1-1-1976

The Use of Spatial Characteristics for the Improvement of Multispectral Classification of Remotely Sensed Data

D. J. Wiersma

D. Landgrebe

Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

THE USE OF SPATIAL CHARACTERISTICS FOR THE IMPROVEMENT OF MULTISPECTRAL CLASSIFICATION OF REMOTELY SENSED DATA

D. J. Wiersma and D. Landgrebe

The authors are affiliated with the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
The research reported in this paper was supported in part by NASA Contract NAS9-14016.

ABSTRACT

Two parallel and overlapping approaches to classification of remotely sensed data with the aid of spatial information are underway at the present time. The image processing approach attempts to model after the human visual system, while the second approach is primarily numerical. The technique of texture features[1,2], representing the image processing approach, and the sample classifier ECHO[3,4], representing the numerical approach are compared. The numerical approach is demonstrated to be superior in classification accuracy as well as being more efficient computationally.

INTRODUCTION

There is much work underway at the present time to learn how to use a machine to analyze various types of image data. This work is proceeding along two parallel and somewhat overlapping lines. In one, techniques are being developed which, generally, attempt to model the manner in which human intelligence addresses the problem.[7] The term image processing is often used in connection with this work, and an example effort is that directed towards developing suitable image texture measures.

The second branch is less associated with or modelled after the manner of human intelligence. The intent here has been to concentrate upon the unique capabilities of the machine for processing purposes. We will refer to this class of studies as the numerical approach, and an example effort of this type is the attempt to develop so-called sample classifiers.[9]

Relative to the analysis of scenes of the surface of the earth both approaches have been shown to improve classification accuracy by a significant amount. In the case of image processing techniques (e.g. the development of textural features) a distinct advantage is the ability to use perceptual concepts such as smoothness, contrast, and linearity to aid in the analysis. Numerical techniques on the other hand, often can be demonstrated to be significantly faster computationally. Although both approaches have been under development for some years, direct comparisons between them have not appeared in the literature. This is the subject of this paper.

The extraction of information from scene data is fundamental to the image processing and the numerical approach. In considering the problem of analyzing the energy distribution emanating from a scene in order to obtain information about the scene, one perceives immediately that the information-bearing aspects may exist in the spectral, spatial, or temporal variation of this energy distribution. From an image processing vantage point it is clearly important to understand the spatial variations; as humans we know that we can discriminate between two objects which have the same color but a different textural quality, for example. Concepts such as smoothness, coarseness, size, shape and contrast are often used to describe objects as we perceive them and to differentiate them from other objects. From an information theory point of view, one would suspect that the fact that the groups of resolution cells from the same class are next to each other would provide information which would be useful in identifying that class of objects. These latter two statements form the basis for the two approaches to machine analysis of image data. The critical task in the former is the assignment of algorithmic procedures to these concepts (texture, size, shape, contrast, etc.) such that a digital computer is able to process them. In the numerical approach, at this stage of development we rely primarily on the fact that the group of resolution cells which comprise an object in the scene have some distinguishing (multivariant) distribution which would be recognizable by machine.

The purpose of the present study is to examine these two approaches and obtain a comparison between them in order to increase our understanding of their potential and probable utility. The numerical techniques selected as an example here is the so-called Extraction and Classification of Homogeneous Objects[3,4] (ECHO). The image processing technique used is that of textural features developed by Haralick et al.[1,2]. Haralick has successfully used textural features calculated from LANDSAT satellite data and improved classification accuracy from 74-77% using only the spectral information, to 84% with the addition of textural features. We will next review these two procedures briefly.

TEXTURE FEATURES

The textural information in an image is contained in the spatial relationship between the grey-tones. Thus, Haralick defined a set of matrices termed grey-tone spatial-dependence matrices and evaluated them for each block or subimage. The elements P_{ij} of the matrices are the frequencies with which a picture element with grey-tone i is a distance 1 from a picture element with grey-tone j in a block. If the blocks are quantized into n grey-tones where n is an image the matrices are n x n. The size of the block will be a factor in the computation of the matrices. Also, the spatial information may be different in each of the spectral channels that are available, therefore, the three variables which have a significant effect on the make-up of the grey-tone spatial-dependence matrices are the block size, the number of quantization levels and the spectral band from which the features are computed.

The texture features themselves are computed from the grey-tone spatial-dependence matrices. Haralick, et al.[1,2], proposed a set of 28 texture features. Four of these were selected for our current study on the basis of their computation efficiency and their similarity to perceptual concepts. The features are angular second moment, contrast, correlation, and entropy. Angular second moment is a measure of the homogeneity of the image. The expression for calculating angular second moment is:

$$ASM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (P_{ij}/R)^2 \qquad (1)$$

where N_g is the number of grey-tones into which the image was quantized; P_{ij} is an element of the grey-tone spatial-dependence matrix and R is the number of possible neighboring cells.

Contrast is a measure of the degree of local variations in the image and is given by:

$$CONT = \sum_{n=o}^{N_g-1} n^2 \left(\sum_{|i-j|=n} P_{ij}/R \right) \qquad (2)$$

High contrast in the image implies a large difference in grey-tone values for neighboring picture elements.

Correlation is a measure of the grey-tone linear dependencies in the image. An expression of the same form as a correlation coefficient is used.

$$CORR = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij(P_{ij}/R) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad (3)$$

The quantities μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of the marginal distributions associated with the elements of the grey-tone spatial-dependence matrix.

Entropy is a measure of the variability of the image and has the logarithmic form:

$$ENT = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}/R \ \log(P_{ij}/R) \qquad (4)$$

These four features are thus calculated from the grey-tone spatial-dependence matrices and may be used to aid classification of multispectral data.

The interpretation of the texture features can be illustrated by examining the means of the features evaluated over an agricultural area as compared to those over water. See Table 1.

TABLE 1. Means of texture features taken from the data to be discussed in Section 4. The features were calculated from spectral band 4 with a block size of 15 x 15 pixels.

	AG	WATER
ASM	.0494	.7641
Contrast	2.6692	.9481
Correlation	.5140	.5458
Entropy	1.4584	.3270

Water which appears very uniform in the scene has a much higher value of angular second moment than does ag, which at the spatial resolution used in this study has more of a "patchwork" pattern, i.e. more texture. The entropy value for ag on the other hand is higher than for water since there is more variability in the agricultural areas of the image than over the more uniform water. Consequently, ASM and ENT are negatively correlated with a correlation coefficient on the order of -0.9. There is considerable contrast in the ag areas due to the bright and dark fields being near each other. This contrast shows up in the higher value of CONT for the ag class. The CORR feature can be used to indicate the amount of linear structure in an image. For example, a row crop will have a higher correlation than water which has very little correlation. However, for the data set used in this study there was not a significant difference in linear structure between classes to be able to use correlation effectively. The random arrangement of the ag fields showed slightly less linear structure than water.

THE ECHO CLASSIFIER

A different approach to the use of spatial information to improve scene analysis arises as follows. When thought of from the standpoint of a pixel-by-pixel multispectral classifier, there must be unused information in the fact that two spatially adjacent pixels are more likely to be from the same class than two widely separated ones. Furthermore, if instead of being forced to make a decision as to class membership on each pixel individually, one could accumulate n pixels for which there is high probability that they are members of a common class, it is more likely the correct class can be selected from the list of candidates based on the n pixels than would be possible based on one. Thus, one is led to the idea of partitioning the scene into "objects" or groups of pixels which are members of a common class, then classifying each object based on the distribution of the pixels of which it is composed. Thus, the name Extraction and Classification of Homogeneous Objects (ECHO).

Note that the objects need to be composed of a "homogeneous" set of pixels, not necessarily a uniform one, i.e. the variance within an object can be large or small, so long as it is consistant. There have been several sample (or object) classifier approaches used. Minimum statistical distance methods[9] have been thoroughly tested and such a classifier has been implemented in LARSYS for a number of years. The one used in this study is a maximum likelihood sample classifier, however. It has the advantage that as n declines to one, an object made up of a single pixel, the sample classifier becomes a conventional pixel classifier.

Both boundary seeking and object seeking scene partitioning algorithms have also been investigated. Within the object seeking approach, both conjunctive and disjunctive algorithms have been tested. The partitioning algorithm in ECHO is of the conjunctive object seeking type.[3,4]

EXPERIMENT

Multispectral data taken by the LANDSAT satellite on June 9, 1973 was used to compare the spatial analysis techniques. The scene used covered an area in Monroe County, Indiana and included the city of Bloomington and the Monroe Reservoir. Near the reservoir there is considerable forested area. Also, the terrain allows large agricultural areas. Four classes, Urban, Forest, Ag, and Water were selected. It was anticipated that the spatial variability of these four classes would be sufficiently different to provide additional information for classification.

There are two major steps to performing analysis and classification using texture features. First, the features must be computed from the data. The second step is to train the classifier and perform the classification. Only the latter step is required for ECHO and in this case the same training areas were to be used for ECHO as for the texture analysis.

Before calculating the texture features the channel from which the features were to be computed was chosen. Channel 2 (.6 to .7μm) and channel 4 (.8 to 1.1μm) were selected to compare the effect of channel selection on the results. The size of the block over which the features were computed was varied to test the effect of the block size on the results. Square blocks of seven, eleven, and fifteen picture elements on a side were used to compare the effects of the block size on the results. The image was re-quantized, using an equal-probability quantization algorithm, into eight grey levels. The texture features were computed for one block and assigned as four additional channels to the center picture element of the block. The block was then shifted over one element, and a new set of features was computed. Data for the four spectral channels plus the four textural channels was then stored on magnetic tape.

Classification was performed using LARSYS processing functions.[5] Training fields were selected by clustering several areas in the image to determine spectral classes. These classes were identified from aerial photographs. Care was taken to avoid isolated training points where the texture of the area surrounding the points would not be representative of the texture of the classes to which the points belong. A separate set of test fields was selected using topographical maps and aerial photography. There were 4464 pixels in this set as compared to 2179 in the training set.

Means and covariance matrices of the training data were computed by the STATISTICS processor[5] and then the CLASSIFYPOINTS processor[5] was used to classify the image pixels. This processor used a maximum likelihood Gaussian criterion. Using the test fields, the accuracy of the classification was computed. A variety of combinations of spectral and textural channels were used to compare and evaluate the texture features with one another and with ECHO.

RESULTS

A number of comparisons were conducted using various features and parameter combinations. Figure 1 shows several of these. The error rate obtained using the four spectral bands only is given first as a reference. The others all involve the use of some computed spatial information in addition to the spectral data, and in each case, as expected, the error rate for test pixels was decreased over the reference classification. The improvement was less in the case of the CORRelation feature and greatest with the ECHO procedure.

The subset of features labeled as optimum are a subset of five which were picked from the eight by computing the transformed divergence[8] for all class pairs and picking the subset whose average transformed divergence was maximum. The set thus selected was made up of the first two spectral bands plus ASM, CONT and ENT.

Figure 2 shows the training pixel error rate for the same feature sets as Figure 1. Results from the training pixels tends to indicate that the training set was a valid one for all feature subsets based on the consistency shown.

Figure 3 shows the results for the individual classes AG and URBAN. The divergence calculation indicated that these would be the most difficult to separate.

Figure 4 gives a performance comparison for different block sizes and where different spectral bands are used from which to derive the texture features. The four bands of the LANDSAT multispectral scanner have the following spectral ranges: 0.5-0.6, 0.6-0.7, 0.7-0.8, and 0.8-1.1 micrometers. Thus, the first two are in the visible region which the latter two are in the reflective infrared. It has been observed[6] that adjacent bands within these two regions tend to be correlated with one another, but that the correlation tends to be significantly less for bands which are in different regions (e.g. the second and third in this case). One band in each region was therefore thought to be adequate for this study and the second and fourth were arbitrarily selected. There does not appear from Figure 4 to be a strong dependency on this choice. There does appear to be some preference for the larger block sizes, however.

CONCLUSIONS

Results of a comparison between two procedures for analyzing earth observational data have been presented. The two procedures involved different implementations of the use of spatial information as an adjunct to spectral information. In the first the procedure involved the generation of additional features intended to quantify the texture perceived by a human observer when the data is placed in an image form. The second is a more mathematically abstract procedure in which the data is first partitioned into objects which are homogeneous in the sense that all multivariant pixels assigned to a given object are members of the same statistical population. In order to increase the validity of the comparison, a number of different texture algorithms were used and several other parameters were varied.

The results clearly show the value of computed spatial information when used in conjunction with multispectral data, and although only this single data set has been used for test so far, the performance of the ECHO procedure does appear to be better than that involving any of the textural features. It should also be pointed out that the ECHO type of procedure is computationally much more efficient than the calculation of textural features. ECHO usually requires about the same amount of computer time per pixel classified as the pixel by pixel classifier and, depending on the distribution of object sizes in the scene, sometimes less. Texture features, on the other hand, require calculation over and above that of the pixel by pixel classifier. In the case of this test, the additional computation required was considerable although the emphasis in the current study was to evaluate fully the potential for improved analysis results rather than to minimize the computation time.

REFERENCES

1. R. M. Haralick and K. S. Shanmugam, "Combined spectral and spatial processing of ERTS imagery data," SYMPOSIUM on Significant Results Obtained From the Earth Resources Technology Satellite-1. Vol 1., Goddard Space Flight Center, Maryland. March, 1973. National Aeronautics and Space Administration.

2. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst., Man, Cybern., Vol. SMC-3, pp 610-621, Nov., 1973.

3. R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," IEEE Trans. Geoscience Electronics, GE-14, No. L., Jan., 1976.

4. R. L. Kettig and D. A. Landgrebe, "Computer classification of remotely sensed multispectral image data by extraction and classification of homogeneous objects," LARS Information Note 050975. Purdue University, West Lafayette, IN., 1975.

5. T. L. Phillips, Ed., LARSYS Version III User's Manual, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN., 1973.

6. "Remote Multispectral Sensing in Agriculture," Vol. 3. Laboratory for Agricultural Remote Sensing, Annual Report, 1968, Purdue University, West Lafayette, IN.

7. A. Rosenfeld, "Picture Processing by Computer," Academic Press, New York, 1969.

8. P. H. Swain, "Pattern Recognition: A basis for remote sensing data analysis," LARS Information Note 111572, Purdue University, West Lafayette, IN., 1973.

9. A. G. Wacker and D. A. Landgrebe, "The minimum distance approach to classification," LARS Information Note 100771, Purdue University, West Lafayette, IN., 1971.

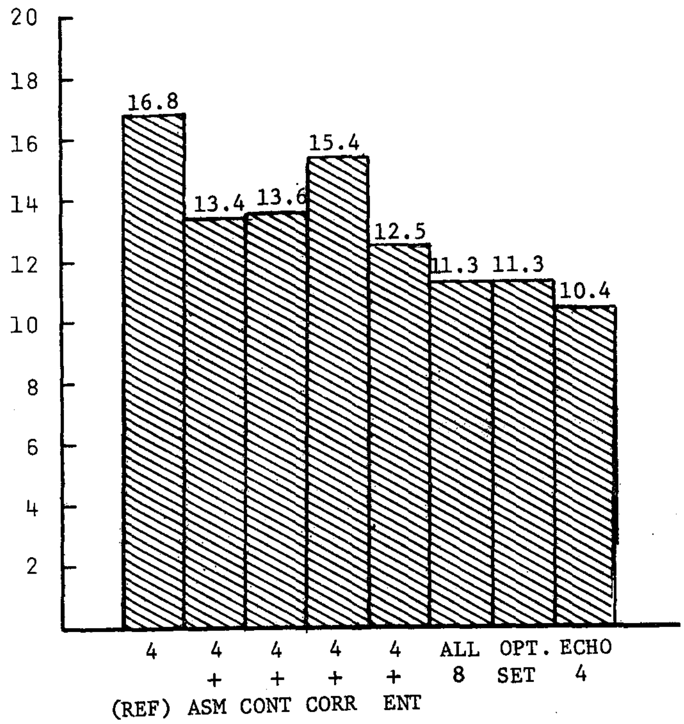

FIGURE 1. Comparison of classification performance between various combinations of four spectral and four textural features. Texture features computed from channel 4 with a block size of 11. The optimum set consists of the first two spectral bands plus ASM, CONT, and ENT.

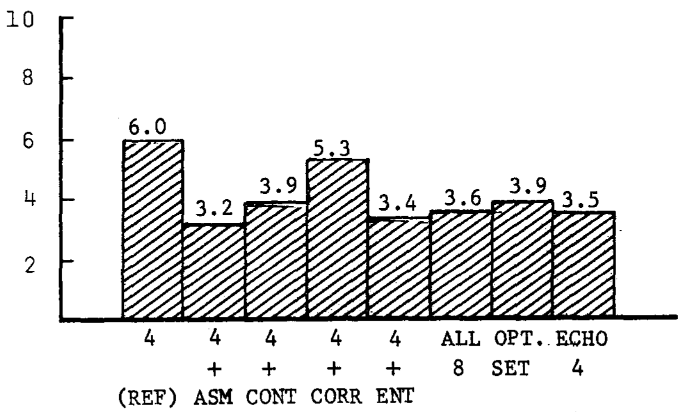

FIGURE 2. Comparison of training performance between various combinations of four spectral and four textural features. Texture features computed from channel 4 with a block size of 11. The optimum set consists of the first two spectral bands plus ASM, CONT, and ENT.

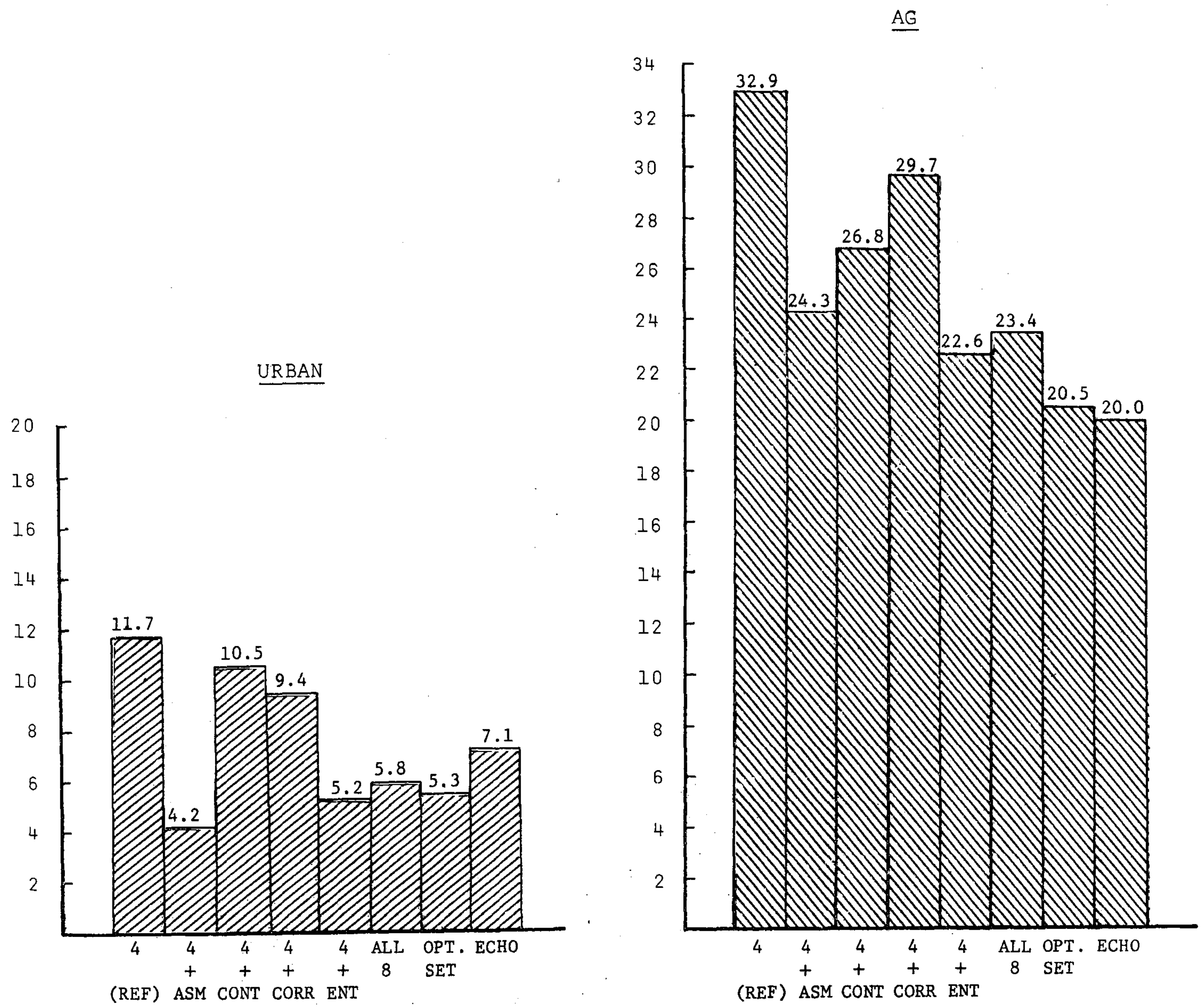

FIGURE 3. Text pixel performance for the classes Urban and Ag. Texture features computed from channel 2 with block size of 11.

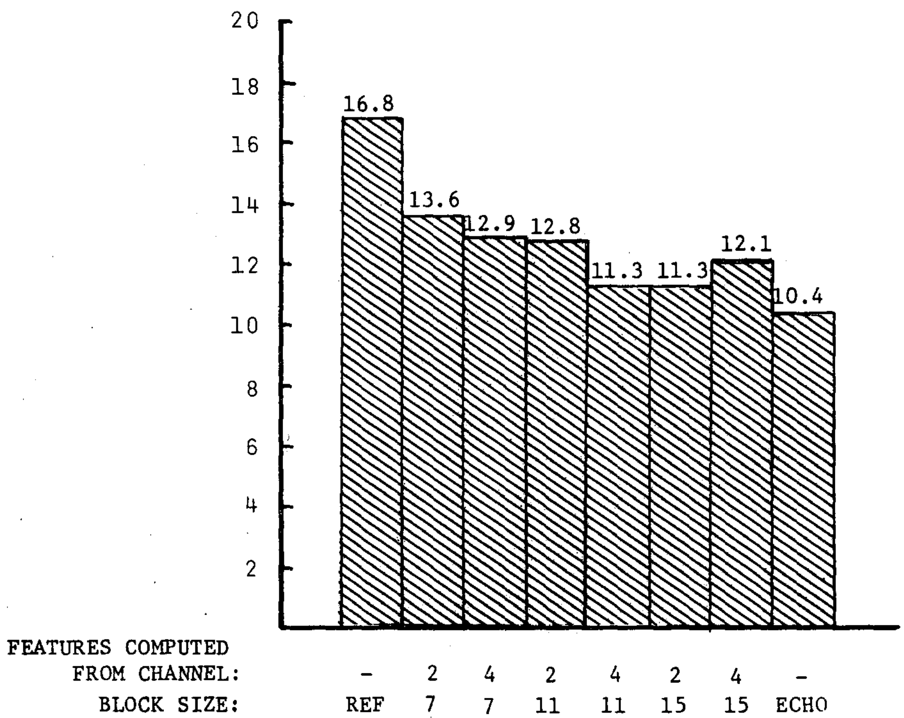

FIGURE 4. Test pixel classification performance for different block sizes and utilizing different spectral bands from which the texture features were computed. For REF and ECHO only the four spectral bands were used. All spectral and textural channels were used for the remaining classifications.