Performance Analyses of Probabilistic Relaxation Methods for Land-Cover Classification

Peng Gong and Philip J. Howarth

Department of Geography, University of Waterloo

The performances of two probabilistic relaxation (PR) classification methods, a standard and a modified version, are assessed in terms of classification accuracy measured by the Kappa coefficient and the CPU time required to carry out the computation. The classification results obtained with these methods are compared with results obtained using conventional maximum-likelihood classification (MLC). Experiments indicate that the modified PR method significantly improves upon the classification results generated by the MLC method. The modified PR method saves up to 70% of the CPU time, compared with the standard PR method, and also gives slightly better classification accuracy.

INTRODUCTION

Recent studies of land classification in the rural–urban fringe using satellite data have indicated that the commonly-used per-pixel classifiers do not produce higher classification accuracies from higher spatial resolution data (Toll, 1984; Irons et al., 1985; Martin et al., 1988). Such classifiers include the maximum-likelihood classifier (MLC) and the minimum-distance classifier (MDC), which only use the spectral information from each pixel on an individual basis. The large amount of spatial information existing in an image is thus ignored. To increase classification accuracies using higher spatial resolution data, it is desirable to incorporate both spectral and spatial information into the classification process.

Three approaches for incorporating spatial information in a classification can be identified. Each approach is used at a different stage of the process. The first group consists of the preclassification approaches which use spatial information extracted from the original images as additional bands of data in the classification process. Such methods include the use of filtered images (e.g., Dutra and Mascarenhas, 1984; Cushnie, 1987), images containing structural information such as edge density (Hlavka, 1987; Gong and Howarth, 1990) and region characteristics such as the mean and variance of gray levels, size, shape, and compactness (Ketting and Landgrebe, 1976; Egawa and Kusaka, 1988). Texture features (e.g., Haralick et al., 1973; Weszka et al., 1976; Hsu, 1978; Jensen, 1979) are also included in this group. The second group of approaches may be categorized as post-classification methods. They use logic filters to reduce noise in the classified results (e.g., Gurney, 1981; Townshend, 1986). The third group involves use of contextual classifiers and compound decision rules (e.g., Welch and Salter, 1971; Fu and Yu, 1980; Landgrebe, 1980; Swain et al., 1980; Owen, 1984;

Address correspondence to P. J. Howarth, Dept. of Geography, Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada.

Received 29 May 1989; revised 11 September 1989.

Haralick and Joo, 1986), or the use of probabilistic relaxation (e.g., Richards et al., 1982; Harris, 1985; Gong and Howarth, 1989). Although the effects of the first two approaches are well known in remote sensing, the third type of approach has received relatively little attention.

In this study, we focus on the evaluation of two probabilistic relaxation (PR) methods for land-cover classification at the rural–urban fringe. The first PR method is the standard one, proposed by Rosenfeld et al. (1976) and further developed and tested by several others including Peleg and Rosenfeld (1978), Eklundh et al. (1980), Richards et al. (1981a), Kalayeh and Landgrebe (1984), and Lee and Richards (1989). In this relaxation procedure, all the probabilities for each pixel are used in the probability updating. The second method is a modified PR approach (Gong and Howarth, 1989) in which only a few significant probabilities for each pixel are used. In addition, a probability threshold is set in this second method which excludes a pixel from probability updating when its largest probability is higher than a specific threshold.

The objectives of this study are:

- To compare the effectiveness of the two PR methods in improving the classification results when compared with the MLC method.
- To examine, using the modified PR approach, the effects of different thresholds on the classification result.
- To evaluate the effects on the classification result of using weighting factors in the PR process.

The criteria for evaluating these results are the classification accuracies measured by means of the Kappa coefficient (Cohen, 1960) and the CPU time required to undertake the classification.

STUDY AREA, DATA, AND CLASSIFICATION SCHEME

The rural–urban fringe of northeastern Metropolitan Toronto (43°49′N; 79°10′W) was selected as the site for this study. Urban expansion has been occurring rapidly in recent years with large tracts of agricultural land being converted into built-up areas. The area is therefore of interest for studies of rural-to-urban land conversion and land-cover classification (Johnson and Howarth; 1987; Howarth et al., 1988; Martin et al., 1988; Martin, 1989; Gong and Howarth, 1989; 1990).

A 512×512 pixel subscene of SPOT HRV multispectral (XS) data acquired on 4 June 1987 was used in the study. In order to make it easier to compare the satellite imagery with other information such as aerial photographs and maps of the area, a geometric correction was applied to transform the original XS data to the UTM projection with 20 m$\times$20 m resampled pixels. This was achieved using a third-order polynomial and cubic convolution interpolation. As the study area is relatively small and flat, both topographic and atmospheric conditions were assumed to be homogeneous. Therefore, no further radiometric correction was made to the XS data. In order to reduce the amount of computation, the three original bands of XS data were transformed into two bands through principal component analysis (PCA). The two new bands of data contain over 99% of the total variance of the original data.

Twelve land-cover classes were used in the study. They are residential roof, paved surface, industrial and commercial roof, cleared land, lawn and tree complex, cultivated grass, deciduous tree, coniferous tree, crop cover, new crop and pasture, bare field, and water surface. A more detailed discussion of the characteristics of these land-cover classes is to be found in Gong and Howarth (1990).

METHOD AND ALGORITHM

The Standard Probabilistic Relaxation Approach

The PR model can be illustrated using land-cover classification as an example. Assume that a digital image with a size of N pixels is to be classified into m classes $(c_1, c_2, \ldots, c_m)$. The probability that a pixel i is classified into class c_j is defined as $P_i(c_j)$. For a whole image there are $N \times m$ probabilities. They satisfy the condition

$$\sum_{j=1}^{m} P_i(c_j) = 1, \quad \text{for all } i, \text{ and for all } j \text{ and } i. \quad 0 \le P_i(c_j) \le 1, \tag{1}$$

The values of $P_i(c_j)$ can be determined in several ways. Most frequently, there is no other source of data and thus they are derived from a simpler classification procedure such as the maximum-likelihood classification, as discussed in more detail later (Eklundh et al., 1978; Harris, 1985). Commencing with these values as initial values, the relaxation process employs probability information from predefined neighboring pixels, iteratively, to create new probabilities for each pixel. It is thus expected that any ambiguities that occur in the original classification will be reduced. The new probability for pixel i with class label c at the $(k+1)$th iteration is estimated through the updating rule

$$P_{i,k+1}(c) = \frac{P_{i,k}(c)[1+q_{i,k}(c)]}{\sum_{j=1}^{m} P_{i,k}(c_j)[1+q_{i,k}(c_j)]}, \quad (2)$$

where k indicates the iteration, with $k=0$ representing the initial probabilities obtained from a simpler classification. The denominator ensures that the newly created probability satisfies Eq. (1). The term $q_{i,k}(c)$ is the updating factor or neighborhood operator which expresses the influence of all the predefined neighboring pixels on the pixel i. It is determined from the equation

$$q_{i,k}(c) = \sum_{j=1}^{n} w_{ij} \sum_{c_j=c_1}^{c_m} r_{ij}(c, c_j) P_{j,k}(c_j), \quad (3)$$

where n is the number of the predefined neighborhood pixels and w_{ij} is the weighting factor from neighborhood pixel j to pixel i (j may be i itself). Often the weights of all the neighborhood pixels for pixel i are assumed to be equal (Peleg and Rosenfeld, 1978; Eklundh et al., 1980). According to Richards et al. (1981b), it is also reasonable to assume that the weight of pixel i (whose probabilities are to be updated) will be higher than each of its neighbors. The term $r_{ij}(c, c_j)$ is the compatibility coefficient between pixel i with class c and one of its neighborhood pixels j with class c_j. It is this coefficient that combines the information from the neighboring pixel. Detailed discussion on the calculation of the compatibility coefficients can be found in Peleg and Rosenfeld (1978) and Haralick (1983). One commonly used method, namely correlation compatibility coefficients, is applied in this study.

A detailed explanation of the relaxation process is given by Harris (1985). The relaxation process [Eq. (2)] continues until a certain criterion is met. The criterion can be the number of iterations, the fixed point which means that no changes between two successive iterations are observed (Haralick et al., 1980), or when all the global context has been taken into account (Haralick, 1983).

Derivation of the m Initial Probabilities for Each Pixel

According to Richards (1986), a suitable discriminant function for each pixel X with class j in its natural logarithm form is

$$g_j(X) = \ln P(c_j) - 0.5\ln|\Sigma_j| - 0.5(X - M_j)^T \Sigma_j^{-1}(X - M_j), \quad (4)$$

where X is the vector of spectral reflectance for the pixel to be classified, $g_j(X)$ stands for the discriminant function for class j, and $P(c_j)$ is the *a priori* probability of class j. Usually measured values of the *a priori* probabilities are not available at the beginning of the classification, so that they are assumed for each class. The term $|\Sigma_j|$ is the determinant of the variance–covariance matrix of class j, Σ_j^{-1} is the inverse of the variance–covariance matrix for class j, and M_j is the mean of class j. The terms $|\Sigma_j|$, Σ_j^{-1}, and M_j are obtained from training data. The value of $g_j(X)$ is the logarithm of the probability for pixel X having class j, under the assumption that the data are class-conditionally Bayesian. To initialize the relaxation, Eq. (1) has to be satisfied. A value for $P_i(c_j)$ can be derived by normalizing the antilogarithms of $g_j(X)$ obtained in Eq. (4). It should be noted that the antilogarithm of $g_j(X)$ is proportional to the real probability for pixel X to have a label c_j.

The Modified Probabilistic Relaxation Approach

Gong and Howarth (1989) showed that the PR process may not necessarily require all the probabilities for each pixel to be labeled. Because the magnitudes of the probabilities for each pixel differ significantly, it was suggested that only a few of the largest probabilities for each pixel be included in the PR process. To obtain these larger probabilities, a sorting process has to be undertaken after all the m probabilities for each pixel have been calculated. From this point on, only the larger probabilities are stored and used to calculate the initial probabilities. All other probabilities are considered as zero.

The second difference between the modified PR model and the standard one is that a thresholding option is included in the modified model. As the purpose of the PR process is to reduce the class labeling ambiguity for each pixel, it is reasonable to exclude a pixel from probability updating if there is less ambiguity for this pixel. The higher the largest probability of a pixel to be assigned to a specific class, the lower is the classification ambiguity. Bearing this in mind, a probability threshold can be set for the PR process. If a pixel's largest probability is higher than the threshold, the probabilities for this pixel will be fixed and thus will no longer be updated in the relaxation. In this case, the amount of computation will be reduced compared to the standard PR model in which every pixel is taken into account.

The last change made to the standard PR process is the determination of weighting factors. In the standard model, the influences of neighborhood pixels (except the one for which the probabilities are to be updated) are equally weighted. This has the disadvantage that too much uncertainty from the neighboring pixels may be added to the probability updating. It is expected that neighboring pixels which have lower classification certainties should be weighted less than those having higher certainties. Such a certainty factor for a pixel can be derived from the sum of the $g_j(X)$'s obtained in Eq. (4). The reason for doing this is based on the observation that if a pixel has a higher probability of being classified into a certain class (i.e., is less ambiguous), then the sum of its probabilities tend to be larger. The weighting factor for a neighborhood pixel j to pixel i is therefore modified in the form

$$W_{ij} = a_j S_j / \sum_{k=1}^{n} a_k S_k, \tag{5}$$

where the terms w_{ij} and n were introduced earlier. S_j is the certainty indicator for pixel j, k denotes all the n neighborhood pixels for pixel i, and a is the standard weighting factor. As pixel i is usually included in the neighborhood, its weight is a_i. If the weights of all the other pixels are equal, each weight is $(1-a_i)/(n-1)$.

With these modifications, it is expected that the PR process will be able to handle larger size images and to obtain more classes faster, in comparison to the standard PR model.

TESTS

In this study, the procedures described above have been implemented on a VAX11/785 computer using the FORTRAN 77 programming language. The amounts of CPU time calculated in this paper are therefore dependent on this particular machine and the programming efficiency of the authors. Thus, CPU times should only be interpreted in a relative manner.

A neighborhood of nine pixels was used in the probability updating. The neighborhood contains the pixel i, whose probabilities are to be updated, and its eight surrounding neighbors. Four different values for the weighting factor a_i were tested. They were 0.11, which indicates equal weights in the neighborhood for the standard PR process, 0.15, 0.20, and 0.25. The higher figures gave pixel i more weight, indicating an increasing influence from pixel i. Four thresholds (0.5, 0.7, 0.9, and 1.0) were tested in the modified PR model. The last figure (i.e., 1.0) in fact means that no thresholding is employed in the modified PR process. The classification results and CPU time used by the modified PR model with the threshold 1.0 were compared to the results and times associated with the standard PR model. The number of significant probabilities for each pixel used in the modified PR process was four. The use of four different weighting factors constituted four trails for the standard PR model. The combination of the four thresholds and the four weighting factors resulted in 16 trials for the modified PR model. For both the standard and the modified PR models, 20 iterations were tested within each trial.

The initial probabilities for each pixel were obtained through a supervised maximum-likelihood classification. Training samples for each land-cover class consisted of 60 pixels. They were selected on a single pixel basis with the aid of 1:8,000 scale aerial photographs obtained in April 1987.

Test pixels for accuracy assessment were obtained through a stratified systematic unaligned sampling strategy (Jensen, 1983). A 16×16 block was used. As a result, 1024 pixels were obtained for the test sample. The identity or land-cover class of each sample pixel was recorded by the analyst. It should be noted, however, that even with the help of the aerial photographs, the analyst encountered difficulties in labeling some of the mixed pixels. Using such data in the accuracy assessment may result in an underestimate of the classification accuracies. A comparison of the reference or ground data and the classification results for each iteration of the PR process permitted a confusion matrix to be produced.

The Kappa coefficient $\hat{K}$ (Cohen, 1960) and its variance $\hat{V}$ (Fleiss et al., 1969) were then calculated for each confusion matrix. For an $m \times m$ confusion matrix, let p_{ij} be the proportion of subjects placed in the i,jth cell; let p_{i+} and p_{+j} be the proportions of subjects placed in the ith row and jth column, respectively. Then, with

$$p_0 = \sum_{i=1}^{m} p_{ii} \quad \text{and} \quad p_c = \sum_{i=1}^{m} p_{i+} p_{+i}, \qquad (6)$$

the Kappa coefficient $\hat{K}$ is defined by

$$\hat{K} = \frac{p_0 - p_c}{1 - p_c}, \qquad (7)$$

where p_0 and p_c indicate the proportion of units which agree and the proportion of units for expected chance agreement, respectively. With the above definition, Fleiss et al. (1969) showed that the most appropriate method to estimate the variance of $\hat{K}$ is

$$\hat{V} = \frac{1}{N(1-p_c)^4} \Bigg\{ \sum_{i=1}^{m} p_{ii} \big[(1-p_c) - (p_{i+} + p_{+i})(1-p_0) \big]^2 + (1-p_o)^2 \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij} (p_{i+} + p_{+j})^2 - (p_0 p_c - 2p_c + p_0)^2 \Bigg\}. \qquad (8)$$

To determine the difference between two $\hat{K}$'s, the significance test proposed by Cohen (1960) for comparing two classification results was adopted. With this method, the difference between two Kappa coefficients resulting from two classifications is first obtained. The square root of the sum of the variances $\hat{V}$ between the two classifications is then calculated. A Z-value can be determined by dividing the difference by the square-root. A Z-value above 1.960 indicates that the two classification results are significantly different at the 95% confidence level.

The $\hat{K}$ has been recommended by Rosenfield and Fitzpatrick-Lins (1986) as a suitable accuracy measure in thematic classification for representing the whole confusion matrix. It takes all the elements in the confusion matrix into consideration, rather than just the diagonal elements which occurs with the calculation of overall classification accuracy. The variance was used when significance tests were made.

RESULTS AND DISCUSSION

Before the relaxation procedure was initiated, a maximum-likelihood classification was applied to the two PCA images, using the training samples described above. As a result, a Kappa coefficient of 0.394 was obtained. This can be compared with the results obtained by the PR models. In order to present the results, five tables (Tables 1–5) were generated with Table 1 containing the results for the standard PR process, and the remainder listing the results obtained with the modified PR process using thresholds of 0.5, 0.7, 0.9, and 1.0. Within each table, the Kappa coefficient was calculated for each iteration. Kappa coefficients obtained by the use of each weighting factor are listed in individual columns. The average Kappa coefficient for each column has also been calculated to determine the effects of the change of weighting factor on the classification results. The CPU time used is listed for each weighting factor (i.e., each trial)

Table 1. Kappa Coefficients Obtained and CPU Time Used by the Standard PR Model

	Weighting Factor			
Iterations	*0.11*	*0.15*	*0.20*	*0.25*
1	0.406	0.407	0.407	0.407
2	0.413	0.410	0.411	0.413
3	0.422	0.424	0.423	0.415
4	0.427	0.426	0.424	0.424
5	0.433	0.428	0.429	0.425
6	0.435	0.436	0.431	0.428
7	0.434	0.431	0.435	0.432
8	0.433	0.436	0.432	0.435
9	0.434	0.432	0.435	0.433
10	0.4361	0.433	0.436	0.433
11	0.436	0.436	0.436	0.437
12	0.433	0.4371	0.433	0.436
13	0.436	0.436	0.436	0.4372
14	0.433	0.434	0.436	0.436
15	0.430	0.434	0.4370	0.4384
16	0.429	0.436	0.4370	0.436
17	0.431	0.433	0.434	0.436
18	0.429	0.434	0.433	0.435
19	0.430	0.430	0.435	0.435
20	0.425	0.430	0.434	0.436
CPU time (h:min.s)	14:30.26	14:24.52	14:34.26	14:34.46
Significance test	1.684	1.724	1.724	1.780

Table 2. Kappa Coefficients Obtained and CPU Time Used by the Modified PR Model with a Threshold of 0.5

Iterations	*Weighting Factor* 0.11	0.15	0.20	0.25
1	0.402	0.402	0.402	0.402
2	0.408	0.409	0.409	0.407
3	0.413	0.412	0.412	0.412
4	0.413	0.413	0.413	0.413
5	0.413	0.413	0.413	0.413
6	0.412	0.412	0.412	0.411
7	0.413	0.413	0.413	0.412
8	0.413	0.413	0.413	0.413
9	0.413	0.413	0.413	0.413
10	0.415	0.415	0.415	0.415
11	0.415	0.415	0.415	0.415
12	0.415	0.415	0.415	0.415
13	0.413	0.413	0.415	0.415
14	0.413	0.413	0.413	0.415
15	0.413	0.413	0.413	0.413
16	0.413	0.413	0.413	0.413
17	0.413	0.413	0.413	0.413
18	0.413	0.413	0.413	0.413
19	0.413	0.413	0.413	0.412
20	0.413	0.413	0.413	0.412
CPU time (h:min.s)	1:06.39	1:08.02	1:09.14	1:04.55
Significance test	0.845	0.845	0.845	0.845

Table 3. Kappa Coefficients Obtained and CPU Time Used by the Modified PR Model with a Threshold of 0.7

Iterations	*Weighting Factor* 0.11	0.15	0.20	0.25
1	0.404	0.403	0.404	0.404
2	0.417	0.416	0.413	0.413
3	0.429	0.428	0.428	0.427
4	0.431	0.429	0.430	0.427
5	0.434	0.434	0.435	0.432
6	0.434	0.435	0.437	0.435
7	0.439	0.437	0.435	0.435
8	0.436	0.434	0.436	0.437
9	0.440	0.438	0.437	0.435
10	0.443	0.444	0.444	0.440
11	0.441	0.443	0.442	0.444
12	0.440	0.439	0.441	0.443
13	0.4431	0.44433	0.440	0.442
14	0.442	0.442	0.44430	0.443
15	0.439	0.440	0.442	0.44430
16	0.438	0.438	0.441	0.441
17	0.438	0.438	0.438	0.441
18	0.436	0.438	0.438	0.438
19	0.438	0.437	0.437	0.438
20	0.437	0.438	0.437	0.438
CPU time (h:min.s)	4:12.07	4:20.18	4:28.57	4:36.57
Significance test[a]	1.966*	2.014*	2.013*	2.013*

[a] * indicates the test is significant at the 95% confidence level.

Table 4. Kappa Coefficients Obtained and CPU Time Used by the Modified PR Model with a Threshold of 0.9

Iterations	*Weighting Factor* 0.11	0.15	0.20	0.25
1	0.404	0.403	0.403	0.402
2	0.417	0.416	0.413	0.412
3	0.429	0.428	0.428	0.427
4	0.431	0.429	0.429	0.428
5	0.435	0.434	0.433	0.432
6	0.435	0.435	0.436	0.436
7	0.437	0.440	0.439	0.436
8	0.441	0.439	0.440	0.439
9	0.441	0.440	0.438	0.438
10	0.4426	0.4439	0.4431	0.441
11	0.440	0.441	0.443	0.4428
12	0.4426	0.441	0.440	0.4425
13	0.440	0.440	0.443	0.442
14	0.442	0.442	0.441	0.441
15	0.440	0.441	0.441	0.442
16	0.441	0.441	0.440	0.4427
17	0.439	0.440	0.441	0.442
18	0.437	0.438	0.440	0.440
19	0.435	0.437	0.438	0.440
20	0.437	0.436	0.436	0.438
CPU time (h:min.s)	7:58.56	8:02.11	8:22.02	8:30.51
Significance test[a]	1.944	2.002*	1.970*	1.952

[a] * indicates the test is significant at the 95% confidence level.

within each table. The largest Kappa coefficient in each trial is underlined and is referred to as the "trial Kappa maximum." In each trial, the trial Kappa maximum and the number of the iteration at which the maximum is obtained are of importance. The trial Kappa maximum reflects how much improvement in accuracy a PR process can achieve. On the other hand, the number of the iteration at which the maximum is achieved indicates how soon the best result will occur in the PR process. From the point of view of the CPU time, the sooner the better. Finally, the difference between the trial Kappa maximum and the Kappa coefficient obtained from the MLC has been transformed into a standard normal distribution score, which is used for the significance test. A score of 1.96 is the lowest for accepting a significant difference at the 95% confidence level.

Tables 1 and 5 display the results obtained from the standard and the modified PR model, both with a threshold of 1.0. A comparison of the two tables indicates that the standard model used about 10–30 min more CPU time than the modified model in the four different trials, while each trial maximum obtained by the standard model

Table 5. Kappa Coefficients Obtained and CPU Time Used by the Modified PR Model with a Threshold of 1.0

	Weighting Factor			
Iterations	*0.11*	*0.15*	*0.20*	*0.25*
1	0.404	0.403	0.403	0.402
2	0.417	0.416	0.413	0.412
3	0.429	0.428	0.428	0.427
4	0.434	0.430	0.429	0.428
5	0.435	0.434	0.433	0.435
6	0.436	0.436	0.438	0.435
7	0.438	0.440	0.439	0.436
8	0.442	0.441	0.440	0.440
9	0.442	0.442	0.442	0.442
10	0.4427	0.4441	0.441	0.441
11	0.439	0.440	0.4425	0.4428
12	0.441	0.440	0.439	0.441
13	0.440	0.442	0.441	0.439
14	0.440	0.442	0.441	0.441
15	0.439	0.439	0.440	0.442
16	0.440	0.440	0.439	0.441
17	0.438	0.438	0.439	0.438
18	0.437	0.439	0.439	0.439
19	0.435	0.436	0.437	0.438
20	0.438	0.437	0.436	0.437
CPU time (h:min.s)	14:07.18	14:01.42	14:21.56	13:56.46
Significance test[a]	1.948	2.006*	1.940	1.952

[a] * indicates the test is significant at the 95% confidence level.

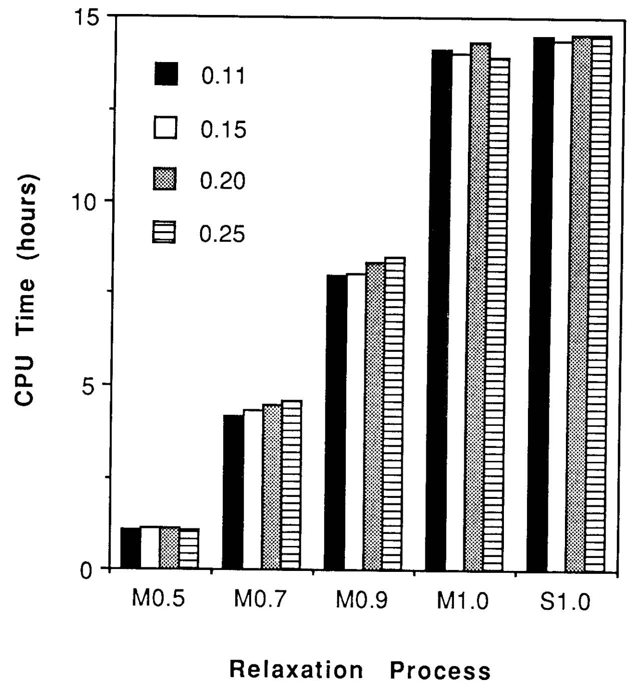

Figure 1. CPU time used by the standard PR model (S) and the modified PR model (M) with thresholds of 0.5, 0.7, 0.9, and 1.0. The weighting factors are 0.11, 0.15, 0.20, and 0.25.

was about 0.005 less than the corresponding value produced by the modified model. The scores for the significance test of the modified model are higher than for the standard one. In Table 5, one trial Kappa maximum (obtained using a weighting factor of 0.15 for the pixel to be updated) even shows a significant improvement over the value for the MLC at the 95% confidence level. It can also be observed from the two tables that the modified model reaches its trial Kappa maximum faster than the standard model.

To assess the effects of different threshold settings in the modified model, the CPU time used at each test threshold has been plotted against that used in the standard model (Fig. 1). It can be seen that by the process of thresholding, the CPU time required in the PR process has been greatly reduced. As one might expect, the lower the threshold, the fewer are the pixels for which the probabilities need to be updated. Consequently, the use of a lower threshold results in a large saving of CPU time. To select an appropriate threshold, however, one needs to consider not only the amount of CPU time saved, but also the trial Kappa maximum. The ideal situation is when the threshold with the lowest CPU time is selected for the highest Kappa value that can be achieved. However, this is not the case in this study. As can be seen from Table 2, although the CPU time reaches a minimum of about 1 h and 10 min at the threshold level of 0.5, the classification results are much poorer than those obtained at the other thresholds (Tables 3, 4, and 5). This is because too many pixels with high ambiguities have been exempted from the probability updating at the low threshold level. The scores for the significance test at this threshold are the lowest shown in all five tables. Usually, one needs to compromise between computing time and accuracy.

By comparing Tables 3, 4, and 5, it can be seen that when thresholds of 0.7, 0.9, and 1.0 are used, there are some minor differences among trial Kappa maxima. The threshold of 0.7 is preferred to the other two. This is supported by two important observations. First, the highest trial Kappa maximum among the three tables is found in Table 3. Second, all the four trial Kappa maxima in Table 3 passed the significance test at the 95% confidence level, while this was achieved in only two tests in Table 4 and one test in Table 5. The CPU time required at the 0.7 threshold level is about 4 h and 30 min, which is approximately half the time required at 0.9, and about 30% of that required at 1.0. Therefore, 0.7 seems to be the optimum threshold for this study when both the CPU time and the accuracy are taken into consideration.

A specific threshold will determine the number of times that probability updating occurs in a trial. The number of times it occurs is approximately proportional to the amount of computation measured by the CPU time in the PR process. It is therefore possible to estimate from the CPU time the number of times that probability updating occurs in the PR process at a certain threshold. If the threshold of 1.0 is taken to mean that 100% probability updating has occurred, then the percentages for thresholds 0.5, 0.7, and 0.9 are approximately 8, 30, and 55, respectively. Hence, for this study, it appears that an optimum threshold can be obtained when about 30% of the probability updating has been undertaken. If the probability distribution for the largest probability of each pixel is known beforehand, the threshold may be selected more easily.

From the five tables, it can be observed that the difference in weighting factors affects the number of iterations required to achieve the trial Kappa maximum. Almost every table shows the same trend, namely, that the use of a smaller weighting factor means that the maximum is achieved faster. This is not surprising when the reasons for setting the weighting factors are considered. The larger the weighting factor for the pixel whose probabilities are to be updated, the smaller are the weights of its neighboring pixels. The smaller weighting factors of neighboring pixels means that these pixels have less influence in the probability updating process. However, it is these neighboring pixels which bring spatial information to the classification in each iteration. Therefore, for spatial information to have a similar degree of effect on the classification results, the PR process needs more iterations with smaller weights from neighboring pixels. By comparing the magnitudes of all the trial Kappa maxima produced by the modified model in Tables 3, 4, and 5, the weighting factor of 0.15 seems to be preferable. This is because in each of these tables the highest trial Kappa maximum came from the trial when the weighting factor of 0.15 was used.

From Figure 1, it is apparent that a standard PR model requires too much computation. One of the major purposes of the modified PR model is to reduce the amount of computation needed in the PR process. From the results obtained in this study, it can be seen that a moderate reduction of computation has been achieved. However, the time involved is still too much for this model to be considered operational on systems such as the VAX 11/785 and the particular software implementation used. To overcome this problem, faster computer systems such as parallel processing systems and contextual classification algorithms requiring less computation are necessary.

In this study, analysis of the thresholds and weighting factors was done in an exhaustive manner because very little is known about the PR process. However, the work does not provide a complete understanding of the entire PR process. For example, the degradation of Kappa values can be observed in every table presented in this study. The reason for this effect is not immediately apparent to the authors. It is suspected, however, that more noise from nonlocal areas could have been introduced into the probability updating process as the number of iterations got larger. This would mean that a larger neighborhood had been incorporated into the analysis.

Further research is required in several directions. First, it is necessary to reduce the iteration time (or the converging time) required to reach the maximum accuracy or to test the use of noniterative methods. Second, it is important to test the use of different compatibility coefficients, such as the nonstationary method (e.g., Kalayeh and Landgrebe, 1984), and the weighting strategies. Third, it is suggested that the PR methods be applied in a range of different landscapes. In this way, researchers should achieve a better understanding of the effects of selecting different thresholds and weighting factors.

CONCLUSIONS

In comparison with the MLC method, the modified PR model can significantly improve classification accuracy. The modified PR model performed slightly better than the standard model in this study. It is possible to select an optimum threshold for the PR process, which will not only reduce the computation requirement considerably, but will also produce better classification results. However, there is no efficient way to select the optimum threshold. Weighting factors affect the time required to achieve maximum accuracy. The PR process involves many parameter options and is computationally complicated. Further research is needed to completely understand the entire PR process.

The authors gratefully acknowledge the assistance of SPOT Image Corporation of France and the Canada Centre for Remote Sensing in supplying the SPOT data used in this study as part of the Programme d'Évaluation Préliminaire SPOT (PEPS), Project No. 229. This research is funded by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science and NSERC Operating Grant A0766 awarded to P. J. Howarth. Mr. Gong's studies are supported by the International Development Research Centre (IDRC) of Ottawa. Thanks to Joe Piwowar for reviewing a draft of this paper.

REFERENCES

Cohen, J. (1960), A coefficient of agreement for nominal scales, *Educ. Psychol. Meas.* 20(1):37–46.

Cushnie, J. L. (1987), The interactive effect of spatial resolution and degree of internal variability within land cover types and classification accuracies, *Int. J. Remote Sens.* 8(1):15–29.

Dutra, L. V., and Mascarenhas, N. D. A. (1984), Some experiments with spatial feature extraction methods in multispectral classification, *Int. J. Remote Sens.* 5(2):303–313.

Egawa, H., and Kusaka, T. (1988), Region extraction in SPOT data, *Geocarto Int.* (3):25–30.

Eklundh, J. O., Yamamoto, H., and Rosenfeld, A. (1980), A relaxation method for multispectral pixel classification, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-2(1):72–75.

Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969), Large sample standard errors of Kappa and weighted Kappa, *Psychol. Bull.* 72(5):323–327.

Fu, K. S., and Yu, T. S. (1980), *Statistical Pattern Classification Using Contextual Information*, Research Studies Press, Chichester, England, 191 pp.

Gong, P., and Howarth, P. J. (1989), A modified probabilistic relaxation approach to land cover classification, in *Proceedings of IGARSS'89/12th Can. Symp. Remote Sens.*, Vancouver, BC, pp. 1621–1624.

Gong, P., and Howarth, P. J. (1990), The use of structural information for improving land-cover classification accuracies at the rural-urban fringe, *Photogramm. Eng. Remote Sens.*, forthcoming.

Gurney, C. M. (1981), The use of contextual information to improve land cover classification of digital remotely sensed data, *Int. J. Remote Sens.* 2(4):379–388.

Haralick, R. M. (1983), An interpretation for probabilistic relaxation, *Comput. Vision, Graphics, Image Process.* 22:388–395.

Haralick, R. M., and Joo, H. (1986), A contextual classifier, *IEEE Trans. Geosci. Remote Sens.* GE-24(6):997–1007.

Haralick, R. M., Mohammed, J. L., and Zucker, S. W. (1980), Compatibilities and the fixed points of arithmetic relaxation processes, *Comput. Graphics Image Process.* 13:242–256.

Haralick, R. M., Shanmugan, K., and Dinstein, I. (1973), Texture features for image classification, *IEEE Trans. Syst. Man. Cybern.* SMC-3(6):610–621.

Harris, R. (1985), Contextual classification post-processing of Landsat data using a probabilistic model, *Int. J. Remote Sens.* 6(6):847–866.

Hlavka, C. A. (1987), Land-use mapping using edge-density texture measures on thematic mapper simulator data, *IEEE Trans. Geosci. Remote Sens.* GE-24(6):104–108.

Howarth, P. J., Martin, L. R. G., Holder, G. H., Johnson, D. D., and Wang, J. (1988), SPOT imagery for detecting residential expansion on the rural-urban fringe of Toronto, Canada, in *SPOT 1: Image Utilization, Assessment, Results* (CNES, Eds.), Cepadues-Editions, Toulouse, France, pp. 491–498.

Hsu, S. (1978), Texture-tone analysis for automated landuse mapping, *Photogramm. Eng. Remote Sens.* 44(11): 1393–1404.

Irons, J. R., Markham, B. L., Nelson R. F., Toll, D. L., Williams, D. L., Latty, R. S., and Stauffer, M. L. (1985), The effects of spatial resolution on the classification of TM data, *Int. J. Remote Sens.* 6(8):1385–1403.

Jensen, J. R. (1979), Spectral and textural features to classify elusive land cover at the urban fringe, *Prof. Geogr.* 31:400–409.

Jensen, J. R. (1983), Urban/suburban land use analysis, in *Manual of Remote Sensing*, (R. N. Colwell, Ed.), American Society of Photogrammetry, Falls Church, VA, Chap. 30, pp. 1571–1666.

Johnson, D. D., and Howarth, P. J. (1987), The effects of spatial resolution on land cover/use theme extraction from airborne digital data, *Can. J. Remote Sens.* 13(2):68–74.

Kalayeh, H. M., and Landgrebe, D. A. (1984), Adaptive relaxation labeling, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6(3):369–372.

Ketting, R. J., and Landgrebe, D. A. (1976), Classification of multispectral image data by extraction and classification of homogeneous objects, *IEEE Trans. Geosci. Electron.* GE-14(1):19–26.

Landgrebe, D. A. (1980), The development of a spectral-spatial classifier for earth observation data, *Pattern Recognition* 12(3):165–175.

Lee, T., and Richards, J. A. (1989), Pixel relaxation labeling using a diminishing neighbourhood effect, in *Proceedings of IGARSS'89/12th Can. Symp. Remote Sens.*, Vancouver, BC, pp. 634–637.

Martin, L. R. G. (1989), Accuracy assessment of Landsat-based visual change detection methods applied to the rural-urban fringe, *Photogramm. Eng. Remote Sens.* 55(2):209–215.

Martin, L. R. G., Howarth, P. J., and Holder, G. H. (1988), Multispectral classification of land use at the rural-urban fringe using SPOT data, *Can. J. Remote Sens.*, 14(2):72–79.

Owen, A. (1984), A neighborhood-based classifier for Landsat data, *Can. J. Stat.* 12(3):191–200.

Peleg. S., and Rosenfeld, A. (1978), Determining compatibility coefficients for curve enhancement relaxation processes, *IEEE Trans. Syst. Man Cybern.* SMC-8(7):548–555.

Richards, J. A. (1986), *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, Berlin, 281 pp.

Richards, J. A., Landgrebe D. A., and Swain, P. H., (1981a), Pixel labeling by supervised probabilistic relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-3(2):188–191.

Richards, J. A., Landgrebe D. A., and Swain, P. H. (1981b), On the accuracy of pixel relaxation labeling, *IEEE Trans. Syst. Man Cybern.* SMC-11(4):303–309.

Richards, J. A., Landgrebe D. A., and Swain, P. H. (1982), A means for utilizing ancillary information in multispectral classification, *Remote Sens. Environ.* 12:463–477.

Rosenfeld, A., Hummel, R. A., and Zucker, S. W. (1976), Scene labeling by relaxation operations, *IEEE Trans. Syst. Man Cybern.* SMC-6(6):420–433.

Rosenfield, G. H., and Fitzpatrick-Lins, K. (1986), A coefficient of agreement as a measure of thematic classification accuracy, *Photogramm. Eng. Remote Sens.* 52(2):223–227.

Swain, P. H., Seigel, H. J., and Smith, B. W. (1980), Contextual classification of multispectral remote sensing data using a multiprocessor system, *IEEE Trans. Geosci. Remote Sens.* GE-18(2):197–203.

Toll, D. L. (1984), An evaluation of simulated TM data and Landsat MSS data for determining suburban and regional land use and land cover, *Photogramm. Eng. Remote Sens.* 50(12):1713–1724.

Townshend, F. E. (1986), The enhancement of computer classification by logical smoothing, *Photogramm. Eng. Remote Sens.* 52(2):213–221.

Welch, J. R., and Salter, K. G. (1971), A context algorithm for pattern recognition and image interpretation, *IEEE Trans. Syst. Man Cybern.* SMC-1(1):24–30.

Weszka, J. S., Dyer, C. R., and Rosenfeld, A. (1976), A comparative study of texture measures for terrain classification, *IEEE Trans. Syst. Man Cybern.* SMC-6(3):265–285.