and spatial quantizing.

When texture is an important feature for a category of interest, the dimensionality of the images may be augmented by a texture analysis program which adds dimensions providing texture type information.

Probably, the major workhorse of image data analysis consists of pattern discrimination and clustering techniques. 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, the 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 best as possible, to which environmental category an arbitrary data entity belongs. The disadvantage of the discrimination techniques is that they are sensitive to mis-calibrations. Any slight difference between the sensor calibrations or state of environment for the training data and the new data will cause error.

The advantage of the clustering techniques is that they are not sensitive to calibration problems. Two small-area patches of corn growing in the same field are going to be detected as being similar because they have similar grey tone associated with them. The disadvantage of the clustering techniques is that they are not able to identify the distinct environmental structures they determine.

KANDIDATS has available iterative and chaining clustering programs and pattern discrimination programs. The pattern discrimination programs use a variety of decision rule types including a distribution-free Bayes rule which can only be used on coarsely quantized data, a Bayes decision rule assuming the category conditional probabilities are of some given type of multivariate distribution, a linear decision rule, or a nearest neighbor decision rule.

Appendices I, II, and III summarize a few of the things we are doing with KANDIDATS.

CONCLUDING REMARKS

Two systems for processing remotely sensed image data have been discussed. The first system, IDECS, is a near real-time hardware system oriented towards processing multi-image data sets quickly and economically. The IDECS has convenient film input and color display output capabilities and implements simple kinds of decision rules. The second system, KANDIDATS, is a software system capable of performing many of the more sophisticated processing methods. Because of its monitor which handles all bookkeeping and its modular design, KANDIDATS easily allows

the testing of new automatic processing techniques. After a new technique has been proven on KANDIDATS, it may be simplified and hard-wired in IDECS, thereby keeping the volume processing of remotely sensed data always up to the current state-of-the-art.

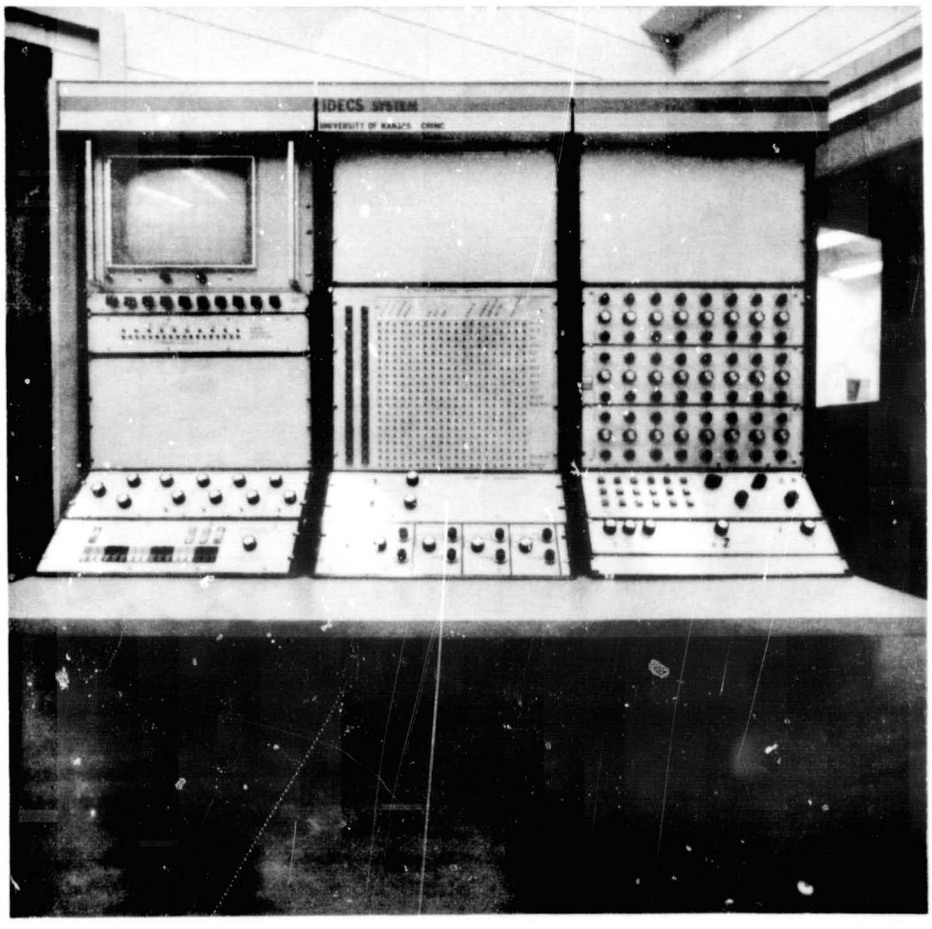


FIGURE 1. IDECS

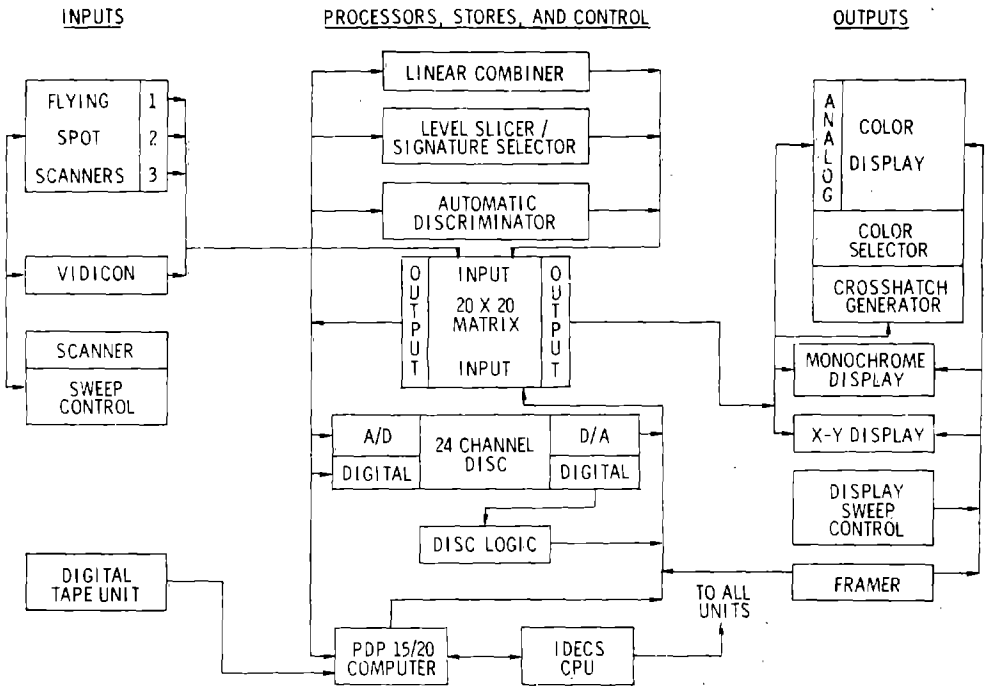


FIGURE 2. IDECS FUNCTIONAL BLOCK DIAGRAM

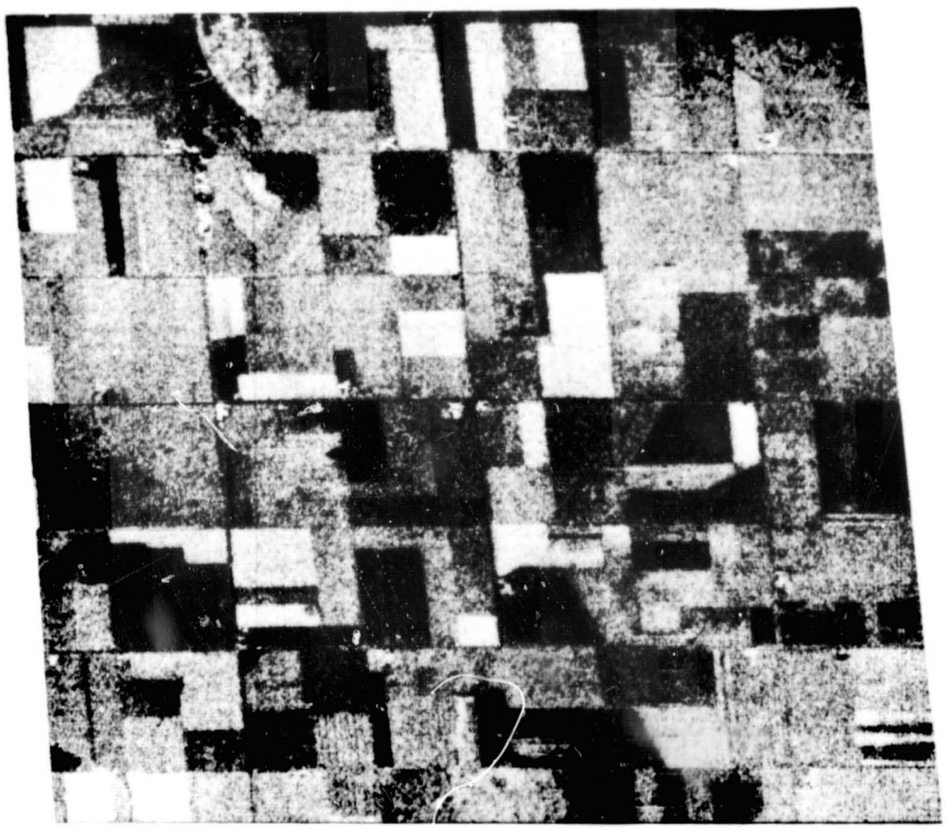
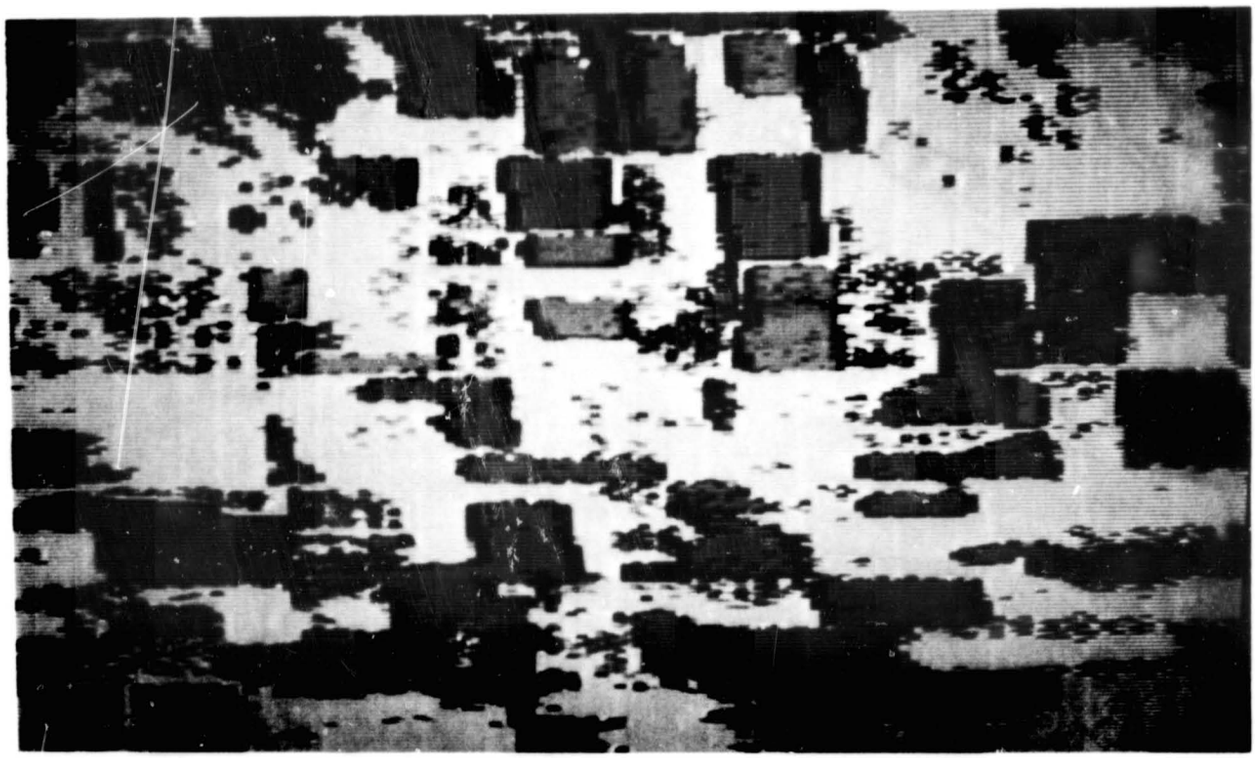


FIGURE 3a. RADAR IMAGE, HH POLARIZATION TAKEN OVER GARDEN CITY, KANSAS, JULY, 1966 BY WESTINGHOUSE AN/APQ 97.



Green Bare Ground

Blue Sugar Beets

Purple Sorghum and Alfalfa

Black No Decision

FIGURE 3b. THEMATIC LAND USE MAP PRODUCED BY IDECS FOR AN HH HV RADAR IMAGE PAIR TAKEN OVER GARDEN CITY, KANSAS, JULY, 1966, BY A WESTINGHOUSE AN/APQ 97. THE RESULTS SHOW 100% CORRECT IDENTIFICATION ON SUGAR BEETS AND BARE GROUND AND 85% CORRECT IDENTIFICATION ON SORGHUM AND ALFALFA.

BASIC KANDIDATS ORGANIZATION

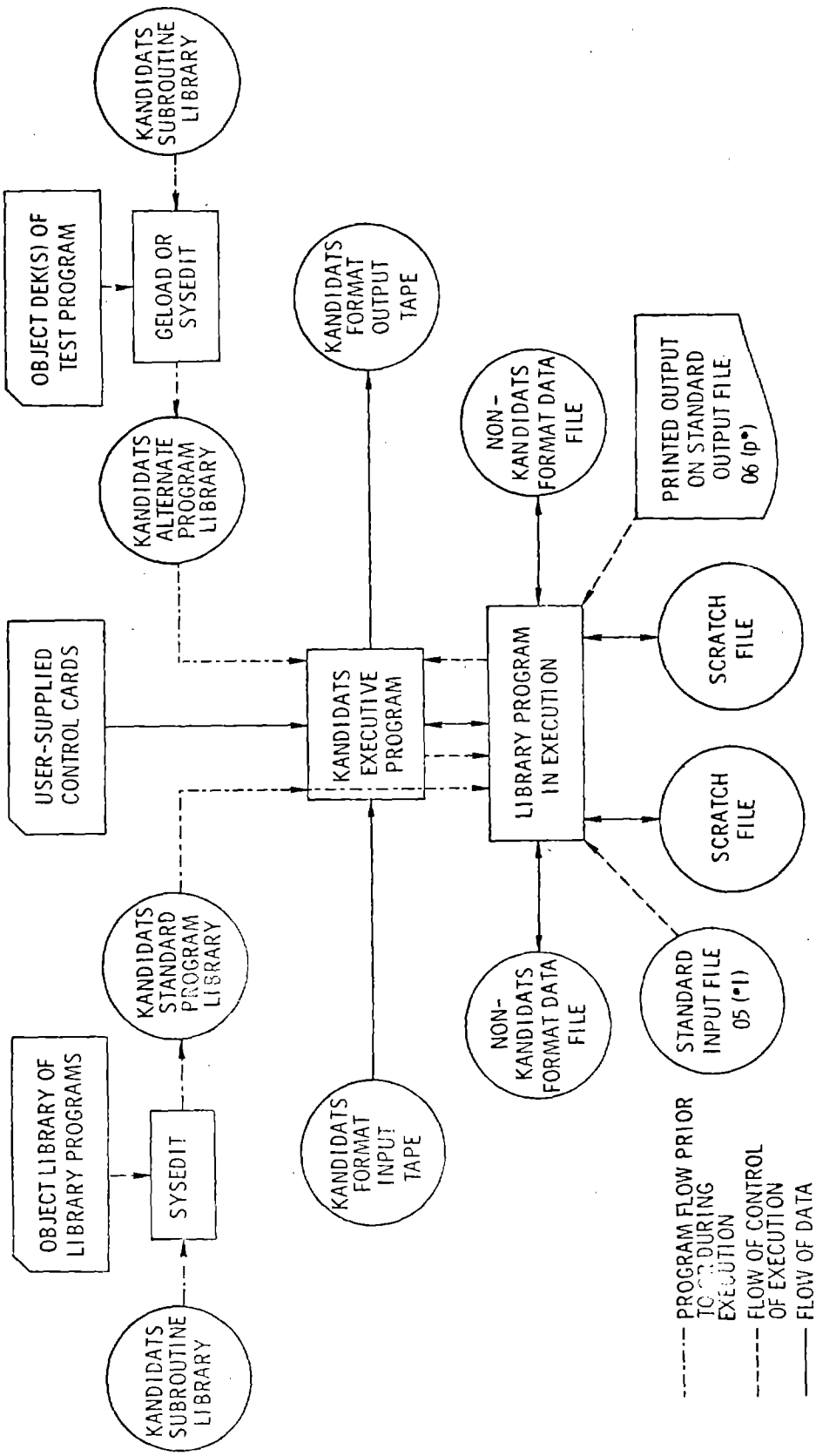


FIGURE 4.

APPENDIX I

by

Percy Batlivala
Center for Research, Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas

ENHANCEMENT AND NORMALIZATION OF RADAR IMAGE TEXTURE

Texture has been of interest to engineers and geoscientists alike because of its potential as a useful discriminant in image category identification. Hence, one important preprocessing operation must be concerned with the enhancement and normalization of image texture. Such an operation must bring out in normal form grey tone variation due to texture and exclude grey tone variations due to look angles or flight parameter fluctions.

Antenna patterns and flight parameter fluctuations have been two factors most responsible for degradation of radar imagery. If we regard the degradation as additive noise, enhancement of the image would, in a sense, be appropriate if there were means of removing the added noise. The 'streaks' parallel to the line of flight in an image could be due to flight parameter fluctuations, scratches caused by handling of the image before digitization, or due to antenna pattern and perpendicular 'streaks' could be due to scan lines.

Given below is a mathematical formulation, which in essence is the enhancement technique.

Let Lx and Ly be the x and y spatial domains, G be the set of grey tones and $P: Lx \times Ly \rightarrow G$ be some digital picture function of some more or less "homogeneous' object O, $O : Lx \times Ly \rightarrow G$.

The relationship between P and O is assumed to be of the following form:

$$P(i,j) = O(i,j) + \alpha(i) + \beta(j) \qquad (1)$$

where $\alpha(i)$ and $\beta(j)$ can be thought of as additive row and column distortion respectively. If we are interested in the texture of O, the average grey tone is not important and a function $\hat{P}(i,j)$ can be determined such that

$$\hat{P}(i,j) = P(i,j) - \hat{\alpha}(i) - \hat{\beta}(j) \qquad (2)$$

where

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \left[\hat{P}(i,j)\right]^2 \qquad (3)$$

is minimized.

The problem now is to estimate $\alpha(i)$ and $\beta(j)$ by $\hat{\alpha}(i)$ and $\hat{\beta}(j)$ so that (3) is minimized.

Let

$$F = \sum_{i=1}^{I} \sum_{j=1}^{J} [P(i,j) - \hat{\alpha}(i) - \hat{\beta}(j)]^2 \qquad (4)$$

To minimize F with respect to $\hat{\alpha}(i)$ and $\hat{\beta}(j)$ we take the partial derivatives, thus

$$\frac{\partial F}{\partial \hat{\alpha}(i)} = \sum_{j=1}^{J} [P(i,j) - \hat{\alpha}(i) - \hat{\beta}(j)](-2) \quad ; \; i = 1, 2, \cdots I \qquad (5)$$

and

$$\frac{\partial F}{\partial \hat{\beta}(j)} = \sum_{i=1}^{I} [P(i,j) - \hat{\alpha}(i) - \hat{\beta}(j)](-2) \quad ; \; j = 1, 2, \cdots J \qquad (6)$$

Setting the partial derivatives to zero.

$$P_{i\cdot} - J\hat{\alpha}(i) - \hat{\beta}_{\cdot} = 0 \quad ; \; i = 1, 2, \cdots, I \qquad (7)$$

$$P_{\cdot j} - \hat{\alpha}_{\cdot} - I\hat{\beta}(j) = 0 \quad ; \; j = 1, 2, \cdots, J \qquad (8)$$

Summing (7) over all i and multiplying (8) by J, we have

$$P_{\cdot\cdot} - J\hat{\alpha}_{\cdot} - I\hat{\beta}_{\cdot} = 0 \qquad (9)$$

$$JP_{\cdot j} - J\hat{\alpha}_{\cdot} - JI\hat{\beta}(j) = 0 \quad ; \; j = 1, 2, \ldots, J \qquad (10)$$

Subtracting equation (10) from (9) we have

$$-JP_{\cdot j} + I[J\hat{\beta}(j) - \hat{\beta}_{\cdot}] = 0 \quad ; \; j = 1, 2, \cdots, J \qquad (11)$$

Similarly by multiplying (7) by I and summing (8) over all j, we have

$$IP_{i\cdot} + J[I\hat{\alpha}(i) - \hat{\alpha}_{\cdot}] = 0 \quad ; \; i = 1, 2, \cdots, I \qquad (12)$$

The solutions to equations (11) and (12) are found to be

$$\hat{\alpha}(i) = \frac{P_{i\cdot}}{J} - \frac{\gamma P_{\cdot\cdot}}{IJ} \quad , \quad \hat{\beta}(j) = \frac{P_{\cdot j}}{I} - (1-\gamma)\frac{P_{\cdot\cdot}}{IJ} \qquad (13)$$

for any arbitrary γ.

Hence

$$\hat{\alpha}(i) + \hat{\beta}(j) = \frac{P_{i\cdot}}{J} + \frac{P_{\cdot j}}{I} - \frac{P_{\cdot\cdot}}{IJ} \tag{14}$$

and the enhanced image $\hat{P}(i,j)$ is then obtained by substituting equation (14) into equation (2) and

$$\hat{P}(i,j) = P(i,j) - \frac{P_{i\cdot}}{J} - \frac{P_{\cdot j}}{I} + \frac{P_{\cdot\cdot}}{IJ} \tag{15}$$

The enhanced image $\hat{P}(i,j)$ is found to have a zero mean, and also each row and column mean is zero.

Figure 1 shows a simulated 5 x 5 'homogeneous' image to which the enhancing technique has been applied. The 5 x 5 image shown in (c) of Figure 1 is the model with additive noise. The 5 x 5 enhanced image shown in (d) of Figure 1 clearly shows a 'diffusion' of the additive noise. For simplicity of representation, image (d) has been quantized, and therefore does not have a zero mean.

Figure 2 shows the enhancement technique applied to a radar image. Part (a) shows a digitized radar image of a sorghum field. This field was isolated from the radar image of a test site selected at Garden City, Kansas. The mission was conducted on September 15, 1965, by Westinghouse. Part (a) of Figure 1 shows a computer output of the original. Streaks running vertically and horizontally show up very clearly on the image. Part (b) shows the pictorial view of the 'noise' which was subtracted out of (a). Part (c) shows the enhanced image. All three images are represented by 13 grey tones and are quantized using an equal probability routine.

Figure 3 shows a larger area of the same test site and is made up of 14 fields. The images shown in the figure are positioned the same, relative to one another, as they were on the ground. Each image in the figure is a representation of the noise subtracted out from it. The streaks occurring in one field carry on into the neighboring fields. The vertical streaks (perpendicular to the line of flight) are almost periodic and, as stated earlier in this section, could be due to scan lines. The horizontal streaks may have been caused due to scratches on the negative or due to antenna pattern, but to pinpoint their cause at this stage, without further research, would be difficult.

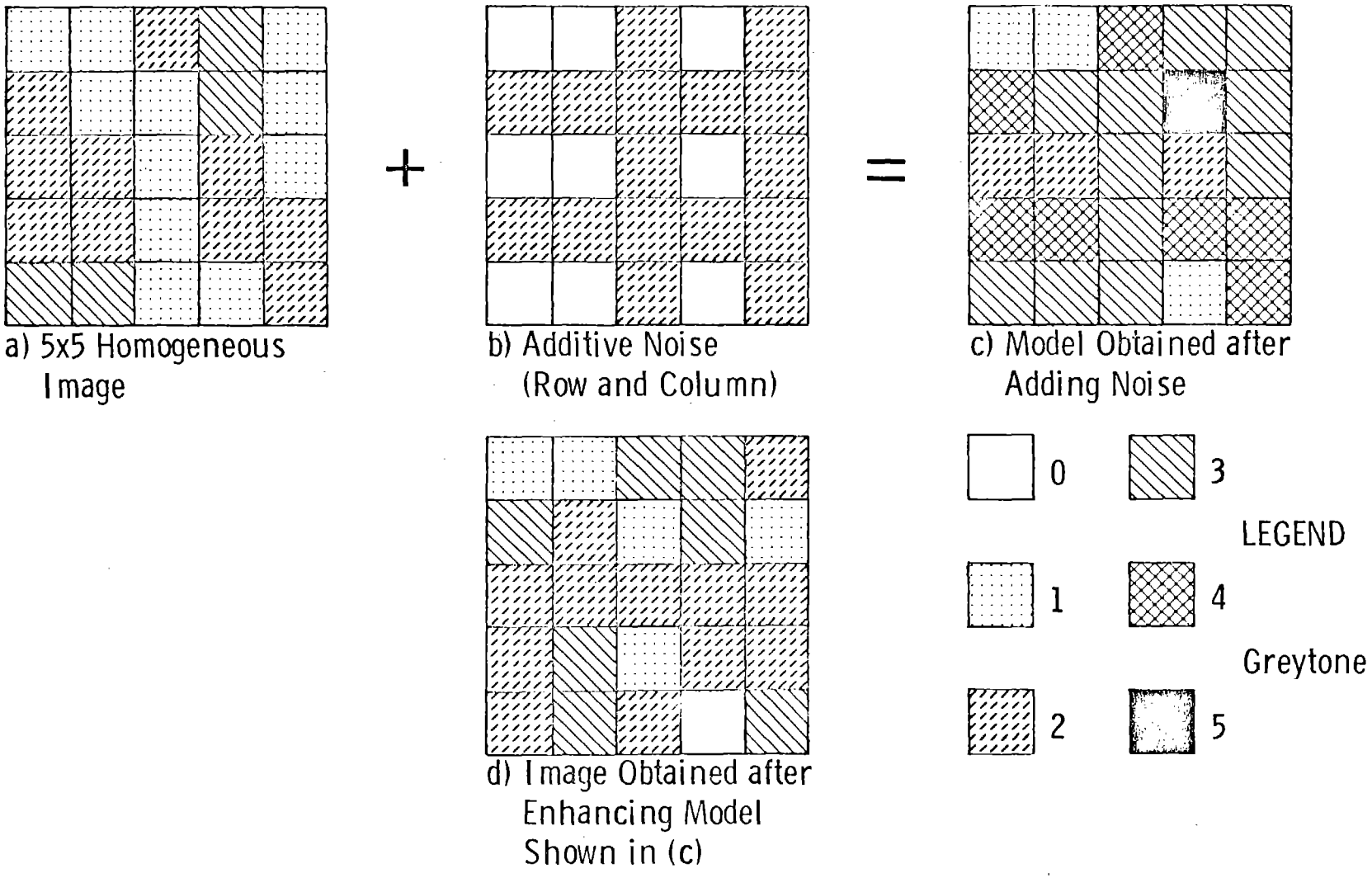
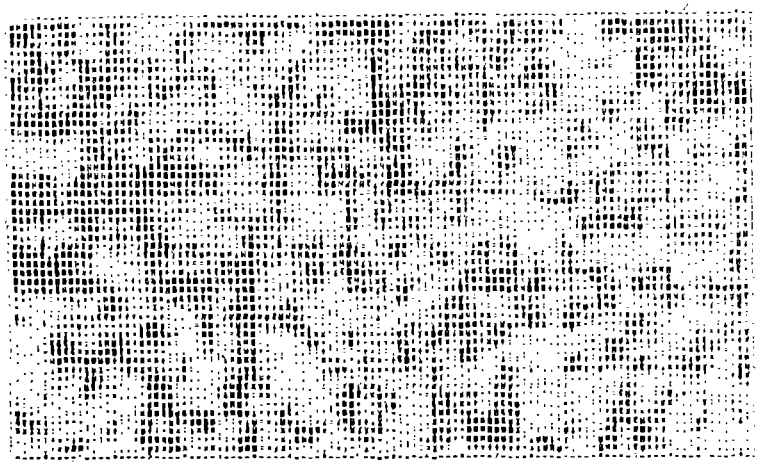
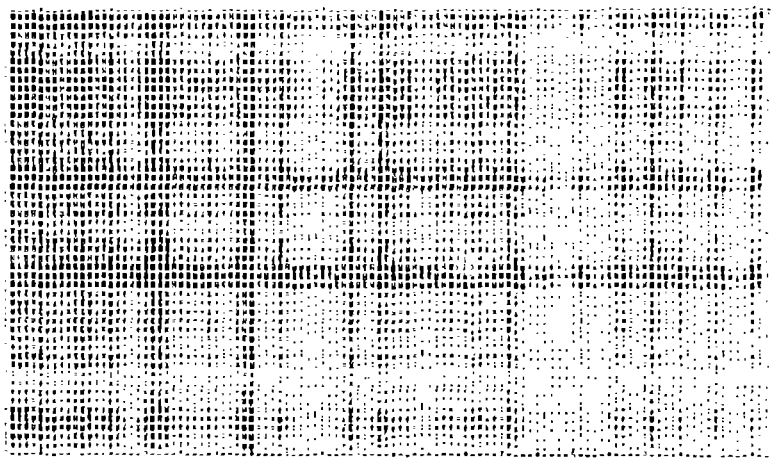


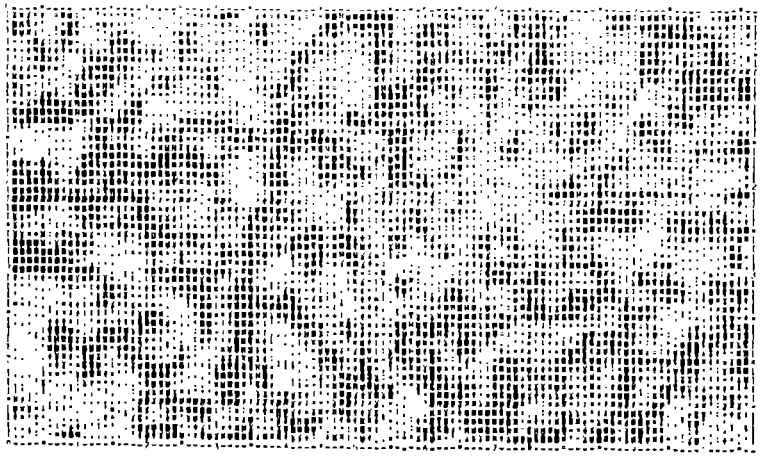
Figure 1, Shows a simulated 5x5 homogeneous image with additive noise (row and column) and its removal by the enhancement technique.



(a) Original image.



(b) Noise subtracted out of original.



(c) Enhanced image

Figure 2, Texture enhancement technique applied to a digitized radar image of a sorghum field (Garden City,Sept.,1965)

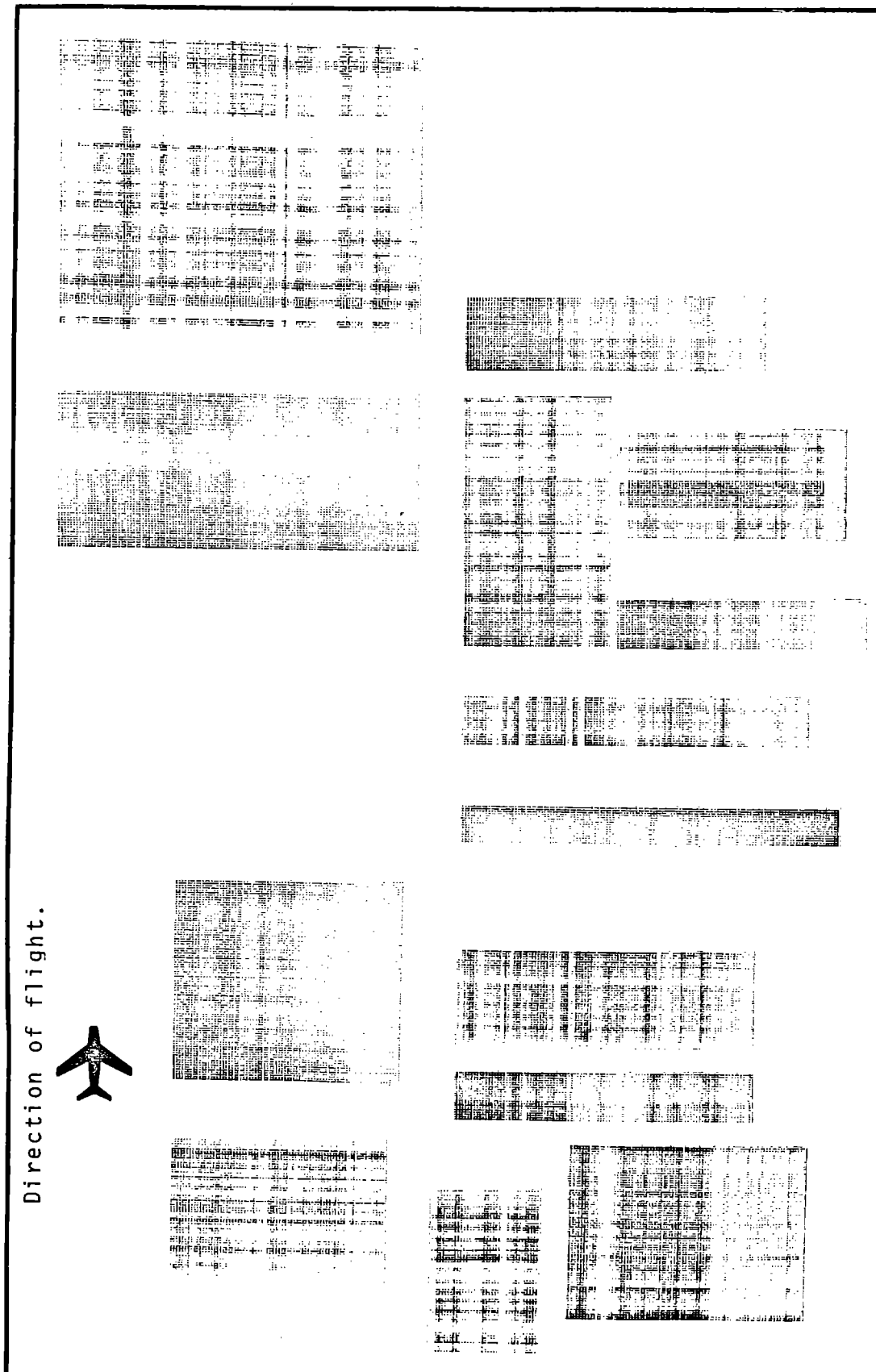


Figure 3, Part of the Garden City test site showing the 'noise' subtracted out of the 14 fields.

APPENDIX II

ON A TEXTURE-CONTEXT FEATURE EXTRACTION ALGORITHM FOR REMOTELY SENSED IMAGERY

R. M. Haralick
Center for Research, Inc.
Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas 66044

ABSTRACT

An image data set of 54 scenes was obtained from 1/8" by 1/8" areas on a set of 1:20,000 scale photography. The scenes which consisted of 6 samples from each of the nine categories scrub, orchard, heavily wooded, urban, suburban, lake, swamp, marsh, and railroad yard was analyzed manually and automatically.

For the automatic analysis, a set of features measuring the spatial dependence of the grey tones of neighboring resolution cells was defined. On the basis of these features and a simple decision rule which assumed that the features were independent and uniformly distributed an identification accuracy of 70% was achieved by training of 53 samples and assigning an identification to the 54th sample and repeating the experiment 54 times. This identification accuracy must be compared with the average 81% correct identification which five photointerpreters achieved with the same scenes, although the 81% correct identification is the accuracy achieved when they used the 9" x 9" photograph to interpret from. Note that the photograph is data of considerably higher resolution having much more context information on it than the small digitized 1/8" x 1/8" area the automatic analysis had available.

INTRODUCTION

The main problem facing us is that of feature selection of textural-contextual information. The features that we may use are limited not only by the catholic constraint of practicality*, but also by our heuristic idea of texture-context information.

In the next section we briefly go over the feature selection problem in general and in subsequent sections present the intuitive ideas behind what we have termed 'texture-context' features. The mathematical details of these features are then explained and some simple examples are shown. The decision making algorithms that were used are discussed; results, including comparison with interpretations by expert photointerpreters, and conclusions are in separate sections.

For this study the data sets were comprised of 54 digitized 1/8" x 1/8" sections of standard 1:20,000, 9" x 9" aerial photography supplied by the United States Army Engineer Topographic Laboratories. Each image was digitized into a 64 x 64 resolution cell matrix (and later into a 58 x 58 one because of some dark border effects encountered from the mask used in the digitization process), and the levels of digitization ranged from 63 to zero. There were six data sets per category and 9 general categories: scrub, orchard, heavily wooded, urban, suburban, lake, marsh, swamp, and railroad yard.

*With regard to this point, it seems appropriate to note here that all feature selection and decision making algorithms were written in FORTRAN IV and implemented on a PDP 15/20 digital computer with 12K core and two DEC tape drives.

PREPROCESSING AND FEATURE SELECTING

The 'classical' black-box description of an automated pattern recognition system is based on four main, not necessarily distinct, subsystems:

(1) the sensors or measuring instruments
(2) the preprocessors
(3) the feature selectors
(4) the categorizor or decision maker.

The data which the sensors or instruments produce are not always in the kind of normalized form with which it makes sense to work. For example, many sensors or measuring instruments produce relative measurements, i.e. the measurements are correct up to an additive or multiplicative constant. Despite calibration efforts, this is particularly true for the camera-film-digitizer system which produce the digital magnetic tape containing the digitized image. Variations in lighting, lens, film, developer, and digitizer all combine to produce a grey tone value which is an unknown but usually monotonic transformation of the "true" grey tone value. Under these conditions we would certainly want two images of the same scene, one image being a grey tone monotonic transformation of the other, to produce the same results from the pattern recognition process. It has been shown that normalization by equal probability quantizing guarantees that images which are monotonic transformations of one another produce the same results. Hence, all the images we used were quantized to 16 levels.

The sensors usually produce many measurements. Simple sensors such as an EKG machine produce $10^3 - 10^4$ sampled values while image sensors produce $10^4 - 10^7$ sensed grey tones. Compared to the huge amount of data produced by the sensors, the category distinctions we need to make are relatively few, say a choice of one out of ten to a hundred categories. This suggests that the pattern recognition system should be able to reduce the data to a more succinct form, eliminating much extraneous information (that information which is, in general, not relative to the discrimination of the given categories). This sort of data reduction which produces the initial features is called preprocessing or feature selecting and unfortunately there exists little or no theory to aid in establishing what this preprocessing or feature selecting should consist of. Rather, this operation is determined intuitively, rationalized heuristically and justified later progmatically and empirically. In the case of our texture-context problem, the use of various moments of the spatial grey tone dependence matrices corresponds to this sort of preprocessing or initial feature selecting.

Research was supported by U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia, CONTRACT DAAK02-70-C-0388.

It is important to note that a primary characteristic of preprocessing or feature selecting is the number of operations needed to be performed in order to obtain the features. Quick procedures are characterized by a number of operations proportional to the number of data points needing to be processed. All procedures which we develop here are quick in that sense.

The next stage in feature selecting consists of removing redundancies from the initial features. If the initial features are N-dimensional vectors in Euclidean N-space, as they are in our study, then it might be that all the vectors lie in some K-dimensional flat where K is much smaller than N. In this case there are N-K linear constraints to which the initial feature vectors are subject and it is possible to essentially maintain all the information in the features vectors by representing them by their coordinates in the smaller dimensional subspace or flat. Such redundancy removal can be done by principal component analysis or by not using those features which do not contribute additional information for the identification of the given categories. It is this latter approach which we take here.

The various features which we suggest are all a function of distance and angle. The angular dependence presents a special problem. Suppose image A has features a,b,c,d for angles $0^\circ, 45^\circ, 90^\circ, 135^\circ$ respectively and image B is identical with image A except that image B is rotated by say 90° with respect to A. Then B will have features c,d,a,b for angles $0^\circ, 45^\circ, 90^\circ, 135^\circ$ respectively. Since the texture-context of A is the same as B, any decision rule using the angular depence features a,b,c,d must produce the same results for c,d,a,b (a 90° rotation) or for that matter b,c,d,a (a 45° rotation) and d,a,b,c (a 135° rotation). To guarantee this we do not use the angular dependent feature directly. Instead, we use some symmetric function of a,b,c,d: their average, range, and mean deviation.

SPATIAL GREY TONE DEPENDENCE

Let $L_x = \{1,2,\ldots N_x\}$ and $L_y = \{1,2,\ldots,N_y\}$ be the x and y spatial domains and $L_y \times L_x$ be the set of resolution cells. Let $G = \{0,1,\ldots,N_g\}$ be the set of possible grey tones. Then a digital image I is a function which assigns some grey tone to each and every resolution cell; $I: L_y \times L_x \rightarrow G$.*

An essential component of our conceptual framework of texture is a measure, or more precisely, four closely related measures from which all of our texture-context features are derived. These measures are arrays termed angular nearest neighbor grey tone spatial dependence matrices, and to describe these arrays we must re-emphasize our notion of adjacent or nearest neighbor resolution cells themselves. We consider a resolution cell -- excluding those on the periphery of an image, etc.—to have eight nearest neighbor resolution cells as in Figure 1.

*The spatial domain $L_y \times L_x$ consists of ordered pairs whose components are row and column respectively. This convention conforms with the usual two subscript row-column designation used in FORTRAN.

We assume that the texture-context information in an image I is contained in the over-all or "average" spatial relationship which the grey tones in image I have to one another. More specifically, we shall assume that this texture-context information is adequately specified by the matrix of relative frequencies P_{ij} with which two neighboring resolution cells separated by distance d occur on the image, one with grey tone i and the other with grey tone j. Such matrices of spatial grey tone dependence frequencies are a function of the angular relationship between the neighboring resolution cells as well as a function of the distance between them. Figure 2 illustrates the set of all horizontal neighboring resolution cells separated by distance 1. This set along with the image grey tones would be used to calculate a distance 1 horizontal spatial grey tone dependence matrix. Formally, for angles quantized to 45° intervals the unnormalized frequencies are defined by:

$$P(i,j,d,0^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid k-m=0,\ |l-n|=d,\ I(k,l)=i,\ I(m,n)=j\}$$

$$P(i,j,d,45^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k-m=d,\ l-n=-d) \text{ or } (k-m=-d,\ l-n=d),\ I(k,l)=i,\ I(m,n)=j\}$$

$$P(i,j,d,90^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid |k-m|=d,\ l-n=0,\ I(k,l)=i,\ I(m,n)=j\}$$

$$P(i,j,d,135^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k-m=d,\ l-n=d) \text{ or } (k-m=-d,\ l-n=-d),\ I(k,l)=i,\ I(m,n)=j\}$$

Note that these matrices are symmetric; $P(i,j;d,a) = P(j,i;d,a)$. The distance metric ρ implicit in the above equations can be explicitly defined by $\rho((k,l),(m,n)) = \max\{|k-m|, |l-n|\}$.

Consider Figure 3-a, which represents a 4 x 4 image with four grey tones, ranging from 0 to 3. Figure 3-b shows the general form of any grey tone spatial dependence matrix. For example, the element in the (2,1)-st position of the distance 1 horizontal P_H matrix is the total number of times two grey tones of value 2 and 1 occurred horizontally adjacent to each other. To determine this number, we count the number of pairs of resolution cells in R_H such that the first resolution cell of the pair has grey tone 2 and the second resolution cell of the pair has grey tone 1. In Figures 3-c through 3-f we calculate all four distance 1 grey tone spatial dependence matrices.

If needed, the appropriate frequency normalization for the matrices are easily computed. When the relationship is nearest horizontal neighbor (d=1 and $\alpha=0^\circ$), there will be $2(N_x-1)$ neighboring resolution cell pair on each row and there are N_y rows providing a total of $2N_y(N_x-1)$ nearest horizontal neighbor pairs (see Figure 3). When the relationship is nearest right diagonal neighbor (d=1, $\alpha=45^\circ$) there will be $2(N_x-1)$ 45° neighboring resolution cell pairs for each row except the first, for which there are none, and there are N_y rows. This provides a total of $2(N_y-1)(N_x-1)$ nearest right diagonal neighbor pairs (see Figure 4). By symmetry there will be $2N_x(N_y-1)$ nearest vertical neighbor pairs and $2(N_x-1)(N_y-1)$ nearest left diagonal neighbor pairs.

Let us now consider how to use such spatial dependence information. We have suggested generating a homogeneity and unhomogeneity image from the original image on the basis of the grey tone dependence matrix. (The homogeneity image is an enhanced display of all

the homogeneous areas while the unhomogeneity image is an enhanced display of all the unhomogeneous areas.) At any resolution (m,n), the homogeneity image I_h has an integer valued grey tone 0 to 8 depending on how many of resolution cell (m,n)'s 8 nearest neighbors on the original image I have respective grey tones which are "sufficiently similar" to the grey tone at (m,n) on image I. Similarity of two grey tones i and j is determined on the basis of whether the grey tones occur next to each other sufficiently often; that is, if the element P(i,j) of the spatial dependence matrix is large enough. At any resolution cell (m,n), the unhomogeneity image I_u has a grey tone 0,1,2,3,4,5,6,7 or 8 depending on how many of resolution cell (m,n)'s 8 nearest neighbors on the original image I have respective grey tones which are "sufficiently dissimilar" to the grey tone at (m,n) on image I. Dissimilarity of two grey tones i and j is determined on the basis of whether the grey tones occur next to each other sufficiently rarely, that is, if the element P(i,j) of the spatial grey tone dependence matrix is small enough.

The idea of large enough or small enough implies a thresholding of the grey tone dependence matrix and depending on what level the threshold is set the resulting homogeneity and unhomogeneity images appear differently. Thus undesirable arbitrary thresholds must be introduced. Fortunately, it is possible to do away with thresholding. Instead of defining similarity as an all or nothing affair, we can define the similarity between grey tones i and j to be P(i,j), the frequency with which i and j co-occur next to each other, some function of P(i,j) such as $\log P(i,j)$ or perhaps even some function of i and j such as $\frac{1}{1+(i-j)^2}$. Dissimilarity between i and j can be measured by $(i-j)^2$.

Texture-context features are easily derived from the homogeneity or unhomogeneity image. For example, the greater the total homogeneous region area, then the darker the homogeneity image. Hence, the mean grey tone of the homogeneity image provides a measure of the "smoothness" of the original image. The grey tone variance of the homogeneity image provides a measure of how the homogeneous areas are spread out on the image. Low variance would indicate large area uniform homogeneity while high variance might indicate many small area homogeneous regions.

It can be shown that the computation of the average grey tone on the homogeneity or unhomogeneity image J can be done without having to have the image J generated. The average grey tone can be computed directly as a function of the spatial grey tone dependence matrices. In this paper we explore only those features which can be computed directly from the spatial dependence grey tone matrix and do not require the homogeneity or unhomogeneity image to be determined.

In the discussion which follows on the use of the spatial grey tone dependence matrices as texture context features for image data, we shall be concerned with forms such as

$$\sum_{a=1}^{N_g}\sum_{b=1}^{N_g}(a-b)^2 P(a,b)/{}^{\#}R$$, the angular second moment difference (ASMD);

$$\sum_{a=1}^{N_g}\sum_{b=1}^{N_g}\frac{1}{1+(a-b)^2}P(a,b)/{}^{\#}R$$, the angular second moment inverse difference (ASMID);

$$\sum_{a=1}^{N_g}\sum_{b=1}^{N_g}\left[\frac{P(a,b)}{{}^{\#}R}\right]^2$$, the angular second moment (ASM);

$$\frac{COV}{VAR}$$, the correlation between neighboring grey tones (COR);

Note: $^{\#}R$ is the number of neighboring resolution cells.

The ASM feature is the sum of the squared terms of the grey tone spatial dependence matrix normalized by the total number of possible adjacencies, $^{\#}R$, for the given angle. For each spatial dependence matrix, there is a corresponding ASM but there has been a great reduction of data because each ASM (as each ASMD and ASMID) is only a number not an array. The ASMD feature is the sum of the members of a grey tone spatial dependence matrix, each member multiplied by the squared difference of the grey tone values and normalized as before. The ASMID feature is the sum of the members of a grey tone spatial dependence matrix, each member divided by one plus squared grey tone difference. The correlation feature COR is actually the value of the two-dimensional autocorrelation function of the picture where the autocorrelation function is evaluated for a particular distance and angle lag.

Each of these features is a function of the angle and distance between what we consider to be neighboring resolution cells. We consider 4 angles, 0°, 45°, 90°, and 135° at distances of 1, 3, and 9 resolution cells. This provides an initial set of 48 features. The number of features is thus reduced by calculating the mean, range, and mean deviation of each type of feature at a given distance over the four angles. The features which are actually first considered by the decision rule are ASM, ASMID, ASMD, COR evaluated at distances of 1, 3, and 9 resolution cells with the average, range, and mean deviation for each feature and distance calculated over the four angles. This is a total of 36 features. Figure 5 illustrates the calculation of three representative features of the image of Figure 3a.

AUTOMATIC SCENE IDENTIFICATION

Automatic scene identification using the 36 texture context features presents a difficult problem because of the relative sparcity of the data: for each of 9 categories there are only 6 samples with each sample having a 36 dimensional feature vector. The difficulty is really twofold: (1) There are so few samples that it is difficult to learn anything about the patterns which are characteristic of the category, (2) The decision rule must contain a minimum of parameters so that the decision rule does not "memorize" the data. Hence the approach we take relies on the simplest type of data statistics: the minimum and maximum value each feature

can take on for measurements in a given category.

Figure 6 illustrates for each pair of categories which variable will separate them. Figure 6 shows that variable 4, ASMID at distance 1, has its average, range and mean deviation appearing a total of 56 times in separating categories. Of those categories which are not separated by the distance 1 ASMID features, COR at a distance 1 has its average, range and mean deviation appearing a total of 8 times in separating categories. Of those categories which are not separated by the distance 1 ASMID or COR features, ASM at distance 1 has its average, range and mean deviation appearing a total of 4 times in separating categories. Of those categories which are not separated by distance 1 ASMID, COR or ASM features, ASM at distance 3 has its average, range and mean deviation appearing a total of 1 time in separating categories. Hence, of the initial 36 features, we use only the following 12 features:

ASM	AVG	
ASM	RANGE	
ASM	DEV	
COR	AVG	DISTANCE 1
COR	RANGE	
COR	DEV	
ASMID	AVG	
ASMID	RANGE	
ASMID	DEV	
ASM	AVG	
ASM	RANGE	DISTANCE 3
ASM	DEV	

For automatic identification, we use a decision rule which is a maximum likelihood decision rule under the assumptions that the 12 feature variables are independent having uniform distributions. Under this assumption, the density function for the k^{th} category is

$$f(x_1,x_2,\ldots,x_{12}|k) = \prod_{n=1}^{12} \frac{1}{(a_{nk}-b_{nk})} \quad \text{for all}$$

$(x_1,x_2,\ldots,x_{12})$ such that

$$b_{nk} \le x_n \le a_{nk},\ n=1,2,\ldots,12,$$

where b_{nk} and a_{nk} define the minimum and maximum values of the uniform distribution on the n^{th} component.

Hence, a measurement $(x_1,x_2,\ldots,x_{12})$ is assigned to category k if and only if

(1) $b_{nk} \le x_n \le a_{nk}$, $n=1,2,\ldots 12$ and

(2) $$\prod_{n=1}^{12} \frac{1}{(a_{nk}-b_{nk})} \ge \prod_{n=1}^{12} \frac{1}{(a_{nj}-b_{nj})}$$

for all j such that $b_{nj} \le x_n \le a_{nj}$, $n=1,2,\ldots 12$.

If there exists no k such that $b_{nk} \le x_n \le a_{nk}$, $n=1,2,\ldots 12$, then $(x_1,x_2,\ldots x_{12})$ is assigned to category k if and only if

$$\sum_{n=1}^{12} \min\{|x_n-a_{nk}|, |x_n-b_{nk}|\}\ (a_{nk}-b_{nk}) \ge \sum_{n=1}^{12} \min\{|x_n-a_{nj}|, |x_n-b_{nj}|\}\ (a_{nj}-b_{nj}),$$

$j=1,2,\ldots,K.$

The minimum and maximum statistics a_{nk} and b_{nk} are estimated in the following way:

Let α_{nk} = the maximum n^{th} component for all measurements designated in k^{th} category;

β_{nk} = the minimum n^{th} component for all measurements designated in k^{th} category.

Assume that category k has M_k measurements, then

$$b_{nk} = \beta_{nk} - \frac{(\alpha_{nk}-\beta_{nk})}{M_k-1}$$

$$a_{nk} = \alpha_{nk} + \frac{(\alpha_{nk}-\beta_{nk})}{M_k-1}$$

Notice that a_{nk} is larger than the maximum by some fraction of the range and b_{nk} is smaller than the minimum by some fraction of the range. Hence, the range $a_{nk}-b_{nk}$ is larger than $\alpha_{nk}-\beta_{nk}$. Under the assumption that the variable has a uniform distribution, the expected value of $\alpha_{nk}-\beta_{nk}$ is $\frac{M_k-1}{M_k+1}$ · (true range) while the expected value of $a_{nk}-b_{nk}$ is the true range.

The identification experiment was done in two ways using the above decision rule. In the first case the entire set of 54 samples was used to train on, i.e. gather the statistics α_{nk}, β_{nk}, $n=1,2,\ldots 12$, $k=1,2,\ldots 9$, and then on the basis of the α_{nk}'s and β_{nk}'s calculated, each sample was assigned to a category. Figure 7 illustrates the contingency table of the resulting assignments. A total of 53 out of 54 samples were correctly identified. We shall have more to say about the interpretation of 53/54 in a moment.

In the second case, the identification experiment was repeated 54 times, each time using a different set of 53 samples to train on. The 54th sample was assigned to a category on the basis of the minimum maximum statistics gathered from the other 53 samples. Figure 8 illustrates the contingency table resulting from these assignments. A total of 38 out of 54 samples were correctly identified for a percentage of approximately 70%.

To help interpret these results a sequential decision algorithm in the form of a dichotomous key was tried. A dichotomous key successively splits a group of measurements in two based on whether a given component is greater than or less than some value. The dichotomous key itself is illustrated in Figure 9. It takes 13 decision points to perfectly separate the 54 measurements into their designated categories. If the 5 decision points, whose sole function is to correctly separate measurements

which were incorrectly assigned, are removed, then it takes 8 decisions to correctly assign 49/54 measurements. Under the assumption that the twelve variables are independent at each decision stage, and that the two category groups being split have the same uniform distribution, Figure 10 illustrates the contingency table resulting from these assignments. Under the assumption that the twelve variables are independent at each decision stage, and that the two category groups being split have the same uniform distribution, the probability of being able to achieve perfect separation of two categories in two decisions is less than 10^{-12}.

The automatic texture-context scene analysis experiment was compared with the identification which five photointerpreters were able to make with the same data set. The photointerpreters were given the original 9"x9" photographs and were allowed to use as much context information as they could in making the identification. These experiments yielded an average of 81% correct identification for the five photointerpreters.

DISCUSSION AND CONCLUSION

An image data set of 54 scenes consisting of 6 samples from each of the nine categories scrub, orchard, heavily wooded, urban, suburban, lake, swamp, marsh, and railroad yard was analyzed manually and automatically.

For the automatic analysis, a set of features for texture context was defined and on the basis of these features and a simple decision rule, an identification accuracy of 70% was achieved. This identification accuracy must be compared with the average 81% correct identification which five photointerpreters achieved with the same scenes, although the 81% correct identification is the accuracy achieved when they used the 9"x9" photograph to interpret from. The photograph is data of considerably higher resolution having much more context information on it than the small digitized 1/8"x1/8" area the automatic analysis had available.

Furthermore, the 70% correct identification arose in the case when the automatic technique trained on 53 samples and assigned an identification to the 54th sample and repeated the experiment 54 times. This means that for each experiment, there was one category which had 5 samples instead of 6 samples. For this category, there is a probability of 1/3 that for each feature the missing sample had minimum or maximum value over all samples for the category. Hence there is a high probability that the missing sample provided significant information which is not available in the sample without it.

Looking at the situation another way, 100% correct identification was achieved by the optimal dichotomous key which required 13 decision points. The probability that such good identification could happen by chance is very small. In fact, the number of 2 celled partitions which the simple hyperplanes used could generate for N samples in a d-dimensional space is only d(N-1) and this number should be compared to $2^{N-1}-1$, the total number of non-trivial distinct 2 celled partitions possible. In our case d=12, N=54 and the ratio

$$\frac{d(N-1)}{2^{N-1}-1} = \frac{12\times 53}{2^{53}-1} \leq \frac{2^{10}}{2^{53}} = 2^{-43}$$

Hence our ability to perform the category separation with such a small chance of available partitions is significant.

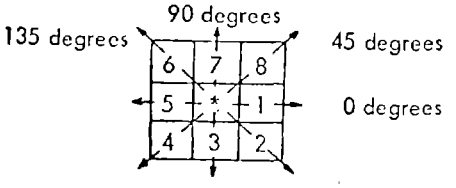


Figure 1. Resolution cells nos. 1 and 5 are the 0-degree (horizontal) nearest neighbors to resolution cell '*', resolution cells nos. 2 and 6 are the 135-degree nearest neighbors, resolution cells 3 and 7 are the 90-degree nearest nearest neighbors, and resolution cells 4 and 8 are the 45-degree nearest neighbors to '*'. (Note that this information is purely spatial, and has nothing to do with grey tone values.)

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

$L_y = \{1,2,3,4\}$

$L_x = \{1,2,3,4\}$

$$P_H = \{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid k-m = 0,\ |l-n| = 1\}$$

$$= \{((1,1),(1,2)), ((1,2),(1,1)), ((1,2),(1,3)), ((1,3),(1,2)), ((1,3),(1,4)), ((1,4),(1,3)), ((2,1),(2,2)), ((2,2),(2,1)), ((2,2),(2,3)), ((2,3),(2,2)), ((2,3),(2,4)), ((2,4),(2,3)), ((3,1),(3,2)), ((3,2),(3,1)), ((3,2),(3,3)), ((3,3),(3,2)), ((3,3),(3,4)), ((3,4),(3,3)), ((4,1),(4,2)), ((4,2),(4,1)), ((4,2),(4,3)), ((4,3),(4,2)), ((4,3),(4,4)), ((4,4),(4,3))\}$$

Figure 2 illustrates the set of all distance 1 horizontal neighboring resolution cells on a 4 by 4 image.

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Figure 3-a.

		Grey Tone			
		0	1	2	3
Grey Tone	0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
	1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
	2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
	3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

Figure 3-b. This shows the general form of any grey tone spatial dependence matrix for an image with integer grey tone values 0 to 3. #(i,j) stands for number of times grey tones i and j have been neighbors.

0°

$$P_H = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Figure 3-c.

90°

$$P_V = \begin{pmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

Figure 3-d.

135°

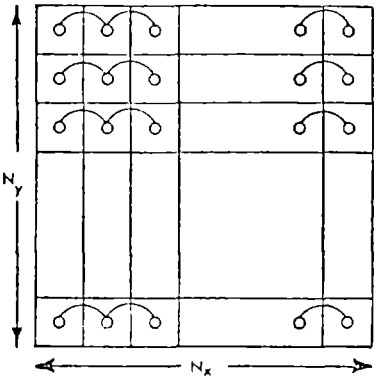
$$P_{LD} = \begin{pmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

Figure 3-e.

45°

$$P_{RD} = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Figure 3-f.



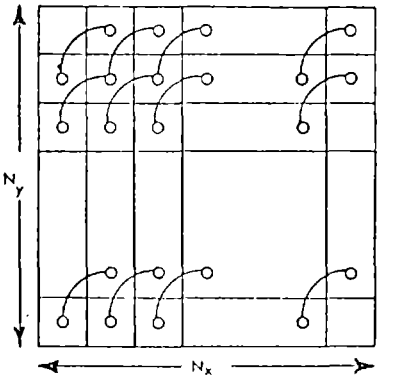


Figure 4 illustrates how the members of horizontal and right diagonal neighboring resolution cells are counted.

$$ASM\ (0^\circ) = \frac{1}{(24)^2}(4^2 + 2^2 + 1^2 + 0^2 + 2^2 + 4^2 + 0^2 + 0^2 + 1^2 + 0^2 + 6^2 + 1^2 + 0^2 + 0^2 + 1^2 + 2^2) = \frac{84}{(24)^2} = .14583$$

$$ASMD(90^\circ) = \frac{1}{24}(6\cdot0 + 0\cdot1 + 2\cdot4 + 0\cdot9 + 0\cdot1 + 4\cdot0 + 2\cdot1 + 0\cdot4 + 2\cdot4 + 2\cdot1 + 2\cdot0 + 2\cdot2 + 0\cdot9 + 0\cdot4 + 2\cdot1 + 0\cdot0) = 1.083$$

$$ASMID(135^\circ) = \frac{1}{18}\left(\frac{2}{1} + \frac{1}{2} + \frac{3}{5} + \frac{0}{10} + \frac{1}{2} + \frac{2}{1} + \frac{1}{2} + \frac{0}{5} + \frac{3}{5} + \frac{1}{2} + \frac{0}{1} + \frac{2}{2} + \frac{0}{10} + \frac{0}{5} + \frac{2}{2} + \frac{0}{1}\right) = .5111$$

$$COR(45^\circ) = \frac{COV(45^\circ)}{VAR(45^\circ)} = \frac{2.111 - 1.4983}{2.333 - 1.4983} = .734$$

where $COV(45^\circ) = \frac{1}{18}(0\cdot0\cdot4 + 0\cdot1\cdot1 + 0\cdot2\cdot0 + 0\cdot3\cdot0 + 1\cdot0\cdot1 + 1\cdot1\cdot2 + 1\cdot2\cdot2 + 1\cdot3\cdot0 + 2\cdot0\cdot0 + 2\cdot1\cdot2 + 2\cdot2\cdot4 + 2\cdot3\cdot1 + 3\cdot0\cdot0 + 3\cdot1\cdot0 + 3\cdot2\cdot1 + 3\cdot3\cdot0) - \frac{1}{(18)^2}(0\cdot5 + 1\cdot5 + 2\cdot7 + 3\cdot1)^2$

$$VAR(45^\circ) = \frac{1}{18}(0^2\cdot5 + 1^2\cdot5 + 2^2\cdot7 + 3^2\cdot1) - \frac{1}{(18)^2}(0\cdot5 + 1\cdot5 + 2\cdot7 + 3\cdot1)^2$$

Figure 5 illustrates the calculation of texture context features at distance 1 for the image of Figure 3a.

SCRU	ORCHR
AVG	1
DEV	7
AVG	9
AVG	12

SCRU	H WOO
AVG	3
AVG	7

SCRU	URBA
AVG	1
AVG	2
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	7
AVG	8
AVG	9
RANGE	9
DEV	9
AVG	12

SCRU	SUBUR
AVG	1
RANGE	2
DEV	2
RANGE	3
DEV	3
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	6
AVG	8
AVG	8
AVG	9
AVG	12

SCRU	LAK
RANGE	3
DEV	3
AVG	3
AVG	5
AVG	7
AVG	9

SCRU	MARS
DEV	2
RANGE	3
DEV	3
AVG	4
DEV	7

SCRU	SWAM
AVG	1
AVG	2
RANGE	2
DEV	2
RANGE	3
DEV	3
AVG	3
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	6
AVG	6
AVG	7
AVG	8
AVG	9
AVG	10
AVG	12

SCRU	RAILY
AVG	1
RANGE	2
DEV	2
RANGE	3
DEV	3
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	7
AVG	8
AVG	9
AVG	10
AVG	12

ORCHR	H WOO
AVG	1
AVG	2
AVG	5
RANGE	8
DEV	8
AVG	8
AVG	9

ORCHR	URBA
AVG	1

ORCHR	SUBUR
RANGE	2
DEV	2
DEV	3
RANGE	4
DEV	4
AVG	4
AVG	8

ORCHR	LAK
AVG	1
AVG	2
DEV	3
AVG	3
AVG	5
DEV	7
AVG	8
RANGE	8
DEV	8
AVG	9
AVG	12

ORCHR	MARS
RANGE	2
DEV	2
DEV	3
AVG	5
RANGE	6
DEV	6
AVG	9
AVG	12

ORCHR	SWAM
AVG	2
RANGE	2
DEV	2
RANGE	3
DEV	3
AVG	3
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	6
DEV	8
AVG	8
AVG	9
AVG	12

H WOO	SUBUR
AVG	1
AVG	2
AVG	3
RANGE	4
DEV	4
AVG	5
RANGE	6
DEV	6
RANGE	8
DEV	8
AVG	8
AVG	9
AVG	12

H WOO	LAK
AVG	5
AVG	9

H WOO	MARS
AVG	4
AVG	8

H WOO	SWAM
AVG	1
AVG	2
RANGE	2
DEV	2
RANGE	3
DEV	3
AVG	3
RANGE	4
DEV	4
AVG	4
AVG	5
RANGE	5
DEV	5
AVG	6
AVG	6
RANGE	6
DEV	6
AVG	7
AVG	8
AVG	9
AVG	10
AVG	12

H WOO	URBA
AVG	1
AVG	2
AVG	3
RANGE	4
DEV	4
AVG	4
AVG	5
RANGE	6
DEV	6
RANGE	8
DEV	8
AVG	8
AVG	9
AVG	12

URBA	RAILY
RANGE	3

SUBUR	LAK
AVG	1
AVG	2
AVG	3
AVG	4
RANGE	4
DEV	4
AVG	5
RANGE	6
DEV	6
RANGE	7
DEV	7
AVG	8
AVG	9
AVG	12

SUBUR	MARS
AVG	1
RANGE	4
DEV	8
AVG	9

SUBUR	SWAM
AVG	2
RANGE	2
DEV	2
RANGE	3
DEV	3
AVG	3
AVG	4
AVG	6
AVG	7
RANGE	8
DEV	8
AVG	8
AVG	9
AVG	12

SUBUR	RAILY
AVG	4
AVG	5
RANGE	6
DEV	6
RANGE	7
DEV	7
AVG	8
AVG	9
AVG	12

LAK	SWAM
AVG	1
RANGE	2
DEV	2
RANGE	3
DEV	3
RANGE	4
DEV	4
AVG	4
AVG	5
RANGE	6
DEV	6
RANGE	8
DEV	8
AVG	8
AVG	9
AVG	12

LAK	RAILY
AVG	1
RANGE	2
AVG	3
RANGE	3
RANGE	4
DEV	4
AVG	4
AVG	5
DEV	6
RANGE	8
DEV	8
AVG	8
AVG	9
AVG	12

MARS	SWAM
RANGE	2
DEV	2
RANGE	3
DEV	3
AVG	4
AVG	6
AVG	8

MARS	RAILY
RANGE	4
DEV	4

SWAM	RAILY
AVG	2
AVG	3
AVG	4
RANGE	4
DEV	4
RANGE	5
DEV	5
AVG	6
AVG	7
AVG	8
RANGE	8
DEV	8

FEATURE	DISTANCE 1	3	9
ASM	1	5	9
ASMD	2	6	10
COR	3	7	11
ASMID	4	8	12

KEY TO VARIABLE NUMBERS

Figure 6a tabulates for each category pair which of the 36 feature variables can separate the category pair.

Figure 6b is a continuation of Figure 6a and tabulates for each category pair which of the 36 feature variables can separate the category pair.

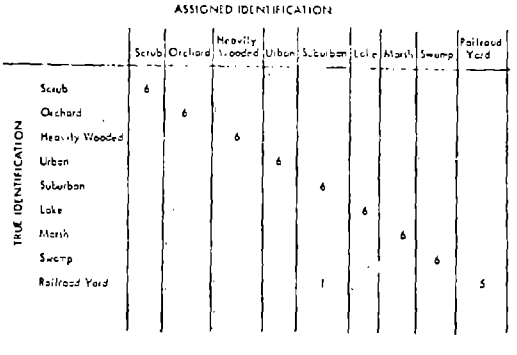


Figure 7 shows contingency table of true identification when statistics are gathered from the full 54 samples and the assignments are made on all the 54 samples.

ASSIGNED IDENTIFICATION

TRUE IDENTIFICATION	Scrub	Orchard	Heavily Wooded	Urban	Suburban	Lake	Marsh	Swamp	Railroad Yard
Scrub	4	1	1						
Orchard		4		1			1		
Heavily Wooded	1		4			1			
Urban				3	2		1		
Suburban				1	3		1		1
Lake			1			5			
Marsh				1		1	4		
Swamp								6	
Railroad Yard									5

Figure 8 shows contingency table of true identification versus assigned identification when the following experiment is repeated 54 times: statistics are gathered from 53 samples and an assignment is made on the 54th sample.

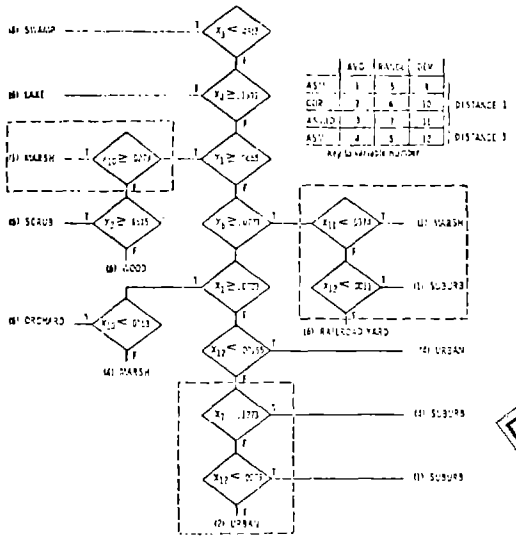


Figure 9 diagrams the Optimal Dichotomous Key (Sequential decision rule).

ASSIGNED IDENTIFICATION

TRUE IDENTIFICATION	Scrub	Orchard	Heavily Wooded	Urban	Suburban	Lake	Marsh	Swamp	Railroad Yard
Scrub	6								
Orchard		6							
Heavily Wooded			6						
Urban				4	2				
Suburban					5				1
Lake						6			
Marsh			1				4		1
Swamp								6	
Railroad Yard									6

Figure 10 shows contingency table of true identification versus assigned identification for the dichotomous key of 8 decision points. The total probability of correct identification is $\frac{49}{54}$ or 91%.

APPENDIX III

by

Robert M. Haralick
James D. Young
Dinesh K. Goel
K. Sam Shanmugam
Center for Research, Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas

A COMPARATIVE STUDY OF DATA COMPRESSION TECHNIQUES FOR DIGITAL IMAGE TRANSMISSION

ABSTRACT

We investigate three methods of data compression for aerial image data: (1) Differential Pulse Code Modulation, (2) Principal Components, and (3) Hadamard Transform. We compare these methods in terms of data compression factor versus rms error. Our comparison indicates that the Principal Components method is uniformly better than the Hadamard Transform method. Furthermore, for compression ratios greater than 5, the Hadamard Transform and Principal Component technques are better than the Differential Pulse Code Modulation. It is only for scenes which are relatively unstructured and compressed at compression ratios of less than 5 that Differential Pulse Code Modulation performs better.

INTRODUCTION

Since there is a high positive correlation between the grey levels of spatially adjacent resolution cells on aerial imagery, the imagery contains a large amount of redundant information. Hence, image data is a good candidate for data compression which would, for example, permit more images to be stored per roll of tape or permit more images to be transmitted per unit time over a given communication channel. And, in fact, several image data compression techniques have been suggested and are in use to eliminate many of the digital bits representing redundant information (see the special issue on redundancy reduction, Proceedings of the IEEE, Vol. 55, 1967; Arguello, 1971; Wilkins and Wintzs' bibliography on data compression, 1971; Claire, et al., 1971). In this paper, we investigate three methods of data compression:

(1) Differential Pulse Code Modulation
(2) Principal Components
(3) Hadamard Transforms

In reviewing some of the picture coding techniques, W. F. Schreiber (1967) had indicated that it is very difficult to compare the different data compression techniques because of the great disparity in the subject matter, methods of evaluation and reproduction of output pictures. The approach used in this study provides an unified method of comparing different data compression techniques since we have used the same images, approximately the same data compression factors, and the same method for evaluation and reproduction of reconstructed images.

DIFFERENTIAL PULSE CODE MODULATION

The statistical relationship between nearby picture elements and the greater sensitivity of the eye to spatial and temporal grey tone differences to absolute grey tone values has led to suggestions for data compression using grey tone differences (Seyler, 1965). The most straight-forward technique is called Differential Pulse Code Modulation (DPCM). Here the data is compressed by transmitting the quantized difference between the correct grey tone at the transmitter and the last reconstructed spatially adjacent grey tone at the receiver (O'Neal, 1966). The data compression is accomplished in the quantizing step since the number of possible quantizer levels used to transmit the grey level differences between spatially adjacent resolution cells is smaller than the number of possible levels used to transmit the actual grey tones at each resolution cell. Because differential quantizing tends to preserve edge information, for any given number of bits per image element, it tends to produce better quality images than ordinary Pulse Code Modulation.

PRINCIPAL COMPONENTS

In the Principal Component method, the image is first split into a number of small mutually exclusive spatial regions and we shall consider the grey tones of these regions to be N-dimensional vectors sampled from source probability distribution. The image is then a collection of these vectors. In the principal component method, these N-dimensional vectors are projected onto some smaller K-dimensional subspace having maximal variance. In this way, the N components of the original vectors may be expressed in terms of K components, thereby achieving some data compression.

An optimal set of coordinates for the K-dimensional subspace is the set of K eigen vectors having largest eigenvalues of the covariance matrix of the sample of N-dimensional vectors. The principle on which this method is based, namely, the Karhunen-Loeve expansion, is well known (Watanabe, 1965). However, this technique has been used on a rather limited basis for image compression applications, even though it has been known to lead to very good comparison performances for analog data such as EKG (Andrews, et al., 1967) and multispectral scanner data (Ready, et al., 1971).

HADAMARD TRANSFORM

In the Hadamard Transform data compression technique, the image is split up into small spatial regions as in principal components. The lower sequencies of the Hadamard Transform of these regions are then transmitted. The image is reconstr-cted at the receiver using the same lower sequency functions. The method works

because the image data there is usually a high positive correlation between adjacent resolution cells and, therefore, the image tends to have a characteristic frequency spectrum with the low frequencies dominating the high frequencies. And although a Hadamard sequency is not the same as the frequency, their general behavior is often similar. Hence, in image data, the low sequencies tend to dominate the high sequencies. Data compression is achieved by use of only the few dominant sequency components. Pratt, et al., 1969, has used the Hadamard Transform for image data compression by transmitting the entire quantized Hadamard Transform of the image.

RESULTS

For comparison purposes a set of sixty-six digital images were processed using the three methods of data compression. These images were obtained by digitizing sections of aerial photographs containing a wide variety of scenes. Eleven scene categories, with six images for each type of scene, were processed. The scenes included both natural scenes such as wooded areas, lakes, and man-made scenes such as urban areas, suburban areas and railroad yards. The digitized images were of 64 x 64 size and the grey levels of individual cells had been quantized into 64 levels.

Using computer programs, the images were transmitted and the RMS errors between the original integer images and the corresponding reconstructed integer images were computed. For each type of scene, the average RMS error was calculated by averaging the rms errors of the six images of the scene. The original and reconstructed images were digitally printed out using 13 grey levels. These digital printouts provide the basis for visual comparison of the original and reconstructed images.

A comparison in terms of data compression factor versus rms error between the original image and the reconstructed compressed image indicate that Principal Components is uniformly better than Hadamard Transforms. Furthermore, for compression ratios greater than 5, Principal Components and Hadamard Transforms are better than Differential Pulse Code Modulation. It is only for relatively unstructured scenes compressed at compression ratios less than 5 that the Differential Pulse Code Modulation performs better. Plots of rms error vs data compression factor for four scene categories are shown in Figure 1.a-d.

Visual comparison of the images compressed by the three methods tend to support the following conclusions:

(1) Of the images compressed by the three procedures, the images compressed by the principal components procedure most resemble the original images.

(2) The images compressed by the Hadamard transform procedure are comparable to the images produced by the principal component procedure. However, these images have a "checkerboard" look.

(3) DPCM procedure tends to "blurr" the boundary lines in the images. At high data compression factors, the images compressed by the DPCM procedure bore very poor resemblance to the original.

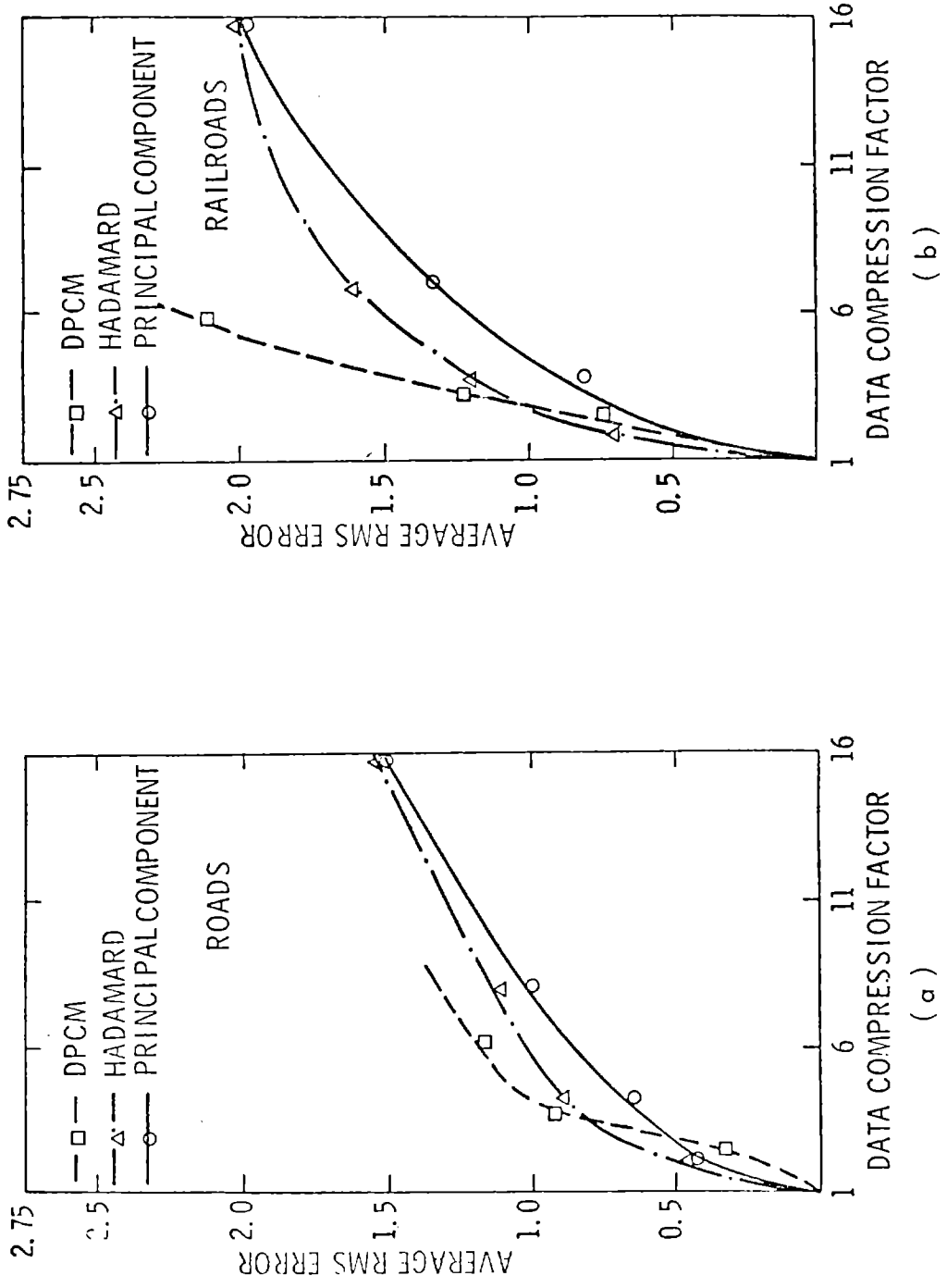


Figure 1. - Data Compression Factor vs. RMS Error

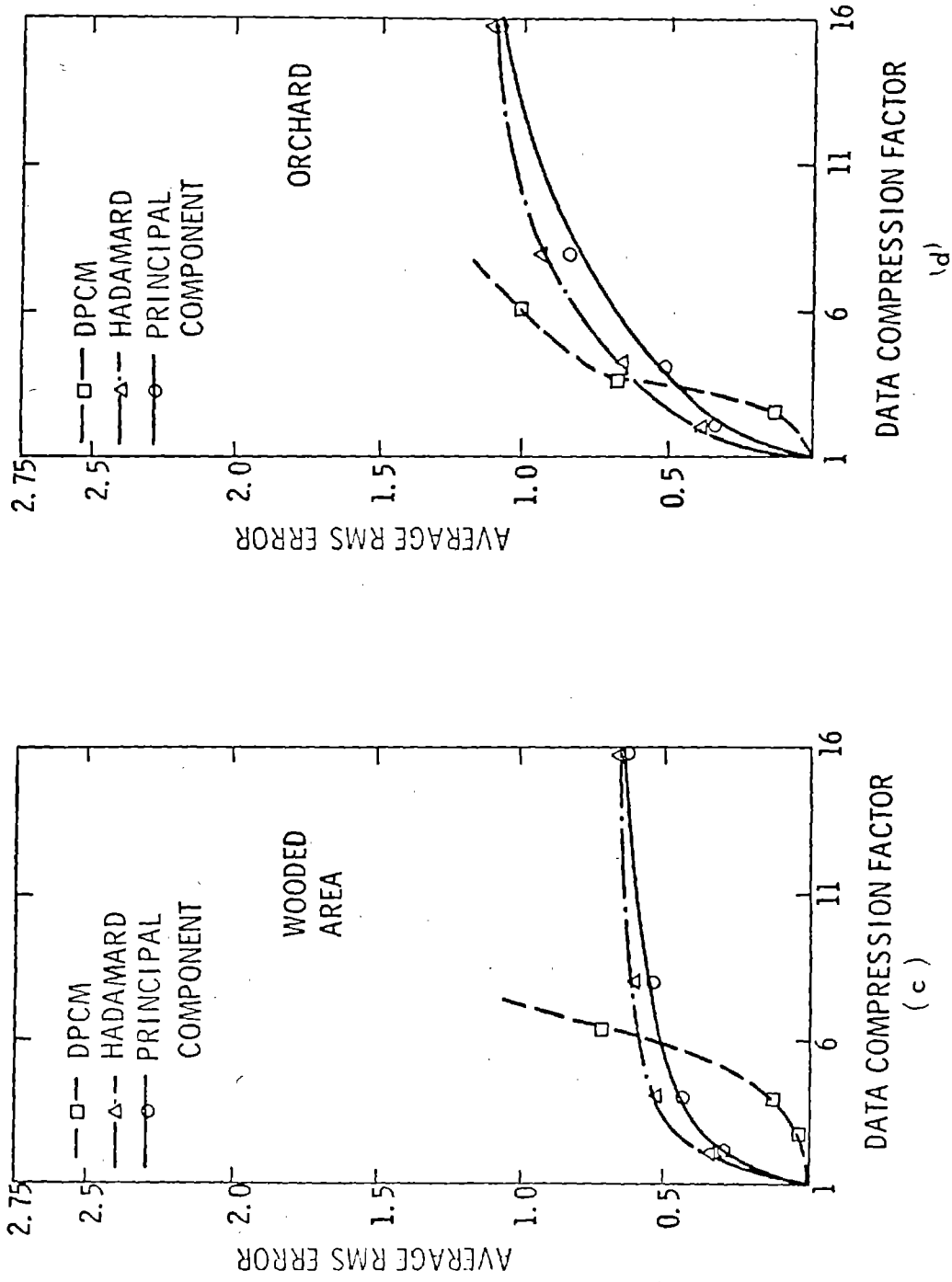


Figure 1. -Data Compression Factor vs. RMS Error

REFERENCES

1. Roger J. Arguello, "Encoding, Transmission and Decoding of Sampled Images," Proceedings of the Perkins-Elmer Symposium on Encoding Transmission and Decoding of Sampled Images, Perkins-Elmer Co., 1971.

2. Proceedings of the IEEE, Special Issue on Redundancy Reduction, vol. 55, pp. 253-401, March 1967.

3. A. J. Seyler, "Statistics of Television Frame Differences," Proceedings of the IEEE, vol. 53, pp. 2127-2128, December 1965.

4. J. B. O'Neal, "Predictive Quantizing for Transmission of TV Signals," Bell Sys. Tech. J., vol. 45, pp. 689-722, October 1966.

5. S. Watnabe, "Karhunen-Loeve Expansion and Factor Analysis," Tran. 4th Prague Conf. on Information Theory, 1965.

6. C. A. Andrews, J. M. Davies and G. R. Schwarz, "Adaptive Data Compression," Proceedings of the IEEE, vol. 55, pp. 267-277, March 1967.

7. William K. Pratt, Julius Kane, Harry C. Andrews, "Hadamard Transform Image Coding," Proceedings of the IEEE, vol. 57, pp. 58-68, January 1969.

8. J. B. O'Neal, "Data Modulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and Television Signals," Bell Sys. Tech. Journal, January 1966.

9. H. L. Vantries, "Detection, Estimation and Modulation," John Weily, New York, 1968.

10. K. W. Cattermole, "Principles of Pulse Code Modulation," American Elsevier Pub. Comp., New York, 1969.

11. R. M. Haralick and D. Anderson, "Texture Tone Study with Applications to Digitized Images," Technical Report No. 182-2, 1972, University of Kansas Center for Research, Inc., Lawrence, Kansas 66044.

12. W. F. Schreiber, "Picture Coding," Proceedings of the IEEE, vol. 55, pp. 320-330, March 1967.

13. H. C. Andrews, "Computer Techniques in Image Processing," Academic Press, 1970, New York.

14. L. C. Wilkins and Paul A. Wintz, "Bibliography on Data Compression, Picture Properties and Picture Coding," IEEE Trans. Inf. Theory, vol. IT-17, pp. 180-197, March 1971.

15. E. J. Claire, S. M. Forber, and R. R. Green, "Practical Techniques for Transform Data Compression/Image Coding," IEEE Trans. on Electromagnetic Compatibility, vol. EMC-13, pp. 1-5, Aug., 1971.

16. P. R. Ready, P. A. Wintz, S. J. Whitsitt and J. A. Landgrebe, "Effect of Compression and Random Noise on Multispectral Data, Proceedings of the Seventh International Symposium on Remote Sensing of Environment, pp. 1321-1342, University of Michigan, Ann Arbor, May, 1971.

17. D. A. Landgrebe, "The Development of Machine Technology Processing for Earth Resources Survey," Third Annual Earth Resources Program Review, NASA MSC Report # MSC 03742, December, 1970.