Purdue University
Purdue e-Pubs

LARS Technical Reports

Laboratory for Applications of Remote Sensing

1-1-1982

Probabilistic Relaxation on Multitype Data

H. M. Kalayeh

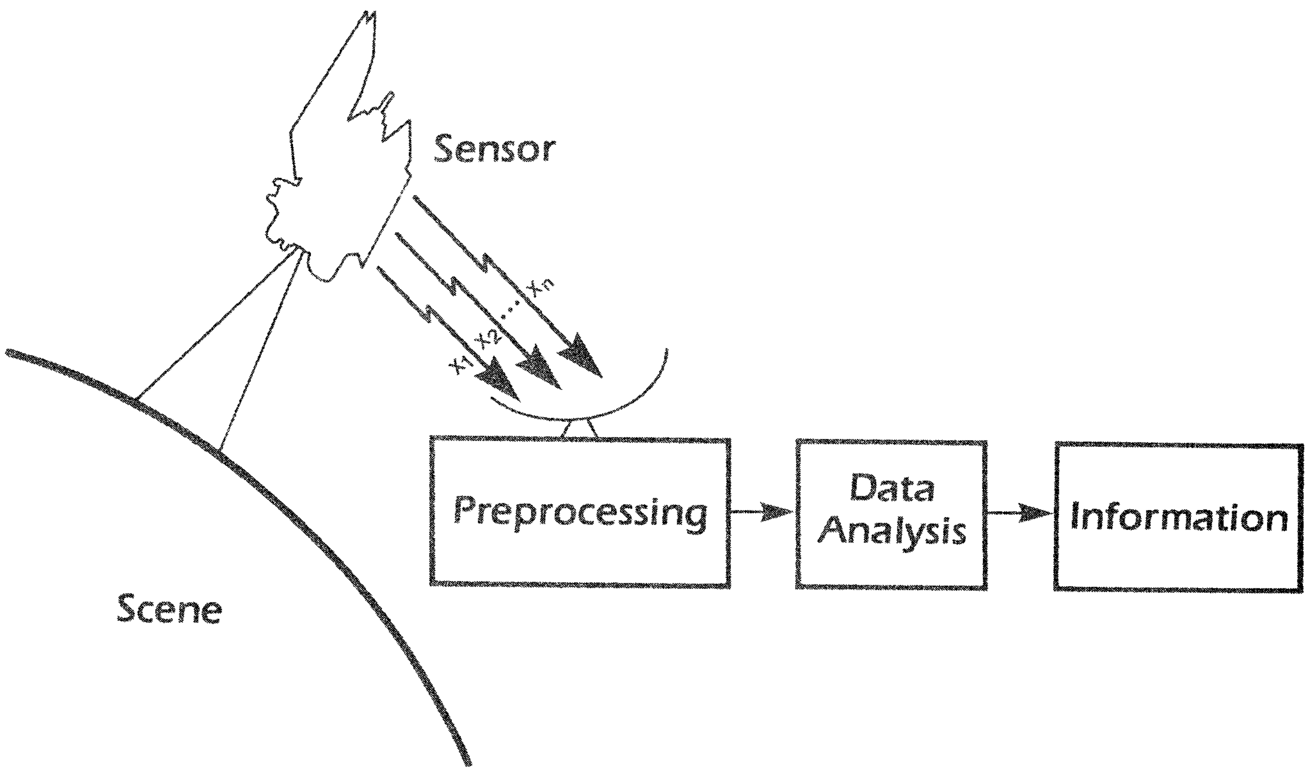
D. A. Landgrebe

Eighth International Symposium

Machine Processing of Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring



July 7-9, 1982

Reprints from Proceedings

Purdue University

Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

PROBABILISTIC RELAXATION ON MULTITYPE DATA

H.M. KALAYEH, D.A. LANDGREBE

Purdue University/Laboratory for
Applications of Remote Sensing
West Lafayette, Indiana

ABSTRACT

Classification of multispectral image data based on spectral information has been a common practice in the analysis of remote sensing data. However, the results produced by current classification algorithms necessarily contain residual inaccuracies and class ambiguity. By the use of other available sources of information, such as spatial, temporal and ancillary information, it is possible to reduce this class ambiguity and in the process improve the accuracy.

In this paper, the probabilistic and supervised relaxation techniques are adapted to the problem. The common probabilistic relaxation labeling algorithm (PRL), which in remote sensing pixel labeling usually converges toward accuracy deterioration, is modified. Experimental results show that the modified relaxation algorithm reduces the labeling error in the first few iterations, then converges to the achieved minimum error. Also a noniterative labeling algorithm which has a performance similar to that of the modified PRL is developed. Experimental results from Landsat and Skylab data are included.

I. INTRODUCTION

Our objective is to develop heuristic algorithms to utilize a combination of spectral, spatial, temporal and ancillary information. In remote sensing, the spectral variations of electromagnetic energy of the scene have been studied extensively. The spectral response, which is a function of wavelength, has been modeled as a random process(1,2). Another source of useful information is the spatial context of a pixel.

The information surrounding an object or pixel is referred to as contextual information. In many pattern recognition problems, there exist spatial characteristics which describe the spatial dependencies among the patterns to be recognized(3). Also, temporal variations in the scene and available ancillary data, such as topographic data, pixel radar response, and classification labeling maps, are known to be information-bearing(4).

II. PROBABILISTIC LABELING

Probabilistic labeling is a process of estimating the initial labeling probabilities. Let X be a point in q-dimensional measurement space containing m classes. Also assume that the probability density function associated with each class is Gaussian. Let $p(X|\omega_k), P(\omega_k)$ be the class-conditional density function and prior probability of the kth class, respectively. To characterize each class, the class mean vector and covariance matrix are estimated from training samples. Then pixel-label probabilities are estimated by calculating the a posteriori probabilities $P(\omega_k|X)$, as follows:

$$P_i^O(\omega_k) \triangleq P(\omega_k|X) = \frac{p(X|\omega_k)\ P(\omega_k)}{\sum_{\ell=1}^{m} p(X|\omega_\ell)\ P(\omega_\ell)} \quad (1)$$

$k = 1, 2, \ldots m$

where $P_i^O(\omega_k)$ is the initial estimate of probability of the ith pixel's label. However, if the initial labeling probabilities cannot be statistically estimated, then we may assign probabilities to the predetermined labels, as follows:

$$P_i^O(\omega_k) = W \quad (2)$$

$$P_i^0(\omega_\ell) = \frac{1-W}{m-1}, \quad \ell = 1, 2, \ldots m \quad (3)$$
$$\ell \neq k$$

where it is assumed that the ith pixel's label is ω_k and $\frac{1}{m} \leq W \leq 1$. This way of estimating the initial labeling probabilities will be referred to as weighting method.

III. UTILIZING SPECTRAL, SPATIAL CHARACTERISTICS BY PROBABILISTIC RELAXATION ALGORITHM

Relaxation labeling processes are an iterative heuristic approach which attempts to extract contextual information in a scene to reduce the ambiguity of a predetermined labeling. Relaxation labeling techniques use two source of information, an initial (ambiguous) labeling and information imbedded in spatial context of a pixel. A block diagram of a post classifier which utilizes probabilistic and supervised relaxation is given in Figure 1.

Let us consider the probabilistic relaxation algorithm which has been suggested by Zucker et al.(5). Let $P_i^n(\omega_k)$ denote the estimate of the probability that on the nth interation the label or class of the ith pixel of a scene is ω_k; k = 1, 2,...m. Then define

$$P_i^{n+1}(\omega_k) = \frac{P_i^n(\omega_k)\ Q_i^n(\omega_k)}{\sum_{\ell=1}^{m} P_i^n(\omega_\ell)\ Q_i^n(\omega_\ell)} \quad (4)$$

where $Q_i^n(\omega_k)$ is called the neighborhood function and is defined by

$$Q_i^n(\omega_k) = \sum_{j=1}^{J} d_{ij} \sum_{\ell=1}^{m} P_{ij}(\omega_k|\omega_\ell)\ P_j^n(\omega_\ell) \quad (5)$$

In this equation $P_{ij}(\omega_k|\omega_\ell)$ is the probability that pixel i is from class ω_k given that pixel j is from class ω_ℓ. The d_{ij} are a set of neighborhood weights which satisfy

$$\sum_{j=1}^{J} d_{ij} = 1 \quad (6)$$

with J as the number of pixels in the neighborhood and m as the number of classes. Examples of J = 5 and J = 9 are given in Figure 2. In all our analysis, the J = 5 neighborhood will be used.

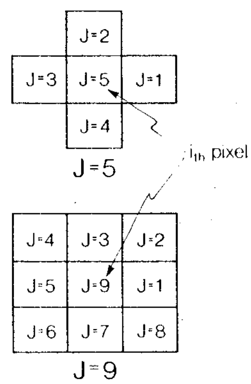


Figure 2. Examples of two different neighbor sets.

IV. UTILIZING SPECTRAL, SPATIAL, ANCILLARY INFORMATION BY SUPERVISED RELAXATION ALGORITHM

The supervised relaxation processes(6,7) are a more general version of probabilistic relaxation methods which attempt to utilize multi-type data characteristics. In the supervised relaxation, first an appropriate likelihood for the label of each pixel is estimated based on the statistical information of available ancillary data. Then the neighborhood function for the label most favored by

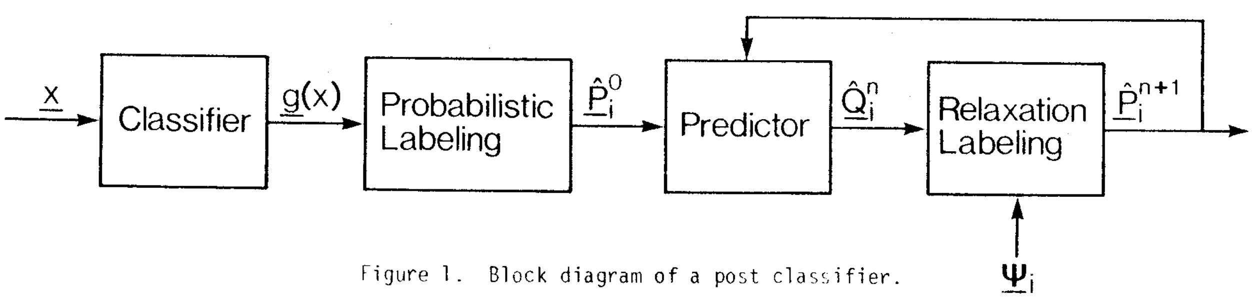


Figure 1. Block diagram of a post classifier.

ancillary data is increased and others decreased in proportion to their support from the ancillary data source. The relaxation algorithm does not know, of course, which are the correct and which are the incorrect labels. It only "knows" which labels are consistent with their neighbors and with the ancillary data. Consequently, an image with initial labeling errors will be iterated until consistency between spectral, spatial and ancillary information is achieved.

Let us consider the supervised relaxation algorithm which is suggested by Richards et al. in (6,7):

$$P_i^{n+1}(\omega_k) = \frac{P_i^n(\omega_k)\ R_i^n(\omega_k)}{\sum_{\ell=1}^{m} P_i^n(\omega_\ell)\ R_i^n(\omega_\ell)} \tag{7}$$

where

$$R_i^n(\omega_k) = Q_i^n(\omega_k)\Psi_i(\omega_k) \tag{8}$$

$$Q_i^n(\omega_k) = \sum_{j=1}^{J} d_{ij} \sum_{\ell=1}^{m} P_{ij}(\omega_k|\omega_\ell)\ P_j^n(\omega_\ell) \tag{9}$$

and

$$\Psi_i(\omega_k) = [1+\beta\ (m\phi_i(\omega_k) - 1)] \tag{10}$$

In the above equations $P_i^n(\omega_k), Q_i^n(\omega_k)$ are the same as we defined earlier and $\Psi_i(\omega_k)$ is an estimate of the likelihood for the ith pixel's label on basis of ancillary data. In the (Eq. 10), $\phi_i(\omega_k)$ is the probability that ith pixel belongs to class ω_k or its label is ω_k, β is a parameter that adjusts the degree of supervision; it is between zero and one. The parameter β is chosen heuristically; however it should reflect our confidence in the ancillary data with comparison to the other sources of information. As before, m is the number of possible classes or labels.

V. PROPOSED ALGORITHMS FOR UTILIZING MULTITYPE INFORMATION

The spatial context of a pixel or dependency among the labels in a neighborhood is incorporated via $P_{ij}(\omega_k|\omega_\ell)$, the transition probability that pixel i is from class ω_k given that pixel j (one of its neighbors) is from ω_ℓ. In practice, $P_{ij}(\omega_k|\omega_\ell)$ is estimated from the result of probabilistic labeling over the whole data set, which means the transition probabilities are assumed constant over the data set. In fact, in an actual data set, they may be expected to vary from place to place. What we are suggesting is, $P_{ij}(\omega_k|\omega_\ell)$ should slowly vary over the data set and the following procedure is suggested to estimate these transition probabilities.

1. Depending on the number of classes, choose a square window of size L x L centered at the ith pixel. For example, for two classes, we have chosen a window of size 5 x 5 and for the tree classes a window of size 6 x 6 may be considered.

2. Estimate the probability of jth pixel's label by

$$P_j(\omega_k) = \frac{1}{L^2} \sum_{r=1}^{L^2} P_{jr}^o(\omega_k) \tag{11}$$

where $P_{jr}(\omega_k)$ is the initial estimate of a pixel's label at location jr of the chosen window.

3. Estimate the transition probability by

$$P_{ij}(\omega_k|\omega_\ell) = \frac{P_{ij}(\omega_k,\omega_\ell)}{P_j(\omega_\ell)} \tag{12}$$

and the joint probability by

$$P_{ij}(\omega_k,\omega_\ell) = \frac{1}{(L-1)^2} \sum_{r=1}^{(L-1)^2} P_r^o(\omega_k)\,[\tfrac{1}{4} \sum_{j=1}^{4} P_{rj}^o(\omega_\ell)] \tag{13}$$

where $P_r^o(\omega_k)$ is the initial estimate of rth pixel surrounding the ith pixel and including ith pixel itself. And $P_{rj}^o(\omega_\ell)$ is the initial estimate of jth pixel surrounding rth pixel but excluding it.

Now, by using this adaptive procedure, the spatial context of each pixel is estimated and incorporated by the neighborhood function to predict the estimate of the probability of each pixel's label.

It is believed this simple algorithm can extract most of the contextual information by only one iteration. The adaptive labeling algorithm is given by:

$$Q_i^n(\omega_k) = \sum_{j=1}^{J} d_{ij} \sum_{\ell=1}^{m} P_{ij}(\omega_k|\omega_\ell) P_j^n(\omega_\ell) \qquad (14)$$

Let $d_{ij} = \frac{1-d_i}{J-1}$, then it can be shown that

$$Q_i^n(\omega_k) = q_i^n(\omega_k) + d_i [P_i^n(\omega_k) - q_i^n(\omega_k)] \qquad (15)$$

where

$$q_i^n(\omega_k) = \sum_\ell P_{ij}(\omega_k|\omega_\ell) \left[\frac{1}{J-1}\sum_{j=1}^{J-1} P_j^n(\omega_\ell)\right] \qquad (16)$$

and

$$P_i^{n+1}(\omega_k) = q_i^n(\omega_k) + d_i\left[P_i^n(\omega_k) - q_i^n(\omega_k)\right] \qquad (17)$$

The new formulation of the probabilistic relaxation will therefore be

$$P_i^{n+1}(\omega_k) = \qquad (18)$$

$$\frac{d_i\left[P_i^n(\omega_k)\right]^2 + (1-d_i) P_i^n(\omega_k) q_i^n(\omega_k)}{\sum_{\ell=1}^{m} d_i\left[P_i^n(\omega_\ell)\right]^2 + (1-d_i) P_i^n(\omega_\ell) q_i^n(\omega_\ell)}$$

In Eq. 17, if we let $d_i = 1-\gamma_i$, then we can write

$$P_i^{n+1}(\omega_k) = P_i^n(\omega_k) + \gamma_i [q_i^n(\omega_k) - P_i^n(\omega_k)] \qquad (19)$$

A summary of all the algorithms is given in Table 1.

In the above algorithms, if $d_i = 0.0$, then the label of the ith pixel will be decided, based on spatial information (assuming its initial label probability is not zero or one). If $d_i = 1.0$, then we are not using any spatial information for the ith pixel.

Table 1. Summary of Probabilistic and Supervised Relaxation Algorithms.

Algorithm 1
Probabilistic Relaxation Labeling (PRL)

$$P_i^{n+1}(\omega_k) = \frac{P_i^n(\omega_k) Q_i^n(\omega_k)}{\sum_{k=1}^{m} P_i^n(\omega_k) Q_i^n(\omega_k)}$$

$$Q_i^n(\omega_k) = \sum_{j=1}^{J} d_{ij} \sum_\ell P_{ij}(\omega_k|\omega_\ell) P_j^n(\omega_\ell)$$

Initial Labeling Probability $\left\{\begin{array}{l}\text{Weighting method} \\ \text{Probabilistic labeling}\end{array}\right.$

Transition Probability $\left\{\begin{array}{l}\text{Over the region} \\ \text{Window}\end{array}\right.$

Algorithm 2
Iterative Adaptive Labeling (IAL)

$$\left\{\begin{array}{l} P_i^{n+1}(\omega_k) = P_i^n(\omega_k) + \gamma_i [q_i^n(\omega_k) - P_i^n(\omega_k)] \\ P_i^{n+1}(\omega_k) = q_i^n(\omega_k) + (1-\gamma_i) [P_i^n(\omega_k) - q_i^n(\omega_k)] \end{array}\right.$$

$$q_i^n(\omega_k) = \sum_\ell P_{ij}(\omega_k|\omega_\ell) [\frac{1}{J-1}\sum_{j=1}^{J-1} P_j^n(\omega_\ell)]$$

$$0 \le \gamma_i \le 1$$

Initial Labeling Probability $\left\{\begin{array}{l}\text{Weighting Method} \\ \text{Probabilistic labeling}\end{array}\right.$

Transition Probability $\left\{\begin{array}{l}\text{Over the region} \\ \text{Window}\end{array}\right.$

Algorithm 3
Non-Iterative Adaptive Labeling (NAL)

The same as Algorithm 2 with only one iteration.

Supervised Relaxation Algorithms

The supervised version of algorithms 1, 2, and 3 will be referred to as algorithms 4 5, and 6, respectively.

As mentioned in Section IV, the supervised probabilistic relaxation algorithms are heuristic techniques which attempt to reduce the ambiguity of a predetermined labeling by measuring consistency of pixel labels based on multitype data characteristics. Labeling consistency is measured by multiplying appropriate label likelihoods, which can be obtained from spectral, spatial and ancillary information. In our analysis, the following ancillary information was utilized:

1. Classification of an image based on elevation data carries information about some main geometric features; therefore, if we constantly remind the relaxation process about these features, then the algorithm will become more intelligent.

2. Also a linear classifier can recognize some geometric features, so its labeling results can be used to supervise the relaxation to recognize other geometric features.

3. The results of classification based on temporal information, for example at time t, can be used to supervise relaxation labeling at time t-1 or vice versa.

VI. EXPERIMENTAL RESULTS

In order to experimentally evaluate the performance of the above heuristic algorithms, two data sets were selected. Data Set 1 was multitemporal spatially registered Landsat MSS data acquired over Henry County, Indiana, in 1978. Data Set 2 was multispectral Skylab S-192 data from northeast of the Vallecito Reservoir region in the Colorado Rockies. This data set was classified into a number of tree species using the maximum likelihood classifier. The classification map so produced was rearranged for simplicity into the two categories of spruce fir and others. For the region elevation, data as well as a probability model for the occurrence of spruce fir vs. elevation were chosen as an ancillary data variable.

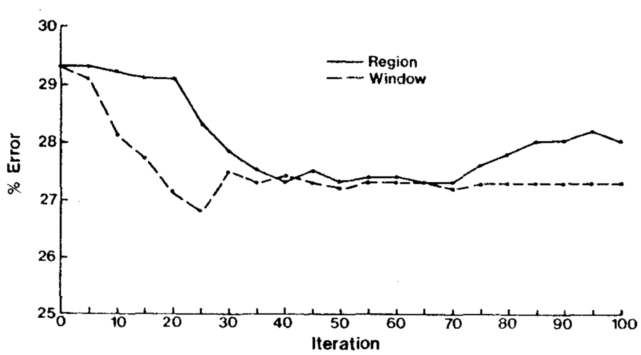


Figure 3a. Comparison of the performance of Algorithm 1 (PRL) with two different ways of estimating the transition probability.

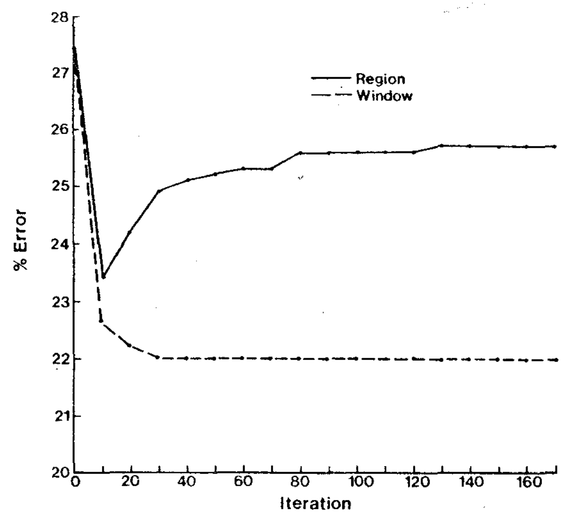


Figure 3b. Comparison of the performance of Algorithm 1 (PRL) with two different ways of estimating the transition probability.

A block of size 40x30 pixels from Data Set 1 collected on August 20, 1978 and a block of size 30x30 pixels from Data Set 1 collected on September 26, 1978 were chosen. The initial labeling probabilities of each block for two labels, corn/soybean and others, were computed. Then the performance of the relaxation labeling algorithm with two different ways of estimating the transition probability (d_i = 0.1) was evaluated. The results are given in Figures 3a and 3b. These results suggest that Algorithm 1 (probabilistic relaxation labeling) with adaptively estimating the transition probability, does not exhibit any deterioration in accuracy.

The performance of iterative relaxation labeling and noniterative algorithms which were applied to a block of 30x30 pixels from Data Set 1 collected on August 20 and September 26 are shown in Figures 4a and 4b. These figures show that the performance of Algorithm 3 (non-iterative adaptive labeling) and Algorithm 1 are almost the same.

The performance of relaxation labeling supervised by temporal, spectral, and elevation data (shown in Figures 5a, 5b, and 5c, respectively) was evaluated by the following experiments:

Experiment 1:

A block of size 30x30 pixels from multitemporal Data Set 1 collected on August 20, 1978 (time t1) and September 26, 1978 (time t2) was chosen. The initial labeling probabilities at times t1 and t2 were estimated by the maximum likelihood method and the transition probabilities were estimated over a window of size 5x5 pixels. Algorithm 1 (PRL) with $d_i=1-\gamma_i=0.1$ was compared to Algorithm 4 (with $d_i=0.0$ and $\beta=0.5$). Information at time t2 was used as ancillary information to supervise Algorithm 1. The objective of this experiment was to preserve some geometric features and therefore improve the performance of Algorithm 1. The results are given in Figure 5a.

Experiment 2:

The objective of this experiment was to improve the performance of Algorithm 1 and Algorithm 2 by supervising them by labeling results of a linear classifier. A block of 40x30 pixels from Data Set 1 was chosen. Then the performances of Algorithms 1, 4, and 5 in estimating initial labeling probabilities by weighting

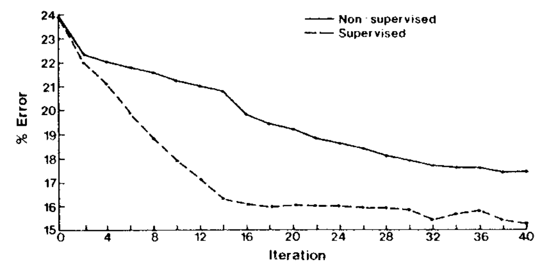


Figure 5a. Comparison of the performance of the supervised and nonsupervised relaxation algorithms. The classification results at a different time were used as ancillary information.

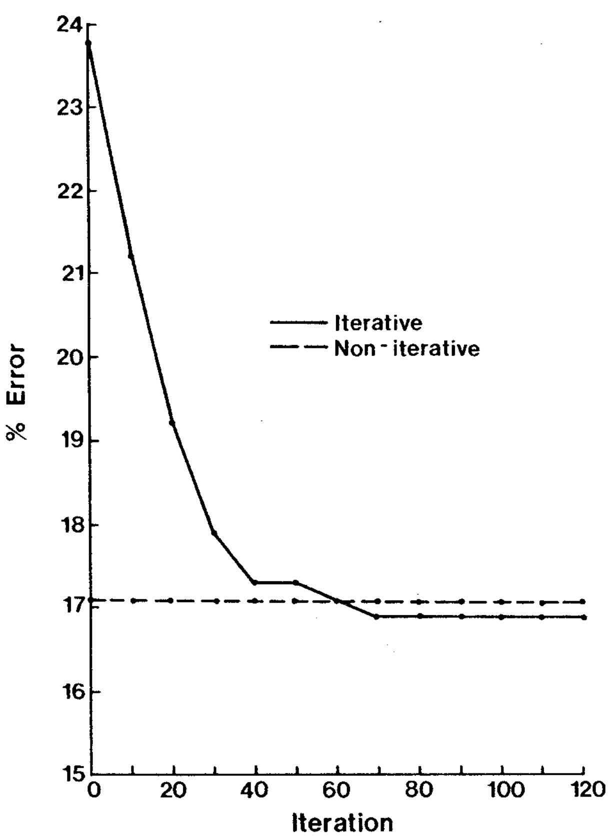


Figure 4a. Comparison of Algorithm 1 (PRL) and Algorithm 3 (NAL).

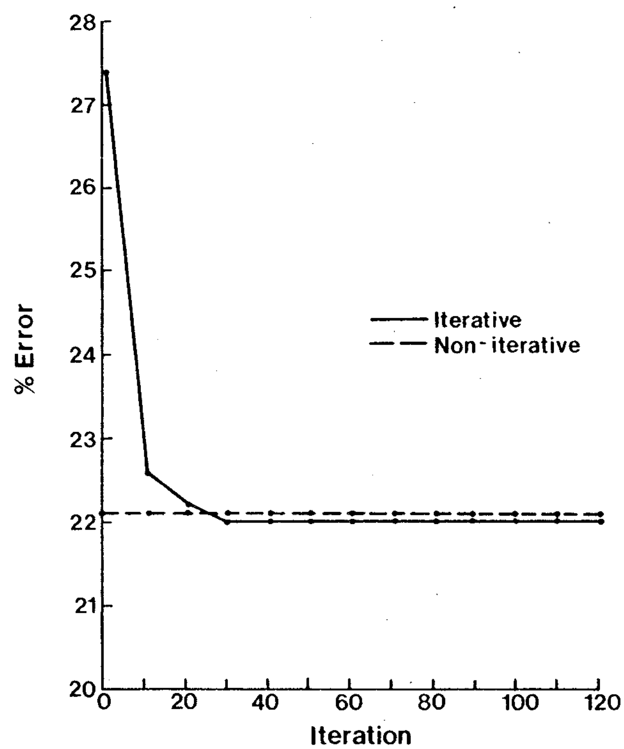


Figure 4b. Comparison of Algorithm 1 (PRL) and Algorithm 3 (NAL).

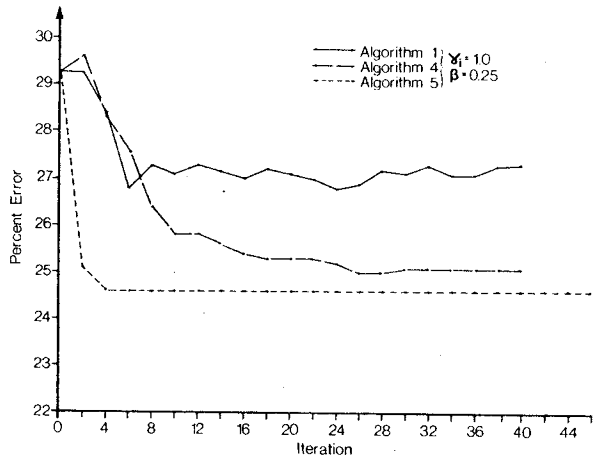


Figure 5b. Comparison of the performance of Algorithms 2, 4, and 5. The classification results by the minimum distance processor were used as ancillary information.

method and estimating the transition probabilities over the chosen block, $d_i=0$ and $\beta=0.25$, were evaluated. The results are given in Figure 5b. The results show that Algorithm 5 has a better performance than Algorithms 1 and 4. Also Algorithm 5 reaches its fixed point or steady state in few iterations.

Experiment 3:

A block of size 129x91 pixels from Data Set 2 was chosen. The accuracy of the labeling was measured by using 88 pixels whose correct labeling was known. Then the performances of Algorithms 1, 3, 4, and 6 in estimating initial probabilities by weighting method and estimating the transition probabilities over the whole region were compared. The results are shown in Figure 5c and suggest the use of a supervised non-iterative approach for reduction of the labeling ambiguity.

VII. CONCLUSION

The probabilistic relaxation technique suggested by Zucker et al.(5) is applied to the remote sensing data as a post classifier. However, the suggested algorithm usually decreases the labeling error (improving phase) passes through a turning point and increases the labeling error (deterioration phase). We have modified the algorithm by assuming that the transition probabilities are slowly varying over the scene and a method to estimate the transition probabilities has been suggested. The experimental results suggest that the modified algorithm does not exhibit a deterioration phase anymore. Also, a non-iterative adaptive labeling algorithm has been developed which performs as well as the modified probabilistic relaxation algorithm. In addition, in order to be able to preserve the geometric features, supervised relaxation labeling was developed. By supervising the process by the available ancillary information, we indeed incorporate "memory" into the labeling process to constantly remind the algorithm about some geometric features which are strongly supported by ancillary information. Finally, it has been shown that by utilizing spectral, spatial and ancillary data, the initial labeling accuracy can be improved.

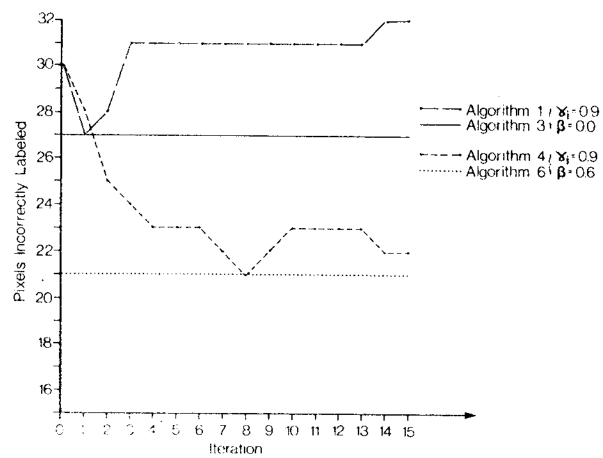


Figure 5c. Comparison of the performance of Algorithms 1, 3, 4, and 6. The elevation data were used as ancillary information.

REFERENCES

1. P.H. Swain and S.M. Davis, eds. Remote Sensing: The Quantitative Approach. McGraw-Hill, New York, 1978.

2. A. Papoulis. Probability Random Variables and Stochastic Processes. McGraw-Hill, New York, 1965.

3. G.T. Toussaint. "The Use of Context in Pattern Recognition. Pattern Recognition, Vol. 10, 1978, pp. 189-204.

4. D.A. Landgrebe, "Analysis Technology for Land Remote Sensing," Proc. IEEE, Vol. 69, No. 5, May 1981.

5. S. Zucker and J. Mohammed. "Analysis of Probabilistic Relaxation Labeling Processes," Proc. IEEE Conf. Pattern Recognition and Image Processing, Chicago, IL, 1978, pp. 307-312.

6. J.A. Richards, D.A. Landgrebe and P.H. Swain, "Pixel Labeling by Supervised Probabilistic Relaxation," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No. 2; (also LARS Technical Report 022580), Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, IN 47906-1399, Feb. 1980.

7. J.A. Richards, D.A. Landgrebe and P.H. Swain, "Supervised Pixel Relaxation Labeling as a Means for Utilizing Ancillary Information in the Classification of Remote Sensing Image Data,"(Submitted for publication in Remote Sensing for Environment.)

HOOSHMAND MAHMOOD KALAYEH is graduate research assistant at the Laboratory for Applications of Remote Sensing (LARS), Purdue University and a candidate for the Ph.D. degree in Electrical Engineering at Purdue. He received his M.S.E.E. from Wayne State University in 1978. A native of Tehran, Iran, he received his B.S.E.E. from Iran College of Science and Technology (ICST) in 1973. From 1973 to 1977 he served as an instructor at ICST. Mr. Kalayeh is a member of IEEE, Tau Beta Pi, and the Purdue Student Society of Professional Engineers.

DAVID ALLEN LANDGREBE is Associate Dean of Engineering at Purdue University and Director of the Engineering Experiment Station. He holds B.S.E.E., M.S.E.E., and Ph.D. degrees from Purdue. He joined the Purdue EE faculty in 1962. He was named Program Leader for the Data Processing program at LARS in 1966 and served as director of LARS from 1969 to 1981. He received the NASA Exceptional Scientific Achievement Medal in 1973. Dr. Landgrebe is a fellow of IEEE and a member of several professional and honorary organizations. He also is an associate editor of the journal Remote Sensing of the Environment, and a member of the administrative committee of the IEEE Geoscience and Remote Sensing Society.