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.

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 pragmatically 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.

*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.

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 $a=0^\circ$), 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, $a=45^\circ$) 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 logP(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);

$$\frac{\text{COV}}{\text{VAR}} = \frac{\displaystyle\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} ab\,\frac{P(a,b)}{{}^{\#}R} - \left[\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} a\,\frac{P(a,b)}{{}^{\#}R}\right]\left[\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} b\,\frac{P(a,b)}{{}^{\#}R}\right]}{\left[\displaystyle\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} a^2\,\frac{P(a,b)}{{}^{\#}R} - \left(\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} a\,\frac{P(a,b)}{{}^{\#}R}\right)\right]^{1/2}\left[\displaystyle\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} b^2\,\frac{P(a,b)}{{}^{\#}R} - \left(\sum_{a=1}^{N_g}\sum_{b=1}^{N_g} b\,\frac{P(a,b)}{{}^{\#}R}\right)\right]^{1/2}}$$

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

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

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

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

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

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

$$\sum_{n=1}^{12} \min\{|x_n - a_{nk}|, |x_n - b_{nk}|\} \, (a_{nk} - b_{nk}) \geq \sum_{n=1}^{12} \min\{|x_n - a_{nj}|, |x_n - b_{nj}|\} \, (a_{nj} - b_{nj}), \quad j=1,2,\ldots,K.$$

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.

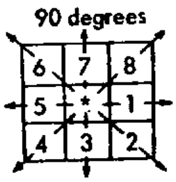

Figure 1. Resolution cells nos. 1 and 5 are the 0-degree (horizontal) nearest neighbors to resolution cell '*', resolution cells nos. 2 and 6 are the 135-degree nearest neighbors, resolution cells 3 and 7 are the 90-degree nearest nearest neighbors, and resolution cells 4 and 8 are the 45-degree nearest 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)

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

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

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

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

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

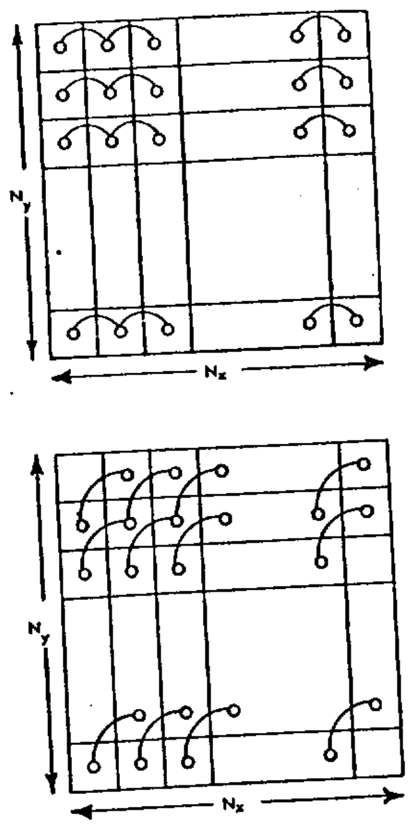

Figure 4 illustrates how the members of horizontal and right diagonal neighboring resolution cells are counted.

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

$$ASMD(90^\circ) = \frac{1}{24}(6\cdot 0 + 0\cdot 1 + 2\cdot 4 + 0\cdot 9 + 0\cdot 1 + 4\cdot 0 + 2\cdot 1 + 0\cdot 4 + 2\cdot 4 + 2\cdot 1 + 2\cdot 0 + 2\cdot 2 + 0\cdot 9 + 0\cdot 4 + 2\cdot 1 + 0\cdot 0) = 1.083$$

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

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

where
$$COV(45^\circ) = \frac{1}{18}(0\cdot 0\cdot 4 + 0\cdot 1\cdot 1 + 0\cdot 2\cdot 0 + 0\cdot 3\cdot 0 + 1\cdot 0\cdot 1 + 1\cdot 1\cdot 2 + 1\cdot 2\cdot 2 + 1\cdot 3\cdot 0 + 2\cdot 0\cdot 0 + 2\cdot 1\cdot 2 + 2\cdot 2\cdot 4 + 2\cdot 3\cdot 1 + 3\cdot 0\cdot 0 + 3\cdot 1\cdot 0 + 3\cdot 2\cdot 1 + 3\cdot 3\cdot 0) - \frac{1}{(18)^2}(0\cdot 5 + 1\cdot 5 + 2\cdot 7 + 3\cdot 1)^2$$

$$VAR(45^\circ) = \frac{1}{18}(0^2\cdot 5 + 1^2\cdot 5 + 2^2\cdot 7 + 3^2\cdot 1) - \frac{1}{(18)^2}(0\cdot 5 + 1\cdot 5 + 2\cdot 7 + 3\cdot 1)^2$$

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

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

```
H WOO RAILY
AVG      1
RANGE    2
DEV      2
RANGE    3
DEV      3
RANGE    4
DEV      4
AVG      4
AVG      5
RANGE    8
DEV      8
AVG      8
AVG      9
URBA SUBUR
AVG     10
URBA   LAK
AVG      1
AVG      2
AVG      3
AVG      4
RANGE    4
DEV      4
AVG      5
RANGE    6
DEV      6
AVG      7
DEV      7
AVG      8
DEV      8
AVG      9
RANGE    9
DEV      9
AVG     12
URBA  MARS
AVG      1
AVG      9
URBA  SWAM
AVG      2
RANGE    2
DEV      2
RANGE    3
DEV      3
AVG      3
AVG      4
AVG      6
AVG      7
AVG      8
DEV      8
AVG      9
AVG     10
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
RANGE    8
DEV      8
AVG      9
DEV      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      9
SUBUR RAILY
LAK    MARS
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    5
DEV      5
AVG      6
RANGE    6
DEV      6
RANGE    7
DEV      7
AVG      7
AVG      8
AVG      9
AVG     12
LAK   RAILY
AVG      1
RANGE    2
AVG      3
RANGE    3
RANGE    4
DEV      4
AVG      4
AVG      5
RANGE    6
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
```

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

KEY TO VARIABLE NUMBERS

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.

ASSIGNED IDENTIFICATION

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

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.

TRUE IDENTIFICATION \ ASSIGNED 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.

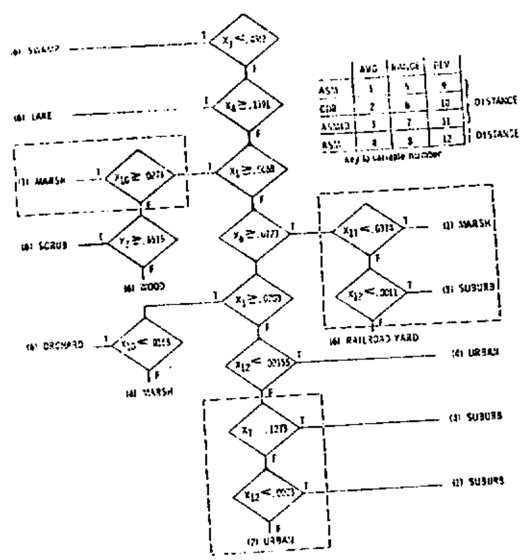

Figure 9 diagrams the Optimal Dichotomous Key (Sequential decision rule).

TRUE IDENTIFICATION \ ASSIGNED 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 93%.