Purdue University
Purdue e-Pubs

LARS Symposia

Laboratory for Applications of Remote Sensing

10-19-1973

Comparative Evaluation of Spatial Features in Automatic Land Use Classification from Photographic Imagery

James H. Herzog
Oregon State University

Roy C. Rathja
Oregon State University

Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

COMPARATIVE EVALUATION OF SPATIAL FEATURES IN AUTOMATIC LAND USE CLASSIFICATION FROM PHOTOGRAPHIC IMAGERY

James H. Herzog and Roy C. Rathja

Department of Electrical and Computer Engineering
Oregon State University
Corvallis, Oregon 97331

I. ABSTRACT

Five spatial features have been evaluated for their applicability in automatic land use classification of photographic imagery. Data arrays of approximately 10,000 square meters were classified on the basis of an 8 by 8 point data array. Statistical features, information features, sequency features and texture features were evaluated using a distance to prototype classifier and an adaptive classifier. The results indicate approximately 70% accuracy in the classification.

II. INTRODUCTION

The successful classification of photographic imagery is strongly dependent upon the feature set utilized in the classification and on the classification process itself. In the case of photographic imagery, the process does not accurately preserve precision radiometric relationships among the entities being classified. As such, features such as point by point optical density have severe limitations in their applicability.

Classical human photointerpretation relies heavily on spatial interrelationships. These relationships are often subtle. In the case of shape detection, the computer is not blessed with the highly diffused parallel processing structure of the human. Shape detection is difficult and costly on a digital computer.

This investigation has chosen to study features which are of intermediate complexity between point features such as optical density and complex spatial features such as shape. These features were studied to determine their usefulness in automatic land use classification of high flight photographic imagery.

The problem selected for this investigation is the classification of land usage in the vicinity of Oregon City, Oregon. This region is located in the Willamette Valley of Western Oregon approximately 20 km (12.5 miles) south of Portland. The Willamette River is vital to the industry, agriculture, and recreation of the area.

Within the region four major land uses were chosen for automatic classification.

1) Forest - Mainly Oak, Douglas Fir and Red Alder

2) Agriculture - Primarily small grains

3) Water resources - Willamette River

4) Urban - Residential and manufacturing

The purpose of the investigation was to determine the ability of a computer using a digitized version of the photograph to correctly classify the four land use regions using the feature sets developed.

This work was supported in part by the National Aeronautics and Space Administration under Contract NAS 5-21831.

III. THE SOURCE DATA

The source photograph, Figure 1, for this investigation was obtained on June 22, 1972 by an RB 57 aircraft operated by NASA using a camera system with a 6 inch focal length lens. The original negative was digitized into 8 bit densities through the cooperation of the University of Oregon Chemistry Department. It was later converted to 6 bits for compatibility with data formats being developed for ERTS satellite imagery. The negative was digitized at 98.5 points per cm (250 points per inch). The distance between digital resolution points was approximately 13 meters (43 feet).

IV. THE AUTOMATIC CLASSIFICATION PROBLEM

Automatic classification of land use may be approached using the techniques of pattern recognition. The classical pattern recognition problem is shown in Figure 2.

Using the raw data, a feature extractor must first be specified to yield a feature vector for the raw data set associated with a Classification Unit (CU). Appropriate feature vectors are usually specified based on heuristic algorithms and insight into the nature of the classification problem.

In this investigation the raw data was grouped into square arrays of 64 data points. This resulted in a classification unit of approximately $1x10^4$ meter^2 ($1.2x10^5$ ft^2).

Five different feature vectors were used in this investigation. Each was evaluated separately to gain insight into its applicability to this problem.

The classifier receives feature vectors as inputs and produces decisions at its output. In developing the classifier, sets of features corresponding to known CU's are used. One set known as the training set is used to abstract information in establishing classification rules. In this investigation two classifiers were used. The Distance to Prototype classifier uses parameters derived from the training set to form prototype vectors for each class. The Adaptive classifier uses artificial intelligence techniques for developing the classification algorithm. Both of these classifiers were evaluated with each of five feature sets.

V. FEATURES

The four land use classifications are characterized by the following spatial characteristics:

Forest - Homogeneous with a mottled texture.

Agriculture - Homogeneous within a field with abrupt intensity changes at field boundaries.

Water - Very homogeneous.

Urban - Very non-homogeneous with many intensity changes.

A. STATISTICAL FEATURES

The mean value and the standard deviation of the 64 data points in the array were calculated. These two features comprise a feature vector F_s which is a measure of average optical density and of the homogeneity of the area.

$$F_s = \begin{bmatrix} \mu \\ \sigma \end{bmatrix} \quad (1)$$

$$\mu = \frac{1}{63} \sum_{i=1}^{8} \sum_{j=1}^{8} a_{ij} \quad (2)$$

$$\sigma^2 = \frac{1}{63} \sum_{i=1}^{8} \sum_{j=1}^{8} (a_{ij}-\mu)^2 \quad (3)$$

B. INFORMATION FEATURES

The information feature is based on the entropy of the data array.

$$F_I = [\rho] \tag{4}$$

$$\rho = -\frac{1}{64}\sum_{i=1}^{64} n(I_i)\ \mathrm{Log}_2[p(I_i)] \tag{5}$$

$n(I_i)$ = Number of data points of intensity I_i

$p(I_i)$ = Probability of intensity I_i

F_I is an indication of the distribution of intensities. It, like the standard deviation, is a measure of the homogeneity of the data. (4)

C. SPATIAL SEQUENCY DISTRIBUTION

The coefficients of the two dimensional Walsh function series were combined to give the Walsh function equivalent of the conventional Power Spectral Density.

The coefficients of the two dimensional Walsh series are given by the matrix operation.

$$H = W^T A W \tag{6}$$

where A is the 8 x 8 data array.

$$W = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & - & -1 & - & -1 & .1 & -1 \end{vmatrix} \tag{7}$$

$$\begin{array}{cc} & \begin{array}{cccccccc} H_0 & & H_1 & & H_2 & & H_3 & H_4 \end{array} \\ H = & \begin{vmatrix} h(00) & h(01) & h(02) & h(03) & h(04) & h(05) & h(06) & h(07) \\ h(10) & h(11) & h(12) & h(13) & h(14) & h(15) & h(16) & h(17) \\ h(20) & h(21) & h(22) & h(23) & h(24) & h(25) & h(26) & h(27) \\ h(30) & h(31) & h(32) & h(33) & h(34) & h(35) & h(36) & h(37) \\ h(40) & h(41) & h(42) & h(43) & h(44) & h(45) & h(46) & h(47) \\ h(50) & h(51) & h(52) & h(53) & h(54) & h(55) & h(56) & h(57) \\ h(60) & h(61) & h(62) & h(63) & h(64) & h(65) & h(66) & h(67) \\ h(70) & h(71) & h(72) & h(73) & h(74) & h(75) & h(76) & h(77) \end{vmatrix} \end{array} \tag{8}$$

H_i of the feature vector was formed by the sum of the squares of the indicated region of the Walsh coefficient matrix shown in equation 8. Each H_i contains information concerning spatial sequency i in the data array. (1)

$$F_W = \begin{vmatrix} H_0 \\ H_1 \\ H_2 \\ H_3 \\ H_4 \end{vmatrix} \tag{9}$$

D. TEXTURE I

The Texture I features were obtained from a compilation of changes in adjacent cells. With the 8 x 8 data array there are 56 adjacent point differences which can be calculated in both the horizontal and vertical direction. This gives a total of 112 differences which form a distribution based on the magnitude of the change. The magnitude of change was arbitrarily divided into 6 features. Their composition is related to the scaling of the data and the thresholds selected for the feature definition. (2)

d_0 = the number of differences of magnitude 0

. . . .

. . . .

d_4 = the number of differences of magnitude 4

d_5 = the number of differences of magnitude 5 or greater.

$$F_D = \begin{vmatrix} d_0 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{vmatrix} \quad (10)$$

E. TEXTURE II

Texture II consists of a single feature. Like Texture I it is based on the differences existing between adjacent data points.

$$d = d_1 + 2d_2 + 3d_3 + 4d_4 + 5d_5 \quad (11)$$

$$F_T = d \quad (12)$$

VI. THE CLASSIFIER

Two standard classification techniques were used in this four class problem to classify Agricultural, Forest, Water, Urban areas based on each of five possible feature vectors described in Section III.

A. DISTANCE TO PROTOTYPE

Mean values of each of the feature vectors were computed for each class using the training set data. These mean vectors were used as prototype vectors for the class.

A distance metric was calculated for an unknown vector X and each prototype.

$$J_P = \sum_{i=1}^{n} (f_{Pi} - x_i)^2 \quad (13)$$

where f_{Pi} is the ith component of the prototype for Class P.

X was assigned to the class having the smallest distance metric.

It is common to include a threshold on the distance metric so that if the distance measure is greater than the threshold, the unknown vector is assigned to a "Don't Know" class.

B. ADAPTIVE CLASSIFIER

The adaptive classifier is an example of a nonparametric classification scheme. Initially nothing need be known about the statistical characteristics of the data.

A weight vector K_P is associated with each class in the problem.

A function J_P is then calculated for the augmented feature vector for each class. (Augmenting consists of attaching a 1 as the last entry in the vector.)

$$J_P = K_P X \quad (14)$$

X is then assigned to the class for which the resulting J_P is the greatest. Adaptive algorithms are used to reward or penalize the weight vector depending on the correctness of the classification of known training vectors.

$$K_p = K_p + \alpha X \quad \text{for a correct classification} \tag{15}$$

$$K_p = K_p - \alpha X \quad \text{for an incorrect classification} \tag{16}$$

The training set was cycled through 25 times with decreasing values of α to get a "best" classification.

VII. EXPERIMENTAL PROCEDURE

The digitized data was used to construct a line printer gray scale printout using over-strike characters. With the aid of ground truth experts training sets and test sets were isolated on both the line printer output and the original photograph.

Sixty classification units (each of 64 points) were selected for each land use class for the training set. Sixty additional CU's were selected as test sets. In many cases the agricultural set included field boundaries. The five feature vectors were calculated for each of the classification units.

The prototype vector was formed from the mean values of the training set data for each class. These prototype vectors were used in the Distance to Prototype classifier. No threshold was used in the classification process.

In the Adaptive classifier the K_p vector initially contained a 1 in each location. α was chosen to have an initial value of 1. Feature vectors corresponding to the four different training sets were sequentially presented to the classifier. Equations 15, 16 were used to change the weight vector after each classification. The 60 training vectors for each of the four classes resulted in 240 changes in the weight vector for a fixed value of α.

After the nth pass of the training set the value of α was replaced by $1/(n+1)$ and the process was repeated. The adjustment and classification was repeated 25 times. At this point classification results were very stable.

VIII. RESULTS AND CONCLUSIONS

Classification results are shown in Table 1 and Table 2 for the training set data and test set data respectively. Inspection of the results from Table 1 and Table 2 indicates that all of the features examined had characteristics suitable for land use classification. Both the parametric and adaptive classifier performed adequately.

Urban areas were easily distinguished, in some cases with no errors. The wide dispersion of optical density coupled with rapid spatial variation was significantly different than the spatial characteristics of the remaining classes.

Forest, water, and agriculture were the most difficult to distinguish. Water presents a nearly homogeneous texture; forest texture and composition is slightly more varied. The characteristic of the agricultural area is greatly dependent on the crop and location of field lines. It was very rare for a sample of these three land uses to be classified as urban.

Of special interest is the difficulty of classifying the agricultural samples. The size of the classifying array was rather small compared to the field sizes in this region. This resulted in some agricultural samples containing field boundaries and others containing homogeneous vegetation. Adjacent fields often contained different crop composition.

In evaluating the effectiveness of each of the features it should be noted that the feature vectors contained a maximum of 6 elements (Texture I) and a minimum of one element (Texture II and Entropy). The Walsh series contained 5 elements and the statistical feature contained two elements. The dimensionality of the feature vector obviously influences the computation time of the classifiers.

The suitability of these features to conditions in which photographic exposure is not accurately controlled has not been established. It would appear that all of the features are more sensitive to changes in the optical density of the photograph than to the magnitude of the density itself. More work needs to be done in this area.

REFERENCES

1. Kawamura, J. G., "Automatic Recognition of Changes in Urban Development from Aerial Photographs," IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-1, July 1971, pp. 230-239.

2. LeShack, L. A., "ADP of Forest Imagery," Photogrammetric Engineering, August 1971, pp. 885-896.

3. Rosenfeld, A. G., "Automatic Recognition of Basic Terrain Types from Aerial Photographs," Photogrammetric Engineering, March 1962.

4. Sutton, R. N., and E. L. Hall, "Texture Measures for Automatic Classification of Pulmonary Disease," IEEE Transactions on Computers, Vol. C-21, No. 7, July 1972, pp. 667-676.

5. Wintz, P. A., "Transform Picture Coding," Proceedings of the IEEE, Vol. 60, No. 7, July 1972, pp. 809-820.

Table 1. Classification Results of Training Set Data

	Class	DISTANCE TO MEAN: Classified As A	F	W	U	% Correct	ADAPTIVE: Classified As A	F	W	U	% Correct
STATISTICAL	A	14	34	8	4	23	18	26	3	13	30
	F	0	60	0	0	100	17	42	1	0	70
	W	13	0	47	0	78	5	27	28	0	47
	U	0	0	0	60	100	19	0	0	41	68
	Total					75					54
INFORMATION	A	23	10	16	11	38	8	33	10	9	13
	F	14	41	5	0	68	0	59	1	0	98
	W	4	5	51	0	85	0	28	32	0	53
	U	0	0	0	60	100	0	0	0	60	100
	Total					73					66
TRANSFORM	A	10	35	8	7	17	35	7	12	6	58
	F	0	60	0	0	100	8	40	9	3	67
	W	31	0	29	0	48	9	6	44	1	73
	U	0	0	0	60	100	2	0	0	58	97
	Total					66					74
TEXTURE I	A	15	18	24	3	25	41	8	11	0	68
	F	14	42	4	0	70	23	33	4	0	55
	W	5	6	49	0	82	25	4	31	0	52
	U	0	0	0	60	100	0	0	0	60	100
	Total					69					69
TEXTURE II	A	22	9	26	3	37	29	11	14	6	48
	F	27	22	11	0	37	43	13	4	0	22
	W	1	6	53	0	88	25	1	34	0	57
	U	0	0	0	60	100	0	0	0	60	100
	Total					65					57

Table 2. Classification Results of Test Set Data

Feature	Class	Distance to Mean: Classified As A	F	W	U	% Correct	Adaptive: Classified As A	F	W	U	% Correct
STATISTICAL	A	19	12	13	16	32	26	9	2	23	43
	F	0	60	0	0	100	17	43	0	0	72
	W	16	0	44	0	73	6	32	22	0	37
	U	0	0	0	60	100	25	0	0	35	58
	Total					76					53
INFORMATION	A	27	6	7	20	45	12	28	4	16	20
	F	15	33	12	0	55	0	59	1	0	98
	W	5	11	44	0	73	1	29	30	9	50
	U	0	0	0	60	100	0	0	0	60	100
	Total					68					67
TRANSFORM	A	9	13	22	16	15	39	5	6	10	65
	F	0	60	0	0	100	7	37	14	2	62
	W	26	0	34	0	57	17	8	35	0	58
	U	0	0	0	60	100	4	1	0	55	92
	Total					68					69
TEXTURE I	A	22	22	13	3	37	45	6	6	3	75
	F	13	41	6	0	68	24	31	5	0	52
	W	7	6	47	0	78	23	4	33	0	55
	U	0	0	0	60	100	0	0	0	60	100
	Total					71					70
TEXTURE II	A	37	7	13	3	62	27	16	6	11	45
	F	24	24	12	0	40	45	13	2	0	22
	W	4	10	46	0	77	29	1	30	0	50
	U	0	0	0	60	100	0	0	0	60	100
	Total					70					54

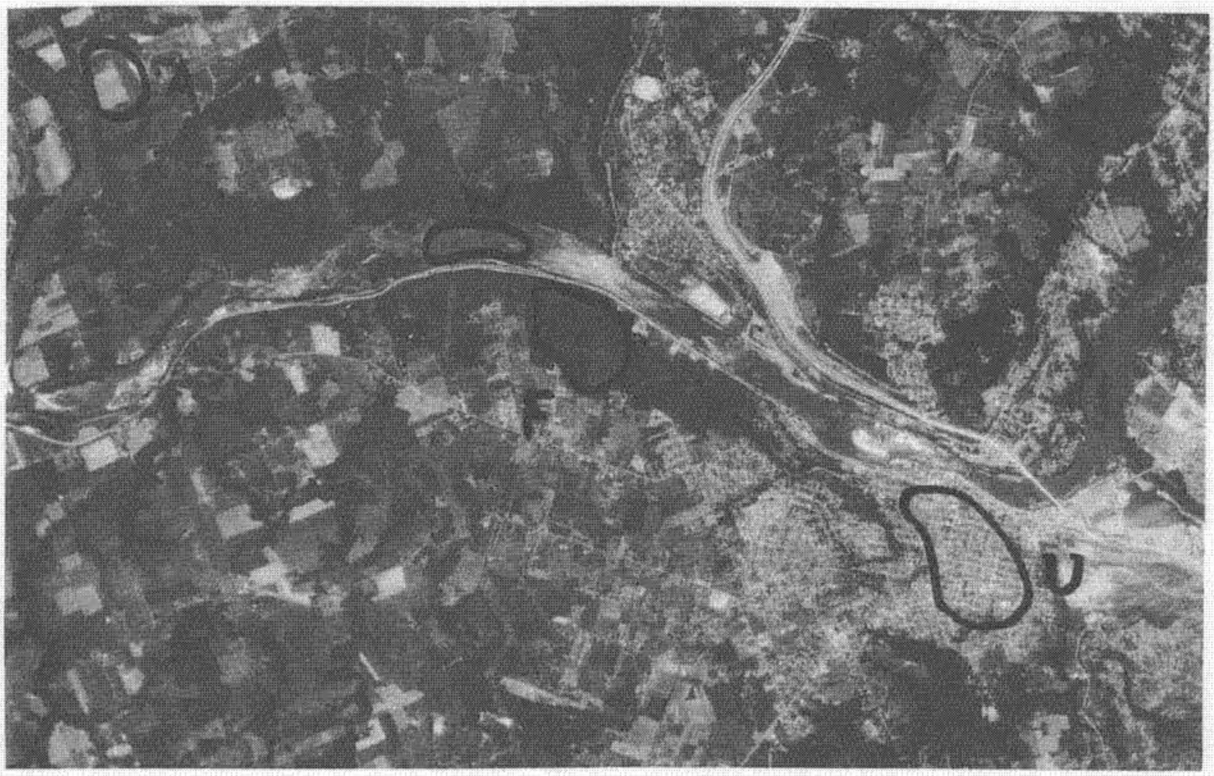

Figure 1. The source photograph for land use classification showing typical training sets and test sets.

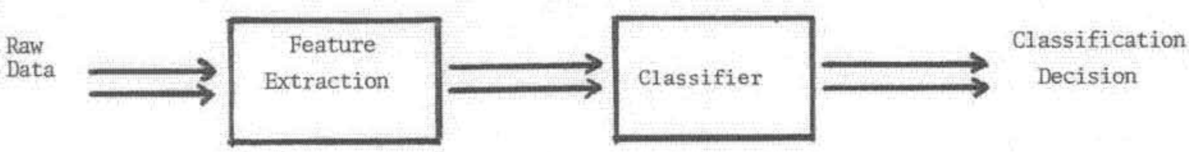

Figure 2. The pattern recognition problem.