Paper I 16

COMBINED SPECTRAL AND SPATIAL PROCESSING OF ERTS IMAGERY DATA

Robert M. Haralick and K. Sam Shanmugam, *University of Kansas Center for Research, Inc., Space Technology Laboratories, Lawrence, Kansas*

ABSTRACT

A procedure for extracting a set of textural features for ERTS MSS data is presented. The textural features were combined with a set of spectral features and were used to develop a classification algorithm for identifying the land use categories of blocks of digital MSS data. The classification algorithm was derived from a training set of 314 blocks and tested on a set of 310 blocks. The overall accuracy of the classifier was found to be 83.5% on seven land use categories.

1. INTRODUCTION

Spectral, textural and context features are three fundamental pattern elements used in human interpretation of imagery data. Spectral features describe the average band to band tonal variations in a multiband image set where as textural features contain information about the spatial distribution of tonal values within a band. Context features contain information derived from areas surrounding the sub-image region being analyzed. When small image areas are independently processed by a machine, only the textural and spectral features are available to the machine.

In much of the automated procedures for processing image data from small areas, such as in crop classification studies, only the spectral features are used for developing a classification algorithm. Other than simple averaging of spectral values within an image area to eliminate irregularities, textural features are generally ignored. Because the areal characteristics of texture carry so much information it is important to use the textural features in automated image processing schemes, except in applications where the poor resolution of the imagery does not provide meaningful textural information.

Other than some work with Fourier, Hadamard transforms (1), (2) and auto-correlation function, there exists little or no theory to aid in establishing what the textural features should consist of. We have developed a set of quickly computable textural features for automatic analysis of MSS and other imagery. We used these features to perform a land use classification of a 5100 square mile area covered by part of an ERTS-A, MSS band 5 image consisting of over 600 blocks of size 64 x 64 resolution cells. The land use categories identified consisted of coastal forest, wood lands, annual grass lands, water bodies, urban areas, small and large irrigated fields. Up to 70% of the image blocks were identified correctly based on the textural features, compared to an accuracy of 74 to 77% for a scheme based on the spectral characteristics of the 4 band MSS image set. When the combined textural and spectral features were used as inputs to the classifier, up to 83.5% of the image blocks were identified correctly.

2. TEXTURAL FEATURES FOR ERTS-MSS DATA

2.1 Spatial Grey Tone Dependence Matrix

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 $G = \{1,2,\ldots,N_g\}$ 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 matrix, or more precisely, four closely related matrices from which all texture-context features are derived. These matrices are termed angular nearest neighbor grey tone spatial dependence matrices.

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

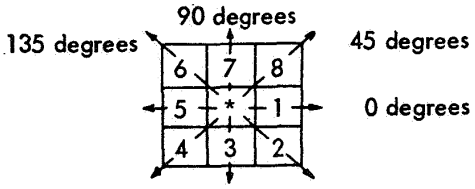


Figure 1. Resolution cells nos. 1 and 5 are the 0-degree (horizontal) nearest neighbors to resolution cell '*', resolution cells nos. 2 and 6 are the 135-degree nearest neighbors, resolution cells 3 and 7 are the 90-degree nearest neighbors, and resolution cells 4 and 8 are the 45-degree nearest neighbors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

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 (see Figure 1). 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. Formally, for angles quantized to 45° intervals the unnormalized frequencies are defined by:

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

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

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

$$P(i,j,d,135^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k-m=d,\ l-n=d) \text{ or } (k-m=-d,\ l-n=-d),\ I(k,l)=i,\ I(m,n)=j\} \quad (1)$$

These spatial grey tone dependence matrices can be normalized by dividing each entry in the matrix by the total number of resolution cell pairs used in computing the entries. Details of the procedure used in computing the entries in the matrices may be found in (3).

2.2 Textural Features

From each of the grey tone dependency matrices we extract a set of 32 textural features. The equations which define these features are given in (4). For illustrative purposes we define three of these features:

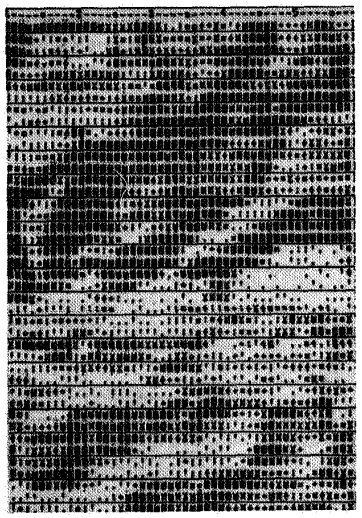
$$f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(\frac{P(i,j)}{\#R}\right)^2 ; \quad f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{\sum_{|i-j|=n}\left(\frac{P(i,j)}{\#R}\right)\right\} ; \quad f_3 = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{ij\,P(i,j)}{\#R} - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{2}$$

where, $\#R$ is the number of resolution cells pairs, and μ_x, μ_y and σ_x, σ_y are the means and standard deviations of the marginal distribution P_x and P_y obtained by summing the rows and colums of $P(i,j)/\#R$. To explain the significance of these features, let us consider the kind of values they take on two different land use category images. Figure 2 shows the digital printout of two sub-images from the California frame. The image shown in 2(a) belongs to the grass land category and image shown in Figure 4(b) is mostly water. Values of the features f_1, f_2, and f_3 are also shown for these images in Figure 2.

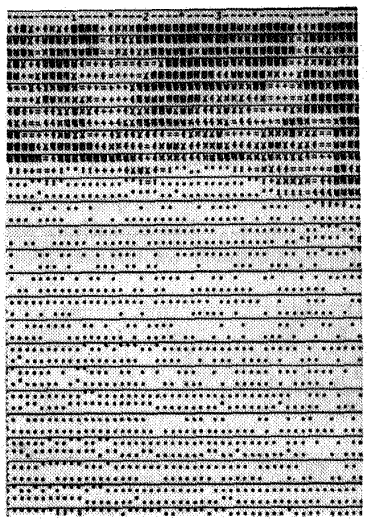
The angular second moment features (ASM), f_1, is a measure of homogeneity of the image. In a homogeneous image, such as shown in 2(b), there are very few dominant grey tone transitions. Hence, the P matrix for this image will have fewer entries of large magnitude. For an image like the one shown in Figure 2(a), the P matrix will have a large number of small entries and hence the ASM feature which is the sum of squares of the entries in the P matrix will be smaller. A comparison of the ASM values given below the images in Figure 2 shows the usefulness of the ASM feature as a measure of the homogeneity of the image.

The contrast feature f_2 is a difference moment of the P matrix and is a measure of the contrast or the amount of boundaries present in an image. Since there is a large amount of boundaries present in the image 2(a) compared to the image shown in 2(b), the contrast feature for the grassland image has consistently higher values compared to the water body image.

The correlation feature, f_3, is a measure of linear grey tone dependencies in the image. For both the images shown in Figure 2, the correlation features is somewhat higher in the horizontal (0°) direction, along the line of scan. The water body image consists mostly of a constant grey tone value for the water plus some additive noise. Since the noise samples are mostly uncorrelated, the correlation features for the water body image have lower values compared to the grassland image. Also the grassland image has a considerable amount of linear structure along 45° lines across the image and hence the value of the correlation feature is higher along this direction compared to the values for 90° and 135° directions.



a. Grassland



b. Water Body

Angle	ASM	Contrast	Correlation	ASM	Contrast	Correlation
0°	.0128	3.048	.8075	.1016	2.153	.7254
45°	.0080	4.011	.6366	.0771	3.057	.4768
90°	.0077	4.014	.5987	.0762	3.113	.4646
135°	.0064	4.709	.4610	.0741	3.129	.4650
Avg.	.0087	3.945	.6259	.0822	2.863	.5327

Figure 2. Textural Features for Two Different Land Use Category Images.

The various features which we suggest are all functions of distance and angle. The angular dependencies present a special problem. Suppose image A has features a, b, c, d for angles 0°, 45°, 90°, and 135° and image B is identical to A except that B is rotated 90° with respect to A. Then B will have features c, d, a, b for angles 0°, 45°, 90°, and 135° respectively. Since the texture context of A is the same as the texture context of B, any decision rule using the angular features a, b, c, d must produce the same results for c, d, a, b or for that matter b, c, d, a (45° rotation) and d, a, b, c (135° rotation). To guarantee this, we do not use the angularly dependent features directly. Instead, we use two symmetric functions of a, b, c, d, their average, and their range. The textural features used in our study were computed for four angles and for a distance of one.

3. SPECTRAL FEATURES

The spectral features used in our study consisted of the mean and variance of the grey tone values within each image block. These features were computed for each

of the four MSS bands, thus yielding a set of 8 spectral features for each image block. The mean and variance of the spectral data have been used extensively in the past for crop classification studies based on MSS data.

4. LAND USE CLASSIFICATION STUDIES

4.1 Data Set

All of the data used in our present study were derived from parts of the ERTS-I image frame 1002-18134. The date of the flight was July 25, 1972, and the center coordinates of the frame were 37.291N, 120.935W. The area of coverage includes the San Francisco Urban area and the Monterray Bay on the coast line of California. Of the four image strips in this frame, all of strip one and one-half of strip three were digitally processed. The second MSS band (MSS-5) image and the area processed are shown in Figure 3.

As a first step in the digital processing, the image area processed was divided into a total of 648 sub-images of size 64 x 64 resolution cells (each covering a ground area of 8.5 square miles). For each of the sub-images, the digital spectral data from the second MSS band (MSS-5) were extracted. The imagery data for each sub-image was then normalized using an equal probability quantizing algorithm and the 32 textural features for each sub-image were then computed. The spectral features for each image block were computed using the digital data from all four sensor bands.

Ground truth for each of the sub-images were obtained with the help of photointerpreters working on the MSS images and the color composite image. A total of seven land use categories could be identified reasonably well on the image and for 624 of the 648 sub-image blocks we could determine a distinct land use category. The set of 7 land use categories used in our study consist of the following:

4.2 Categories of Land Use Detectable on ERTS Image (1002-18134)

(1) Coastal Forest: This area can be typified as a dense temperatue rainforest sited on the windward side of the Coast Ranges of California. It is composed of needle-leaf and broadleaf evergreen trees which generally form a continuous cover.

(2) Woodlands: These are areas to the leeward side of the Coast Ranges. At higher elevations, these woodlands consist of oaks, mostly evergreen, with varying but less continuous cover. At lower elevations these oak woods grade into chapparal which provides decreasing cover with decreasing elevation until it gives way to annual grasslands.

(3) Annual Grasslands: This is an area of non-native annual grasses which have already completed their life cycle by the end of July (turned brown). This area is primarily, like the preceding two categories, quite mountainous. This is also the natural vegetation which would be found in the Great Valley of California except for anthropogenic vegetation.

(4) Urban: Centers of human activity are extremely important as a landscape component within this frame, they are non-randomly located in a size hierarchy of central places connected by transport links. The primary urban agglomerations within

Figure 3. Two strips whose boundaries are shown on the top were processed from this MSS image set (1002-18134).

the frame are the San Jose-Oakland segment of San Francisco, and parts of the Sacramento-Stockton urban complex. Unlike the preceding categories, one of the main image features of urban areas is non-uniformity. Both pattern and texture, as well as a variety of tones, are associated with urban complexes.

(5) Large Irrigated Fields: For practical purposes all field agriculture in the Great Valley is based on irrigation. The areas assigned to this category would most

likely contain field crops such as cotton, alfalfa, and other crops readily adaptable to high mechanization. Most tree crops and vineyards would also be in this category.

(6) Small Irrigated Fields: These areas would contain high value low mechanization crops typified by vegetables.

(7) Water: Although almost self-explanatory as a term, the category in this instance is used to include: (a) ocean, (b) lakes — natural and man-made, and (c) standing water (in fields and on flood plains).

Out of the 648 sub-images in the frame, the photointerpreters helped us to find a unique land use ground truth category for a total of 624 sub-images. Due to cloud cover and other ambiguities the ground truth for 24 sub-images could not be positively identified.

4.3 Classification Algorithm

The problem of developing an algorithm for identifying the land use categories of sub-image blocks from an ERTS MSS image set can be stated as follows. A set of N measurement pairs $(X_1, \theta^1), (X_2, \theta^2), \ldots (X_N, \theta^N)$ are given as learning observations. A vector measurement (pattern) X_i, where the components of X_i are the spectral and textural features of a sub-image block, comes from an image block whose land use category θ^i is known. θ^i is one of the R land use categories (R=7, in our study) $c_1, c_2, \ldots c_R$. Based on the set of learning observations, we want to develop an algorithm for identifying the land use category of a sub-image block based on the measurement (pattern) vector X it produces.

In a widely used algorithm (Fukunaga (5), Fu and Mendel (6), Miesel (7)), the pattern space is partitioned into a number of regions using a set of hyperplanes (decision boundaries) whose locations are determined by the sample patterns. Each region is dominated by sample patterns of a particular category. When a new pattern is presented for identification, it is assigned a category depending on the region in which it belongs. If the new pattern X is located in a region dominated by sample patterns of category c_j, then X is classified as coming from category c_j.

For the multicategory problem involving N_R categories, a total of $N_R(N_R-1)/2$ hyperplanes are used to partition the pattern space. These hyperplanes are defined by a set of weight vectors W_{ij} $(i = 1,2,\ldots,N_R, j = 1,2,\ldots,N_R; j > i)$ which separates the sample patterns belonging to the i^{th} and j^{th} categories. A regression type algorithm given in Fukunaga (Chapter 4) was used to obtain the weight vectors. After the location of the hyperplanes are determined, the classification of new patterns is done as follows. For each category c_i, the number of hyperplanes, V_i, which give a positive response when the new pattern X is presented are determined using

$$V_i = \sum_{\substack{j=1 \\ j \neq i}}^{N_R} \frac{|W_{ij}^T Z| + W_{ij}^T Z;\ i = 1,2,\ldots,N_R}{2\,|W_{ij}^T Z|}$$

where Z is the augmented pattern vector obtained by adding a component of value 1 to X, i.e.,

$$Z = \begin{bmatrix} 1 \\ X \end{bmatrix} .$$

X is assigned to category c_i if $V_i = \max \{V_i\}$. If there is a tie between categories c_m and c_n, then X is assigned to c_i if $W^T_{mn} Z < 0$ or to c_n if $W^T_{mn} Z \geq 0$. Several modifications of the linear discriminant function method and a multitude of other classification procedures may be found in the references cited.

4.4 Results of Land Use Classification Experiments

The textural and spectral feature vectors for each of the 624 sub-images were divided arbitrarily into training and test sets. The classification algorithm was developed using the information contained in the training set and the samples in the test set were assigned to one of the seven possible land use categories. The accuracy of classification was obtained by comparing the category assigned by the classifier with the known ground truth category and the results are presented in the following paragraphs.

Experiment No. 1: In this experiment, a land use classification scheme was developed using only the textural features of the sub-images on MSS-5 band. Fifty per cent of the samples were arbitrarily selected and used for developing the classification algorithm and the algorithm was tested on the remaining samples. There were seven land use categories and the overall identification accuracy of the classifier on the test samples was 67.5%. With 70% of the samples used for training, the accuracy of the classifier was 70.5%.

Experiment No. 2: In this experiment the classification algorithm was developed and tested using only the spectral features. With 30% of the samples in the test set, the accuracy of the classifier was 74% on the test set and with 50% of the samples in the test set, the accuracy of the classifier was 77%.

Experiment No. 3: Eight textural features were combined with eight spectral features and the algorithm developed and tested using the combined features. The overall accuracy of the land use classification algorithm was 83.5% for test set sizes of 30% and 50%. The contingency table for the classification is shown in Table I.

5. DISCUSSION

The results of our study clearly demonstrates the need for using both the spectral and spatial (textural) characteristics of ERTS MSS data for developing classification procedures. The 83.5% accuracy we have achieved on automatic land use classification procedure is less than the accuracy of human interpreters. The difference in accuracy between the automated procedure and human interpreters can be attributed to the following factors:

TABLE I
CONTINGENCY TABLE FOR LAND USE CLASSIFICATION OF TEST SAMPLES

Assigned Category / True Category	Coastal Forest	Woodlands	Annual Grass-Lands	Urban Area	Large Irrigated Fields	Small Irrigated Fields	Water	Total
Coastal Forest	23	1	2	0	0	0	1	27
Woodlands	0	17	10	0	1	0	0	28
Annual Grasslands	1	3	109	1	1	0	0	115
Urban Area	0	3	10	13	0	0	0	26
Large Irrigated Fields	1	2	6	0	37	2	0	48
Small Irrigated Fields	0	0	4	0	3	24	0	31
Water	0	0	0	0	0	0	35	35
TOTAL	25	26	141	14	42	26	36	310

Number of Training Samples = 34 Number of Test Samples = 310
Accuracy of Classification on Training Set = 84.0%
Accuracy of Classification on Test Set = 83.5%

(1) The context information in the image, on which image interpreters rely heavily, was not taken into account by the automated procedure.

(2) The automated classification algorithm assumed unimodal distribution for the feature vectors from each category. This assumption is not valid for at least some of the categories, such as urban land use category, which consists of many sub-categories and hence a multimodal distribution.

(3) A loss in accuracy is inevitable when the training of the classifier is done using samples from one area of the image and testing on a different area due to environmental changes such as different soil and atmospheric conditions.

We believe that if some of the above deficiencies are taken care of, (we are developing procedures to take care of some of them) the accuracy of automated processing of ERTS MSS data will be quite comparable to the accuracy of human image interpreters.

REFERENCES

1. Rosenfeld, A. and E. Troy, "Visual Textural Analysis," University of Maryland Computer Science Center, TR-70-116, June, 1970.

2. Kaizer, Herbert, "A Quantification of Textures on Aerial Photographs," Boston University Research Labs, Technical Note 121, 1955, AD69484.

3. Haralick, R. M. et al., "Texture-Tone Study with Application to Digitzed Imagery," Technical Reports 182-2, 182-3, U. S. Army Engineer Topographic Laboratories, Contract DAAK02-70-C-0388, University of Kansas Center for Research, Inc., 1971, 1972.

4. Shanmugam, K., R. M. Haralick and R. Bosley, "Land Use Classification Using Texture Information in ERTS MSS Imagery," Technical Report No. 2262-1, University of Kansas Center for Research, Inc., January, 1973.

5. Fu, K. S. and J. M. Mendel, "Adaptive Learning and Pattern Recognition Systems," Academic Press, New York, 1972.

6. Miesel, W., "Computer Oriented Approaches to Pattern Recognition," Academic Press, New York, 1972.

7. Fukunaga, K., "Introduction to Statistical Pattern Recognition," Academic Press, New York, 1972.