E7.3 10816.
CR-133310

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

JAMES H. HERZOG AND ROY C. RATHJA*

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

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 sample set. Statistical features, information features, sequency features and texture features were evaluated using a distance to mean classifier and an adaptive classifier.

(E73-10816) COMPARATIVE EVALUATION OF SPATIAL FEATURES IN AUTOMATIC LAND USE CLASSIFICATION FROM PHOTOGRAPHIC IMAGERY (Oregon State Univ.) 9 p HC $3.00 CSCL 08B G3/13 N73-27291 Unclas 00816

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.

II THE SOURCE DATA

The source photograph for this investigation was obtained on June 22, 1972 by an RB 57 aircraft operated by NASA using a Vinton camera system with a 6 inch focal length lens. For this investigation imagery was used from the vicinity of Oregon City, Oregon. This area incorporates examples of the four land use classifications of interest in this study; forest, urban, water, agricultural.

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.

*Both authors are with the Department of Electrical and Computer Engineering, Oregon State University.

9 pp.

A line printer reconstruction of the original image was used in locating training sets and test sets of the features of interest. In all cases, ground truth experience was sufficient to readily classify the land use by manual photointerpretation.

Sixty square arrays each comprising 64 points were used for each of the four classes in the training set. A second set of 240 arrays was chosen as a test set.

III FEATURES

The features selected were chosen because of their association with each of several ways to represent the 8 by 8 data array, A to be classified.

A. Statistical Features

The mean value and the standard deviation of the 64 data points in the array were calculated. These two features provided a measure of average optical density and of the homogeneity of the area.

$$F_s = \left[\begin{matrix}\mu \\ \sigma\end{matrix}\right] \tag{1}$$

$$\mu = \frac{1}{64} \sum_{i=1}^{8} \sum_{j=1}^{8} a_{ij} \tag{2}$$

$$\sigma^2 = \frac{1}{64} \sum_{i=1}^{8} \sum_{j=1}^{8} (a_{ij}-\mu)^2 \tag{3}$$

B. Information Feature

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

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

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

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

$p(I_i)$ = Probability of intensity I_i

F 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.

$$H = WAW \tag{6}$$

Where A is the 8 x 8 data array.

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

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

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)

$$F_w = \begin{bmatrix} H_0 \\ H_1 \\ H_2 \\ H_3 \\ H_4 \end{bmatrix} \tag{9}$$

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.

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

E. Texture II

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

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

$$F_T = d$$

IV CLASSIFICATION TECHNIQUES

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 Mean

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.

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

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

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

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.)

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

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

$K_P = K_P + aX$ for a correct classification

$K_p = K_p - aX$ for an incorrect classification

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

V RESULTS

Classification results are shown in Table I and Table II for the training set data and test set data respectively. There is no evidence to suggest that the classification techniques chosen were particularly sensitive to any unique characteristic the training set.

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

TABLE II CLASSIFICATION RESULTS OF TEST SET DATA

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

TABLE I CLASSIFICATION RESULTS OF TRAINING SET DATA

VI CONCLUSIONS

Inspection of the results from Table I and Table II 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. P. A. Wintz, "Transform Picture Coding", Proceedings of the IEEE, Vol. 60, No. 7, July 1972, pp. 809-820

2. R. N. Sutton 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

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

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