Multispectral Image Context Classification Using Stochastic Relaxation

MING CHUAN ZHANG, ROBERT M. HARALICK, FELLOW, IEEE, AND JAMES B. CAMPBELL

Abstract—**A new multispectral image context classification, which is based on a stochastic relaxation algorithm and Markov–Gibbs random field, is presented. The implementation of the relaxation algorithm is related to a form of optimization programming using annealing. The authors motivate a Bayesian context decision rule, and a Markov–Gibbs model for the original Landsat MSS (multispectral scanner) image is introduced, and then develop a new contextual classification algorithm, in which maximizing the posterior probability (MAP) is based on stochastic relaxation, an annealing optimization method. Finally, experimental results that are based on simulated and real multispectral remote sensing images to show how classification accuracy is greatly improved are presented. The algorithm is highly parallel and exploits the equivalence between Gibbs distributions and Markov random fields (MRF).**

I. Introduction

CONVENTIONAL automatic classification techniques, in particular for remote sensing data, classify each pixel independently. This type of classification can only exploit spectral or, in some cases, spectral and temporal information. Using coherent spatial information for classification efficiency and accuracy in remote sensing application has long been desired. In recent years some researchers have discussed this realization. A spatial stochastic recursive contextual classification was proposed by T. S. Yu and K. S. Fu [1]; an estimation method of the context function was discussed by J. S. Tilton, S. B. Vardenman, and P. H. Swain [2]; a recursive context classification using dynamic programming was presented by R. M. Haralick, M. C. Zhang, and J. B. Campbell [3].

In statistical pattern recognition, there appears to be two main approaches to the specification of spatial stochastic processes. Whittle [4] proposed a random field model that arises from the joint probability distribution of the variables in a neighborhood. Whittle's definition requires that the joint probability distribution of the variables in a given neighborhood be of the product form

$$P(D_{ij}) = \prod_{ij} Q_{ij}(D_{ij}) \tag{1}$$

where D_{ij} is a set of random variables within the neighborhood of the pixel (i, j) and Q_{ij} is a nonnegative function.

On the other hand Bartlett [5]–[7] proposed a model that arises from the conditional probability distribution of D_{ij}. His definition requires that the conditional probability distribution of D_{ij} depends only upon the values at the neighbors of (i, j).

Most of the existing methods [1]–[3] are based on the model that arises from the conditional probability distribution of D_{ij}.

Besag [8] found that constraints on the conditional probability structure are so severe that they actually dictate particular models.

Under the stochastic model for spatially oriented pixels, image correlations exist between any pair of pixels. Yu and Fu [1] noted that the best way to incorporate these correlations statistically is to consider the joint probability density function of all the site-variables involved. For example a five dimensional joint probability density function will account for all the correlation between any pair of sites in a four-neighbor system. Similarly, a nine-dimensional joint density function can be used for the eight-neighbor system. For practical reasons most of the estimated context decision rules deal with maximizing the conditional probability distribution of pixel (i, j), given all values within the nearest neighborhood of pixel (i, j).

Besag [8] argues that the conditional probability model has a number of disadvantages. First, there is no obvious method of deducing the joint probability structure associated with a conditional probability model. Secondly, the conditional probability structure itself is subject to some unobvious and highly restrictive consistency conditions. Third, it has been remarked by Whittle [4] that the natural specification of an equilibrium process in statistical mechanics is in terms of the joint distribution rather than the conditional distribution of their variables. Similarly the most natural and important quantity in evaluation of the discriminant function in contextual classification is the joint probability density function [1]. The conditional probability approach, however, has served as the basis for a commonly used class of models—the Markov image models.

In this paper we develop an alternative context classification approach, which is based on a stochastic relaxation algorithm and Gibbs distribution. The favorable features

Manuscript received September 26, 1988; revised July 28, 1989.

M. Zhang is with the Dept. of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28215.

R. M. Haralick is with the Dept. of Electrical Computer Engineering, University of Washington, Seattle, WA 98195.

J. B. Campbell is with the Dept. of Geography, Virginia Polytechnic and State University, Blacksburg, VA 34061.

IEEE Log Number 8931981.

of this approach are that its random model arises from the joint probability distribution of the variates in a neighborhood, and that the algorithm is highly parallel. The parallel algorithm, which is performed by a neighborhood operator, might be implementable in special purpose VLSI hardware.

The stochastic relaxation methods are not new concepts: they have been used in statistical physics for many years. There the problem of analyzing the macroscopic properties of a physical system is translated into one of analyzing the global properties of random fields with a given local structure. However only Geman and Geman [9] have introduced these concepts into image restoration, and they have given a few simple results using synthetic images. In this paper we use them in context classification, and we make an analogy between image and statistical mechanics systems. Pixel gray levels and category labels are viewed as states of atoms or molecules in a lattice-like physical system. The assignment of energy functions in a physical system determines its Gibbs distribution. Because of the Gibbs distribution–Markov random field (MRF) equivalence, this assignment also determines a MRF image model.

In this paper, we first motivate a Bayesian context-decision rule, and then we use a Markov–Gibbs model to develop a new contextual classification algorithm, in which maximizing the *a posteriori* probability (MAP) is based on stochastic relaxation, an annealing optimization method. Finally we present experimental results with both simulated and real multispectral remote sensing data to show how classification accuracy is greatly improved.

II. Motivation and Proposed Approach

The contextual information, which we would like to study, is a form of correlation existing among the successive pattern classes in the two-dimensional (2-D) image. Every pixel in the image can be considered as having one random variable associated with a 2-D Markov random field. Two pixels in spatial proximity to one another are unconditionally correlated, with the degree of correlation decreasing as the distance between them increases. All the spatial correlations among "site-variables" on a lattice of an image can be extracted by the specified spatial process. As mentioned earlier the most important quantity in the contextual Bayes' decision problem is the joint density function of all the site-variables within the specified contextual neighborhood. So the best way to incorporate these correlations statistically is to estimate the joint probability density function of all the site-variables involved.

For pratical reasons, most of previous studies deal with some specific cases (in which the four-neighborhood assumption is invariably utilized in the context algorithms) and the contextual information is incorporated by considering the conditional probabilities of pixel (i, j), given its neighbors.

These approaches are certainly based on a realistic premise but it is computationally feasible only for first order neighborhoods. The new contextual decision rule introduced in this section improves this by considering the joint probability of the pixels in the neighborhood of the pixel (i, j) and by using a larger context.

Before presenting this rule, we must first give some notational conventions, and assume that each pixel of the multiband image considered in the paper has a N-tuple of finite gray-tone values.

A. Notation

In order to have a precise framework within which we can describe the stochastic relaxation algorithm, we need some notational conventions.

1) I designates the row index set of an image $I = \{1, \cdots, M_r\}$.
2) J designates the column index set of an image $J = \{1, \cdots, M_c\}$.
3) N_{ij} designates a neighborhood of pixel (i, j).
4) D_{ij} the collection of all measurement vectors in the neighborhood N_{ij} of pixel (i, j).
5) C assigned category labels in the neighborhood.
6) C° assigned category labels in the neighborhood, excluding the central pixel of the neighborhood.
7) $2N_r + 1$ designates the number of rows in a neighborhood.
8) $2N_c + 1$ designates the number of columns in a neighborhood.
9) L_r designates the local row index set of the neighborhood
$$L_r = \{-N_r, \cdots, N_r\}.$$
10) K_c designates the local column index set of the neighborhood
$$K_c = \{-N_c, \cdots, N_c\}.$$
11) (l, k) designates a local position in neighborhood
$$(l, k) \in L_r X K_c.$$
12) d_{lk} an observed measurement vector from pixel (l, k).
13) C_{lk} assigned category label of pixel (l, k) in the neighborhood.
14) Ω set of all possible categories.
15) θ designates a pattern configuration of assigned labels in neighborhood.
16) Q set of all possible pattern configuration of assigned labels in neighborhood
$$Q = \{\theta_1, \theta_2, \cdots, \theta_m\},$$
m is total number of pattern configuration of assigned labels in neighborhood.

17) N designates a neighborhood system on the two dimensional finite set of $I \times J$.

18) L designates a clique in the neighborhood of pixel (i, j).

19) W designates family of cliques in the neighborhood of pixel (i, j). $W = \{L_1, L_2, \cdots, L_k\}$, k is total number of cliques in N_{ij}.

20) V_L designates a potential associated with clique L

$$V_L\colon R^{|L|} \to [0, \infty) \qquad L \in W.$$

21) $U(D_{ij})$ designates an energy function associated with Gibbs distribution $P(D_{ij})$.

22) $U(C)$ designates an energy function associated with Gibbs distribution $P(C)$.

23) $U(D_{ij}, C)$ designates an energy function associated with Gibbs distribution $P(D_{ij}, C)$.

B. Bayesian Context Classification Model

From the Bayesian Model [10], the context classification problem can be stated as follows: to assign labels C to the pixels in the neighborhood of pixel (i, j) that minimizes the expected loss

$$\sum_{C^*} \text{los}(C, C^*) P(C^* | D_{ij}, Q) \tag{2}$$

where $P(C^*|D_{ij}, Q)$ is the probability that the true labeling of the pixels is C^* given: 1) the measurements D_{ij} of the pixels in the neighborhood of pixel (i, j), 2) the prior information Q we have about pixel dependencies, and where $\text{los}(C, C^*)$ is the loss incurred for the assignment of interpretation C to the pixels in the neighborhood of pixel (i, j), when the true interpretation is C^*.

We use the most common zero-one loss function for our study problem. There is no loss for a correct joint assignment and unit loss for any incorrect joint assignment. Here, correct assignment means that each pixel in the neighborhood is assigned correctly. Thus there is no distinction in loss between an incorrect joint assignment in which only one pixel is incorrectly assigned or an incorrect assignment in which all pixels are incorrectly assigned.

Such a loss is defined by

$$\text{los}(C, C^*) = \begin{cases} 0, & \text{where } C = C^* \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

There are two assumptions about the world and pixel measurement process that can simplify the expected loss expression (2).

The first assumption states that the description process is local. When the pixel (i, j) is being examined, no characteristics from any other pixel but pixel (i, j) affect the description obtained from pixel (i, j). Hence

$$P(D_{ij} | C^*, Q) = \prod_{(l,k) \in N_{ij}} P(d_{lk} | C^*, Q). \tag{4}$$

The second assumption states that the n-tuple measurement of pixel (i, j) depends only upon the true interpretation c^*_{ij} associated with pixel (i, j) and does not depend upon any relationships pixel (i, j) may have with other units or upon the interpretation associated with any other pixel. Hence

$$P(d_{ij} | C^*, Q) = P(d_{ij} | c^*_{ij}), \qquad (i, j) \in \Omega. \tag{5}$$

Under these assumptions, the optimal decision rule determines interpretation C for the pixels in the neighborhood that minimize

$$\sum_{C^*} \text{los}(C, C^*) \prod_{(l,k) \in N_{ij}} P(d_{lk} | c^*_{lk}) P(C^*) / P(D_{ij}). \tag{6}$$

With the loss function defined by (3), the best decision procedure chooses interpretation C that satisfies the maximality condition

$$\prod_{(l,k) \in N_{ij}} P(d_{lk} | C_{lk}) P(C) \geq \prod_{(l,k) \in N_{ij}} P(d_{lk} | Z_{lk}) P(Z), \quad \text{for all } Z \in \Omega \tag{7}$$

The choice of C satisfying this maximality condition cannot be independently done pixel by pixel.

III. Markov Random Fields with Nearest Neighbor Assumption

It is clear that any efficient computer algorithm for image analysis, classification, and processing can only be done using the framework of a proper image model. The Markov random field (MRF) and Gibbs model, which is pervasive in the image processing literature, constitutes a promising natural way to capture context assumptions in classification.

Consistent with the 2-D discrete MRF for multispectral image processing applications, we assume a random observation vector d_{ij}, and the pixel position (i, j) is defined on the 2-D finite integer set of $I \times J$.

The MRF model may be defined as the following. Let $\{d_{ij} | (i, j) \in I \times J\}$ be an observed image, and N_{ij} be the appropriate symmetric neighbor set of the pixel (i, j). It is postulated that this is generated by an appropriate 2-D (noncausal) MRF model. The model characterizes the statistical dependency among pixels by requiring that

$$P(d_{ij} | d_{mn}\colon (m, n) \in I \times J, (m, n) \neq (i, j)) = P(d_{ij} | d_{mn}\colon (m, n) \in N_{ij}). \tag{8}$$

If $N_{ij} = \{(0,1), (0,-1), (-1,0), (1,0)\}$ it corresponds to taking the simplest Markov model. By including more neighbors we can construct a higher order Markov model.

Unlike the 1-D discrete time series, where the existence of a preferred direction is inherently assumed, no such preferred ordering of the discrete neighborhood is appropriate for 2-D. In other words, the notion of "past" and "future" as understood in a unilateral 1-D Markov process is restrictive in 2-D as it implies a particular ordering in which the observations are scanned top down and left to right. It is quite possible that an observation at a pixel p

may be dependent on surrounding observations in all directions.

An alternative representation of random fields is by the Gibbs distribution. We say that a random field has a Gibbs distribution if its density function is given by

$$P(D_{ij}) = \frac{1}{Z} e^{\frac{-U(D_{ij})}{KT}} \tag{9}$$

where K is a Boltzmann's constant, T is temperature, and $U(D_{ij})$ is called the energy. Z is the normalizing constant as

$$Z = \sum_{D_{ij}} e^{\frac{-U(D_{ij})}{KT}}. \tag{10}$$

The energy $U(D_{ij})$ is the sum of local potentials $V_L(D_{ij})$ such that

$$U(D_{ij}) = -\sum_{L \in W} V_L(D_{ij}). \tag{11}$$

The $V_L(D_{ij})$ are called local potentials, which are evaluated over each clique on the neighborhood of pixel (i, j), where a clique is a subset in which each pixel is a neighbor of all pixels in the subset L. The W_{ij} designates the family of cliques in the neighborhood of pixel (i, j).

A. Markov Properties

Before describing the relaxation algorithm, it is first necessary to define neighborhood system and MRF, and then to discuss the factorizability property characteristic to Gibbs states with nearest neighbor potentials.

1) Definition 1: A collection of subsets of $I \times J$ described as $N_{ij} \in I \times J$ is a neighborhood system over the 2-D finite set of $I \times J$, if and only if N_{ij}, the neighborhood of pixel (i, j) is such that: 1) (i, j) does not belong to N_{ij} and 2) if $(k, l) \in N_{ij}$, then $(i, j) \in N_{kl}$, for any $(i, j) \in I \times J$, if $(k, l) \in N_{ij}$, (k, l) is said to neighbor (i, j).

We can now formally define a MRF with respect to the neighborhood system N defined over the 2-D finite set of $I \times J$.

2) Definition 2: Suppose N is a neighborhood system defined over the two-dimensional finite set of $I \times J$. A random field $\{d_{ij}: (i, j) \in I \times J\}$ is an MRF with respect to the neighborhood system N if and only if

$$P(d_{ij} | d_{lk}, (l, k) \in I \times J, (k, l) \neq (i, j)) = P(d_{ij} | d_{lk}, (k,) \in N_{ij}), \quad \text{for all } (i, j) \in I \times J. \tag{12}$$

3) Factorizability Property: A set of pixels L is a clique if every pair of pixels in L are neighbors. We include the empty set (no pixel) as a clique. If we considered only the empty set (no pixel), single pixel, and cliques consisting of pairs of pixels, it is called a first order model. If we also consider cliques consisting or triples of pixels, it is called a second-order model.

The factorizability property is the theoretical base of decomposing the potential functions by cliques (see formula 11). It is very useful for finding the canonical potential form in our problem.

The factorization property can be stated as follows. Suppose that the rectangular lattice G, defined as $G = I \times J$ has several rectangular sublattices (connected components) G_m, $m = 1, 2, \cdots$, defined as $G_m \in G$. If $D = \{d_{ij}: (i, j) \in G\}$ is MRF on the rectangular lattice G, and each $D_m = \{d_{ij}: (i, j) \in G_m\}$ is also MRF defined on the rectangular sublattice G_m, then the probability P over the MRF G is the product measure

$$P = \prod_m P_m \qquad (m = 1, 2, \cdots) \tag{13}$$

where P_m is probability measurement of MRF G_m.

From this, we see that the factorization property guarantees the decomposition of the complex potential function (see (11)) into a set of simple potential functions of each clique over the neighborhood.

B. Markov–Gibbs Equivalence

We call two state representations "equivalent," if one of them determines another and vice versa.

Preston [12] proved that the following are "equivalent" for a state π in a 2-D discrete random field.

1) π is an equilibrium state.
2) π is a state of MRF.
3) π is Gibbs state with nearest neighbor potential.

This equivalence, called the Markov–Gibbs equivalence, implies that a purely probabilistic notion of an MRF can be equated to the physically based Gibbs distribution. The Gibbs model describes the interaction of a macroscopic system in thermal equilibrium in the same way the spatial Markov models describe local dependence. For an MRF, the conditional probabilities are expressed in terms of nearest neighborhoods, while for a Gibbs distribution the energy E is the sum of potentials V measured over the same neighborhood.

From this we see that the MRF–Gibbs equivalence provides an explicit formula for the joint probability distribution in terms of an energy function, and it supplies a powerful mechanism for modeling spatial features.

Kinderman [13] proved that a nearest neighbor Gibbs distribution determines an MRF, and provided theoretical formulas to determine the canonical potential from the local characteristics.

The proof that the MRF determines a nearest neighbor Gibbs distribution is given by Preston [12], and Grimmett [14]. They note that there is not a unique potential function. However there is a unique canonical potential that is singled out in the following manner: the states are renumbered $0, 1, 2, \cdots, r$ with 0 playing the role of a preferred state. The potential is then said to be a canonical potential if $V_c(\omega) = 0$ when ω assigns the value 0 to at least one site in C. It is then proved that there is a unique canonical potential for a given MRF.

From the previous discussion we have seen that we are able to determine a Markov field if we have a nearest neighbor potentials. If we have a MRF on a finite 2-D integer set $I \times J$, then the local characteristics do uniquely determine this measure. In fact the canonical potential can be determined from these local characteristics by a formula, which we will discuss in Section IV, and then the Gibbs distribution is determined. We cannot merely choose any set of local characteristics because we have to satisfy the measure and be consistent with these characteristics. The greatest difficulty is in choosing appropriate local characteristics; so the key step in the method becomes to determine the proper potential function. We will discuss this in Section IV.

IV. The Markov–Gibbs Model for Bayes' Context Classification

In this section, we will show how the Markov–Gibbs model is incorporated with the Bayes' context classification, and how the optimal decision rule determines an interpretation C for the pixels in the neighborhood of pixel (i, j) that satisfies the maximality condition (7).

From (7), we know that with the zero-one loss function, the best decision procedure chooses a labeling C that satisfies the maximality condition

$$\prod_{(l,k)} P(d_{lk}|C_{lk})P(C) \geqslant \prod_{(l,k)} P(d_{lk}|Z_{lk})P(Z),$$

for all $Z \in \Omega$.

From the previous section, we can see that

$$P(D_{ij}, C) = \prod_{(ij) \in N} P(d_{ij}|C_{ij})P(C) \tag{14}$$

is a Gibbs distribution, since we can rewrite $P(D_{ij}, C)$ as

$$P(D_{ij}, C) = \frac{1}{Z} e^{-U(D_{ij}, C)/KT} \tag{15}$$

where $U(D_{ij}, C)$ is an energy function associated with Gibbs distribution $P(D_{ij}, C)$, which has the form

$$U(D_{ij}, C) = -\sum \log P(d_{lk}|c_{lk}) + U(C) \tag{16}$$

and $U(C)$ satisfies

$$P(C) = \frac{1}{Z} e^{-U(C)/KT}. \tag{17}$$

Z and K are constants and $U(C)$ is the energy function associated with Gibbs distribution $P(C)$, which has the form

$$U(C) = \sum_{L \in W_{ij}} V_L(C). \tag{18}$$

W_{ij} denotes the family of cliques in the neighborhood N_{ij} of pixel (i, j). Each V_L is a function on N_{ij}, with the property that $V_L(C)$ depends only on those coordinates and assigned labels of pixel (i, j), which are located in the clique L. Such a family $\{V_L,\ L \in W\}$ is called a potential. The V_L functions in (18) represent contributions to the energy $U(C)$ from external fields (singleton cliques), pair interactions (doubletons), etc.

Z is the normalizing constant given by

$$Z = \sum_C e^{-U(C)/KT} \tag{19}$$

where T, stands for "temperature." For our purposes T controls the degree of "peak" in the "density." Choosing T "small" makes it easier to find the minimal energy configurations by sampling; this is the principle of annealing, and will be applied to our procedure in Section V.

The assigned category, in the sense of Bayesian inference, is determined by maximizing (15). This is a maximum posterior estimate. The probability is maximized when the energy is minimized. This is analogous to the situation of thermal equilibrium in statistical physics, in which the most probable molecular configurations occur at the lowest energies. For the case of Bayes context classification, the most probable labeling occurs when the negative exponent is minimized. Using conventional gradient techniques, maximizing posterior probability is virtually impossible for all by the first-order MRF models, because of the existence of many local extrema. However stochastic relaxation, which is a new multivariate or combinatorial optimization technique (finding the minimum of a given function depending on many parameters), developed by Kirkpatrick *et al.* [11], offers a practical solution.

After creating the Gibbs models for Bayes' context classification, the problem now is to find $U(D_{ij}, C)$.

A general form of $U(D_{ij}, C)$ is that

$$\begin{aligned} U(D_{ij}, C) &= -\sum_{(l,k) \in N_{ij}} \log P(d_{lk}|C_{lk}) + U(C) \\ U(C) &= \sum_{(l,k) \in N_{ij}} V_{(l,k)}(C_{lk}) \\ &\quad + \sum_{(l,k) \in N_{ij}} V_{(l,k),(l+1,k)}(C_{lk}, C_{l+1,k}) \\ &\quad + \sum_{(l,k) \in N_{ik}} V_{(l,k),(l,k+1)}(C_{lk}, C_{l,k+1}) \end{aligned} \tag{20}$$

where the summation is over all $(k, l) \in N_{ij}$, and N_{ij} denotes the nearest neighbor. The Ising model [15], which is the earliest and best-known lattice system, can be thought of the special case of (20) in which C is binary ($C_{ij} = (0, 1)$), homogeneous (strictly stationary) and isotropic (rotational invariant). Its potential function is

$$U(C) = \alpha \sum_{(l,k) \in N_{ij}} C_{lk} + \beta \left(\sum_{(l,k) \in N_{ij}} C_{lk}{}^* C_{l} + {}_{1k} + \sum_{(l,k) \in N_{ij}} C_{lk}{}^* C_{lk+1} \right) \tag{21}$$

for some parameters α and β, which measure, respectively, the external field and bonding strengths.

For our contextual classification case, in which $U(C)$ is a function of pattern configurations, the expression for the Ising model (21) is not suitable. But we still assume that the image is homogeneous and isotropic.

Before we derive the canonical potential for the general case, we first describe a useful theorem of potential function [14]. If the random field $\mathcal{C} = \langle c_{ij}, (i, j) \in I \times J \rangle$ is a Markov field, then its potential function is given by

$$V_L(C) = \sum_{L_i \in W} (-1)^{|W - L_i|} \log P(C_{L_i}) \tag{22}$$

where the summation is over all cliques in W, C_{L_i} designates the configuration that agrees with C on L_i, but assigns the value 0 at all sites outside of L_i, and $\mu = (-1)^{|W-L_i|}$ is a Möbius function, and $W - L_i$, the difference of W and L_i, is $\{x | x$ is in W and x is not in $L_i\}$.

From this theorem we see that given an MRF on a finite set $I \times J$, the local characteristics (see Section III-A) do uniquely determine this potential function; the canonical potential can then be determined from these local characteristics.

Three types of potential functions are defined as

$$\begin{aligned}
&\sum_{(l,k) \in N_{ij}} V_{\{(l,k)\}}(C_{lk}) \\
&\quad = \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk}) \\
&\qquad \cdot \sum_{(l,k) \in N_{ij}} V_{\{(l,k),(l+1,k)\}}(C_{lk}, C_{l+lk}) \\
&\quad = 2^* \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk} | C_{l+1k}) \\
&\qquad \cdot \sum_{(l,k) \in N_{ij}} V_{\{(l,k),(l,k+l)\}}(C_{lk}, C_{lk+1}) \\
&\quad = 2^* \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk} | C_{lk+1}).
\end{aligned} \tag{23}$$

These definitions are based on the factorization of an MRF and a neighboring clique assumption. In this model only cliques of size two are involved.

From the previous definitions, the problem in this paper can be stated as follows: the assigned category is determined by minimizing

$$U(D_{ij}, C) = \sum_{(l,k) \in N_{ij}} \log P(d_{lk} | C_c) + U(C)$$

where

$$\begin{aligned}
U(C) = &\sum_{(l,k) \in N_{ij}} \log P(d_{lk} | C_c) + \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk}) \\
&+ 2^* \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk} | C_{l+1k}) \\
&+ 2^* \sum_{(l,k) \in N_{(ij)}} \log P(C_{lk} | C_{lk+1}).
\end{aligned} \tag{24}$$

Because of the existence of many local extrema, the computation cost of maximizing the posterior probability for Bayes classification is usually computationally high. For example if a MSS (multispectral scanner) image has N class categories on a $M \times M$ lattice, the number of configurations is at least N^{M^2}. Hence, the identification of even a near-optimal solution is surprisingly difficulty for such a relatively complex function. In Section V we present the implementation of a stochastic relaxation procedure, which overcomes the computational difficulty remarkably well.

V. Implementation of the Stochastic Relaxation Context Classification

The method used in the stochastic relaxation context classification is essentially a variant of a Monte Carlo procedure, due to Metropolis *et al.* [16]. In the Metropolis procedure samples are randomly generated from a Gibbs distribution at constant temperature. This simulates the behavior of a physical system in thermal equilibrium. The algorithm can be briefly described as follows. For each state D_{ij} of a model D, a random perturbation is made. The change in energy, ΔU is computed. If $\Delta U \leq 0$, the perturbation is accepted, that is the new pattern configuration, which correspondings to the new "energy," $U = U_o + \Delta U$, replaces the original one. If ΔU is positive then the perturbation is accepted with probability

$$P(\Delta U) = e^{-\Delta U / T}. \tag{25}$$

This conditional acceptance is easily implemented by choosing a random number R uniformly distributed between 0 and 1. If $R \leqslant P(\Delta U)$ then the perturbation is accepted; otherwise the existing model is retained. Random perturbation according to these rules eventually causes the system to reach equilibrium, or the configuration θ, corresponding to maximum probability. The technique used here, slowly lowers the temperature T during execution of the iterative procedure. If the system is cooled sufficiently slowly and equilibrium conditions are maintained, the model converges to a state with minimum energy or maximum posterior probability. This was proved by Geman and Geman [9]. Geman also pointed out that the most important aspect of any cooling function is that it be slow, especially near the critical temperature where convergence is rapid. The successful choice of an annealing schedule requires experience; ideally, the procedure would be interactive. As T decrease, samples from the distribution are forced towards the minimal energy configurations. The temperature $T(k)$ used by Geman satisfies the bound

$$T(k) \geqslant \frac{G}{\log(1+k)}. \tag{26}$$

It is employed in executing the kth site replacement (i.e., the kth classified labels are produced in the iteration scheme). For every k, G is a constant independent of k. When $k \to \infty$, the configurations generated by the algorithm will be those of minimal energy.

VI. Summary of the Stochastic Relaxation Context Classification Procedure

In summary the stochastic relaxation context classification procedure can be implemented as follows.

1) Evaluate training statistics; this includes the mean vector and the covariance required for the Gaussian class conditional distribution.

2) Preclassify the image using a pixel independent or context free Bayes classification technique.
3) Evaluate the transition probabilities: $P(C_{ij}|C_{i,j+1})$ and $P(C_{ij}|C_{i+1,j})$ from the preclassification results.
4) Use (24)–(26) to perform the stochastic relaxation context classification. The experimental results with both simulated and real multispectral remote sensing data will be presented in Section IX.

VII. Improved Scheme

Now we have a desirable stochastic relaxation procedure, in which samples are randomly generated from a Gibbs distribution at a controlled temperature T. As T changes, samples from the distribution are forced towards the minimal energy configuration. Geman and Geman [9] proved the convergence properties of this algorithm, and showed how to reduce the computational difficulty. As we mentioned before, for an MSS image which has N class categories and a $M \times M$ size, the number of configuration is at least N^{M^2}. In Gemans' scheme the pattern samples are randomly collected from a huge pattern configuration space. In contrast to our proposed method, his method had nothing to do with reducing the pattern configuration space. Experimental results showed that for a significant improvement in classification accuracy, the number of iterations was still sizeable.

In order to further reduce the computational complexity, it is important to reduce the size of the huge pattern configuration space or to place some constraints on the pattern generation procedure. We now describe how we can use the homogeneous assumption to control the pattern configuration sampling procedure.

Most Landsat and aerial photograph images are divided into a number of elementary regions at the classification stage. Each region is finite, fairly homogenous, and has similar spectral properties over its entire ground surface. These homogeneous regions correspond to uniform objects (categories) on the earth's surface. We believe that some smooth or homogeneous pattern configurations are much more probable than others, and some irregular patterns have low probabilities.

This fact gives us a strategy for the iterative procedure in that we may use these most probable homogeneous patterns at the beginning of the iteration procedure. After that we randomly generate the pattern configuration and skip irregular patterns that have low probability.

We should note that the global procedure is still random; we only set special pattern configurations into initial states, and give some constraints on the iterative procedure. Therefore this scheme is still a stochastic relaxation procedure.

Let $T(t)$ be any decreasing sequence of temperatures for which a) $T(t) \to 0$ as $t \to \infty$, b) $T(t) \geq N\Delta/\log t$, for all $t \geq t_0$ and some integer $t_0 \geq 2$. And let the annealing procedure generate a process $\{D(t),\ t=1,2,\cdots\}$. Geman proved that the distribution of $D(t)$ converges to equilibrium distribution π, as $t \to \infty$ regardless of starting configuration. The only assumption is that we continue to visit every site.

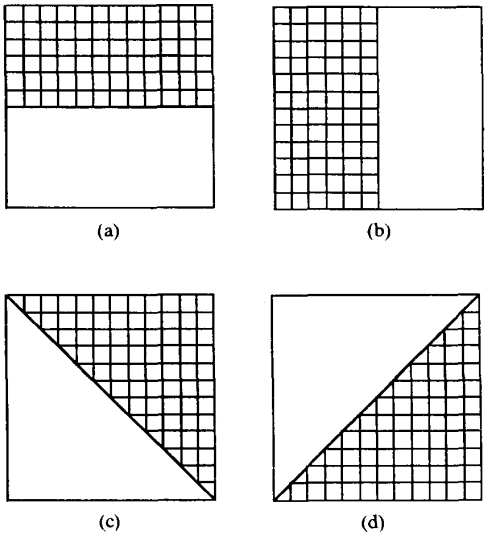


Fig. 1. Four simple pattern configurations for improved scheme; (a) has upper and lower parts; (b) has left and right parts; (c) has upper-right and bottom-left parts; (d) has upper-left and bottom-right parts. Uniform pattern and four simple pattern configurations, which are assumed to have higher occurring probability, are generated and tested at beginning of iterative procedure.

The convergence of our modified procedure can be proved as follows. Let process $\{D_1(t),\ t=1,\ 2,\cdots\}$ be generated by the annealing procedure previously described, and process $\{D_0(1), D_0(2),\cdots, D_0(s)\}$ be a process with limited numbers of states. We create a new process $D(t)$, which has

$$D(t) = \begin{cases} D_0(t), & \text{if } t \leq s \\ D_1(t-s), & \text{if } t > s \end{cases}$$

where $D_0(t)$ is a sequence of special pattern configurations at the first stage, and $D_1(t-s)$ is a process generated by the annealing procedure. Geman states that the distribution of $D_1(t)$ converges to the equilibrium distribution, as $t \to \infty$ regardless of starting configuration. Because $D(t)$ has same statistical property except having different starting configurations, new processes $D(t)$ should converge to the same equilibrium distribution as $D_0(t)$, when $t \to \infty$.

First, we assume that uniform pattern configurations have higher occurring probabilities, and that they are generated and tested at the beginning of the iterative procedure. The assignment of these uniform labels is based on the labels in the neighborhood assigned at the classification stage. So the number of these uniform pattern configurations is equal to the number of categories in the neighborhood.

Subsequent to this testing, we assume that some simple pattern configurations (Fig. 1) also have higher occurring probabilities. These are assigned and tested again.

The pattern in Fig. 1(a) has upper and lower parts. The assignment of labels in each part is also based on the

labels in that part assigned at the preclassification stage. Similarly, Figs. 1(b), (c), and (d) show three other simple patterns.

After these steps, a random pattern generator is introduced in the relaxation procedure.

In order to restrict irregular patterns, we employ a measure of the irregularity as follows:

$$\mathrm{IR} = \frac{\text{number of classes in the neighborhood}}{\text{number of pixels in the neighborhood}}. \quad (27)$$

After we give a threshold, the irregularity measurement of each pattern is calculated and compared with the threshold. If the measurement is larger than the threshold, the pattern is too irregular and the procedure will skip testing and generate the next pattern instead.

VIII. Parallel Algorithm

Although the computational cost of the stochastic relaxation scheme is much more expensive than the conventional context free classification methods, it is highly parallel in the sense that it is implementable by simple and identical neighbor operators.

The performance of each neighborhood operator is independent from other neighborhood operators in the entire image. The amount of time required for each iteration of the entire image is proportional only to the number of pixels in the image, if all operators are executed in parallel.

This important property allows the algorithm to run naturally in a fully parallel architecture.

A more modest degree of parallelism was noted by Geman and Geman [9]. A graph associated with the MRF is divided into collections of sites with each collection assigned to an independently running (asynchronous) processor.

Each such processor would execute a raster scan update of its assigned sites. Communication requirements will be small if the division of the graph respects the natural topology of the scene, provided of course that the neighborhood systems are reasonably local. Such an implementation with five or ten micro- or minicomputers represents a straightforward application of available technology. Chen [17] noted that data flow computer architecture should be useful for the stochastic relaxation (annealing) algorithm for MRFs.

IX. Experiment Results of the Contextual Classification Algorithm

In order to show accuracy in the improvement of stochastic relaxation context classification methods, several experimental results based on both simulated and real multispectral remote sensing data are illustrated.

A. Simulated Data Experiments

Classification accuracy can vary with different kinds of input data sets for a given classification algorithm. It is difficult to evaluate the effectiveness of classification algorithms using small sets of real image data.

A desirable way to evaluate the effectiveness of classification algorithms is to use simple simulated data sets. In this subsection we illustrate a simulated data experiment, which is generated from the ground truth of a real remote sensing image.

The simulated data generating method proposed in [18] is as follows. Use the ground truth (or classification map) and associated estimated mean vectors and covariance matrices of the classes (developed in performing the no-context classification). For each pixel, simulated data vectors are produced by a Gaussian random number generator having the same mean vector and covariance matrix as the class associated with the pixel on the ground truth map. Thus the pixel in the simulated data set has the following characteristics.

1) Each pixel in the simulated data set represents the same class as in the ground truth data.
2) All classes have multivariate Gaussian distributions with parameters typical of those found in the ground truth data.
3) All pixels measurement values are class-conditionally independent of adjacent pixels.
4) There are no mixed pixels.

Data simulated in this manner are an idealization of real remote sensing data, but the spatial organization of the simulated data is consistent with a real world scene, and the overall characteristics of the data are consistent with the contextual classifier assumption.

B. Experimental Results of Real and Simulated Remote Sensing Images

The technique is first illustrated using a simulated image, which is generated from a digital remote sensing data collected by the Landsat MSS. The experimental data, which was a subset of the April 13, 1976 MSS scene of Roanoke, VA, was selected as the first study area. It was classified by a Bayes context free-classification method in order to compare the results. The following ground cover classes were used: 1) Class 1: Urban or built-up land, 2) Class 2: Agricultural land, 3) Class 3: Rangeland, 4) Class 4: Forest land, 5) Class 5: Water (Only a small amount), 6) Class 6: Wetland (Only a small amount), 7) Class 7: Barren land.

Because the study area was selected from the Roanoke, VA mountainous region (longitude from 79 52′ to 80 00′ W; latitude from 37 15′ to 37 23′ N), the land cover of this region is a complex pattern of diverse spectral classes occurring in small parcels. The most easily classified of this land cover class—open water —is not represented in this test area. Thus this area is a difficult area for the conventional pixel independent classification technique. The accuracies of context free classification for such remote sensing images, including Bayesian classifiers and ISODATA [19] are 60 percent. Such accuracies are not unusual for scenes of this complexity.

Fig. 2. First band of MSS scene of Roanoke, VA, April 13, 1976. Image size 151×151.

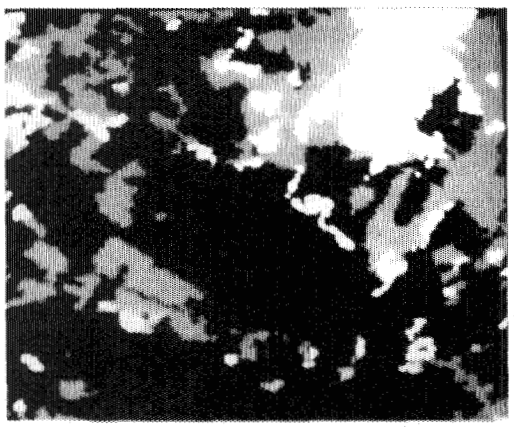


Fig. 3. Ground truth image of Roanoke, VA. Image size 151×151.

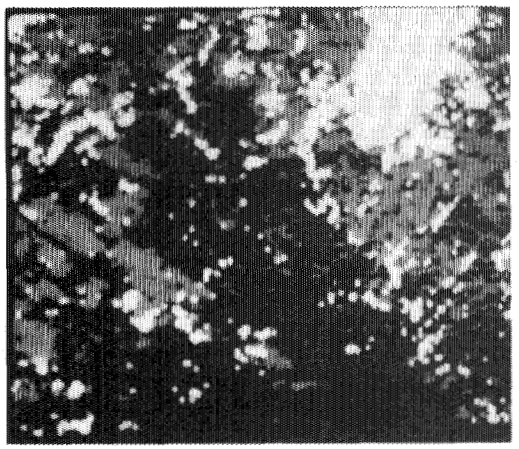


Fig. 4. Bayes preclassification result of MSS, Roanoke, VA.

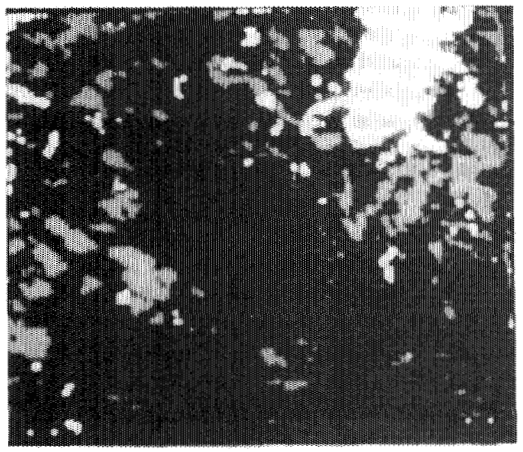


Fig. 5. Context classification result by stochastic relaxation of MSS, Roanoke, VA.

The test image has four bands of digital multispectral data. The mean vectors $\boldsymbol{m}_i$ and the covariance matrices Σ_i for each class i are calculated. Then a simulated image having the approximate characteristics can be generated by the Gaussian model.

As mentioned before, the key step of the contextual classification scheme we presented in this paper is minimizing (24).

The prior probability of each class is calculated from the preclassification results. The transition probability $P(C_{ij})|C_{ij-1})$ $P(C_{ij}|C_{i-1j})$ in (24) can be obtained from maximum likelihood estimation. The conditional probability $P(d_{ij}|C_{ij})$ is estimated from the training sets and ground truth. In this experiment means and covariance matrices of each category were calculated from the ground truth data. The class conditional probability $P(d_{ij}|C_{ij})$ are assumed multidimensional normal

$$P(d|C) = \frac{1}{(2\pi)^{n/2}\left|\sum\right|^{1/2}} e^{-1/2(d-m)^T\Sigma^{-1}(d-m)} \quad (28)$$

where n is the dimension of the feature space.

The context free and context classification results with both simulated and real multispectral remote sensing data are shown in Figs. 2–15 and Tables I–IV. The contingency table and classification images lead us to the conclusion that the context classification provides better results than the context free classification algorithm.

By visually examining these results, one can easily tell how good the performances are within each class, and also along the boundaries between classes. At first sight, we see that the Bayesian context-free classifiers results are quite noisy. The Markov context classification seems to "clean up" the picture significantly. It can be seen that many small isolated pixels are eliminated, and that each area is much more homogeneous in the contextual classification results. Boundaries remain correctly placed. The MSS four-band image, ground truth map, which had been classified by professional analysts, are given in Fig. 2 and Fig. 3, respectively. This comparison and Fig. 5 indicated that a 5–10 percent improvement of accuracy was obtained by the context classification method. Thus in addition to the

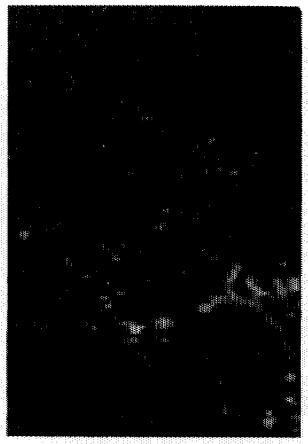


Fig. 6. Third band of MSS scene of California (I). Image size 130×90.

Fig. 7. Bayes preclassification result of MSS from California (I).

Fig. 8. Stochastic relaxation classification result of MSS from California (I).

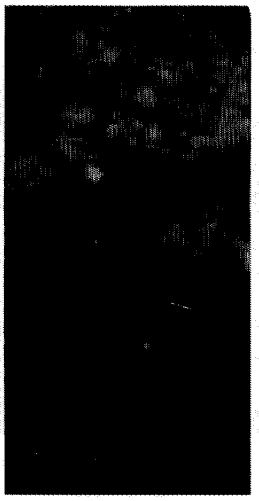


Fig. 9. Third band of MSS scene of California (H). Image size 110×70.

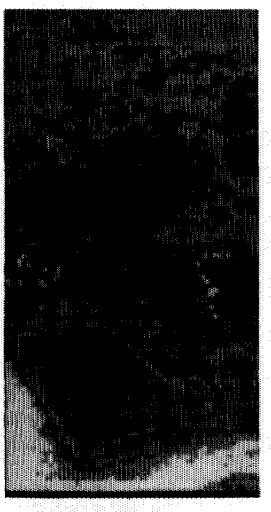


Fig. 10. Bayes preclassification result of MSS from California (H).

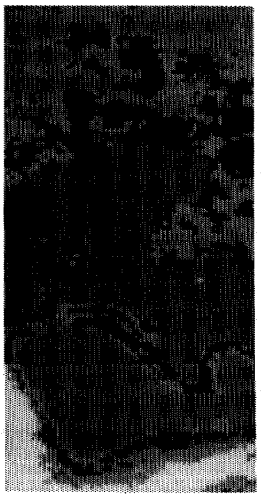


Fig. 11. Stochastic relaxation context classification result of MSS from California (H).

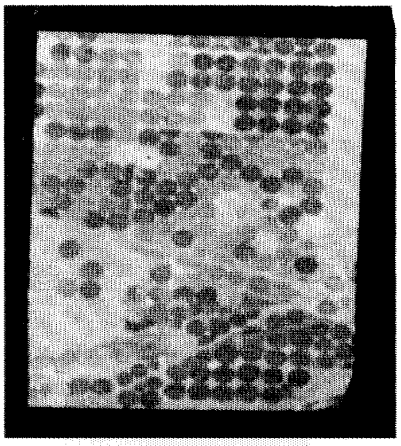


Fig. 12. First band of MSS scene of Clarke, OR, 1982. Image size 150×150.

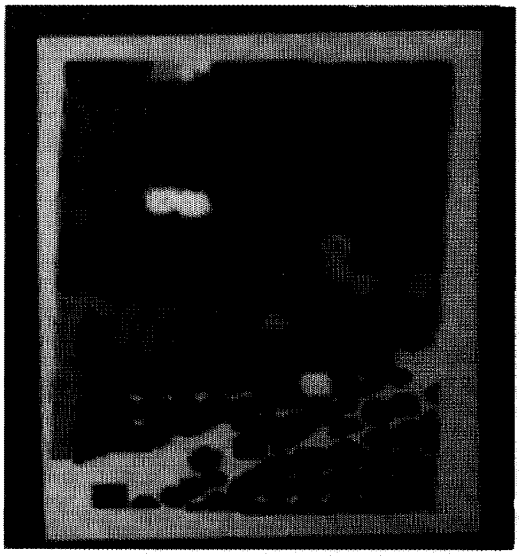


Fig. 13. Ground-truth image of crop field at Clarke, OR, 1982. Image size 150×150.

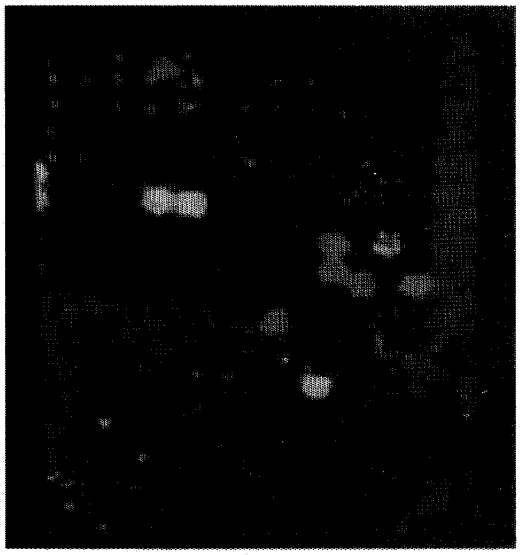


Fig. 14. Bayes preclassification result of MSS, Clarke, OR.

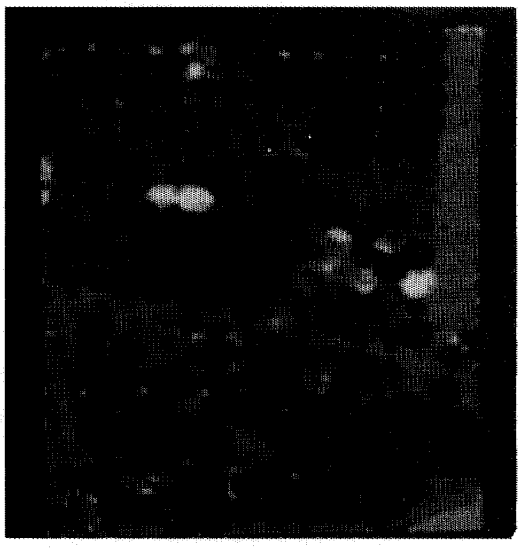


Fig. 15. Contextual classification result by stochastic relaxation, Clarke, OR.

visual improvement, the context classification scheme also improves the classification accuracy.

The second study area is California. Three MSS classification results where sizes of subsets are 130×90, 101×70, 130×60 respectively, are shown in Figs. 6–11. These results show that the algorithm was effective in several different areas with varied categories and preclassification accuracies (these areas had about 90 percent preclassification accuracy).

The third study area is a crop field at Clarke, Oregon (Figs. 12–15). The Landsat MSS image is 12-band-data set (Landsat MSS bands 4–7 from three dates). The accuracies of maximum likelihood classification and canonical analysis method in the same study area were about 75 percent. Our context classification scheme raised the classification accuracy to 80.8 percent.

In order to study the effects of noise, independent zero-mean Gaussian noise $N(0,\sigma^2)$ was added to the 4-bands MSS simulated image at three different noise standard deviations 1, 2, and 3. Then, the noisy image was classified by the pixel independent Bayes classifier and the stochastic relaxation approach to context classification. The overall classification accuracy is measured as the ratio of the number of correctly classified pixels to the total number of classified pixels. It is plotted as a function of the noise standard deviation in Fig. 16.

It can be seen that the Bayes context-free classification is very sensitive to random noise, but the context classification methods are quite opposite and are superior to the context free classifier.

TABLE I
BAYES PIXEL INDEPENDENT CLASSIFICATION RESULTS FOR MSS SCENE OF ROANOKE, VA[a,b]

Class	URB	AGR	RNG	FST	Total	ACC (percent)[c]
URB	760	512	0	162	1437	52.8
AGR	116	379	0	83	578	65.6
RNG	0	0	0	0	0	—
FSN	15	28	0	210	253	83.0
Total	894	919	0	455	2268	59.6[d]

[a]April 13, 1976.
[b]Scale factor of the number of pixels 10**1. Columns represent assigned categories. Rows represent true categories.
[c]Classification accuracy.
[d]Overall classification accuracy: ratio of the number of correctly classified pixels to the number of total classified pixels. URB-urban or built-up land, AGR-Agricultural land, RNG-Rangeland, FSN-Forest land.

TABLE II
STOCHASTIC RELAXATION CONTEXT CLASSIFICATION RESULTS FOR MSS SCENE OF ROANOKE, VA[a,b]

Class	URB	AGR	RNG	FST	Total	ACC (percent)[c]
URB	1108	317	0	12	1437	77.1
AGR	179	385	0	14	578	66.6
RNG	0	0	0	0	0	—
FSN	50	35	0	168	253	64.4
Total	1337	737	0	194	2268	73.3[d]

[a]April 13, 1976.
[b]Scale factor of the number of pixels 10**1. Columns represent assigned categories. Rows represent true categories.
[c]Classification accuracy.
[d]Overall classification accuracy: ratio of the number of correctly classified pixels to the number of total classified pixels. URB-Urban or built-up land, AGR-Agricultural land, RNG-Rangeland, FSN-Forest land.

TABLE III
PIXEL INDEPENDENT BAYES CLASSIFICATION RESULTS OF TEST IMAGE "CLARK"[a]

Class	WHT	ALF	POT	CRN	BNS	APL	PAS	RNG	Total	ACC (percent)[b]
WHT	1017	47	30	5	4	0	10	75	1188	85.5
ALF	71	382	135	10	13	6	12	39	668	57.1
POT	40	32	522	5	29	0	2	32	652	84.6
CRN	1	5	1	65	2	0	0	4	78	83.3
BNS	0	1	0	1	1	0	0	0	3	0
PAS	0	0	0	0	0	0	9	2	11	81.1
RNG	15	12	14	2	4	1	9	335	392	85.4
Total	1146	483	704	89	78	7	42	490	3040	77.5[c]

[a]Scale factor of the number of pixels 10**1. Columns represent assigned categories. Rows represent true categories.
[b]Classification accuracy.
[c]Overall classification accuracy: ratio of the number correctly classified pixels to the number of total classified pixels. WHT-wheat; ALF-alfalfa; POT-potatoes; CRN-corn; BNS-beans; APL-apples; PAS-pasture (irrigated); RNG-rangeland.

TABLE IV
STOCHASTIC RELAXATION CONTEXT CLASSIFICATION RESULTS OF TEST IMAGE "CLARK"[a]

Class	WHT	ALF	POT	CRN	BNS	APL	PAS	RNG	Total	ACC (percent)[b]
WHT	1080	23	25	1	1	0	0	58	1118	90.9
ALF	91	378	155	1	0	0	0	41	666	56.8
POT	54	23	544	1	3	0	0	29	654	83.2
CRN	1	5	1	65	0	0	0	6	78	83.3
BNS	2	5	2	0	35	0	0	4	48	73.9
APL	1	2	0	0	0	0	0	0	3	0
PAS	0	1	0	0	0	0	7	3	11	63.7
RNG	17	11	14	1	0	0	1	349	392	89.1
Total	1643	573	787	72	47	0	10	1158	3040	80.8[c]

[a]Scale factor of the number of pixels 10**1. Columns represent assigned categories. Rows represent true categories.
[b]Classification accuracy.
[c]Overall classification accuracy: ratio of the number correctly classified pixels to the number of total classified pixels. WHT-wheat; ALF-alfalfa; POT-potatoes; CRN-corn; BNS-beans; APL-apples; PAS-pasture (irrigated); RNG-rangeland

X. CONCLUSION

We have developed a new multispectral image context classification algorithm with the MRF, where remotely sensed data are more accurately classified compared to traditional context free classifiers. This new approach of multispectral image context classification is based on a stochastic relaxation algorithm and the Markov–Gibbs random field. The implementation of the relaxation algorithm is one form of optimization using annealing. In this paper we have first motivated a Bayesian context decision rule, then introduced a Markov–Gibbs model for the original Landsat MSS image. Then we developed a new contextual classification algorithm, in which maximizing the posterior probability (MAP) is based on the stochastic relaxation and annealing method. An improved algorithm has been presented to speed the stochastic relaxation procedure. It has greatly reduced the number of iterations by using some special pattern configurations at the beginning of the iterative procedure. The algorithm is highly parallel and exploits the equivalence between Gibbs distributions and MRF's.

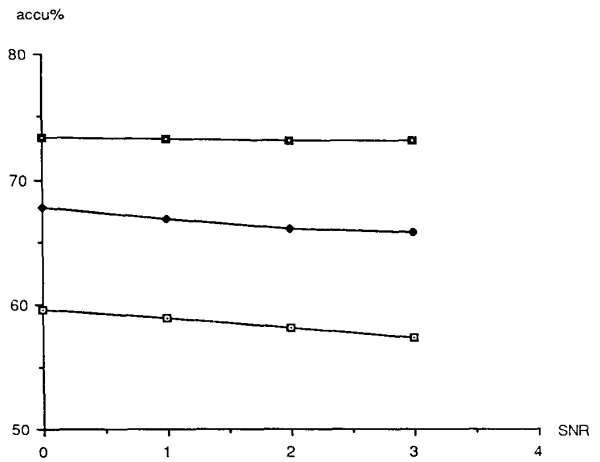


Fig. 16. Overall classification accuracy curves versus noise level. Yellow line: —■— contextual classification by stochastic relaxation. Green line: —◆— Dynamic programming approach to context classification (two-pass forward-backward algorithm). Red line: —□— Pixel independent Bayes classification.

REFERENCES

[1] K. S. Fu, and T. T. Yu, *Statistical Classification Using Contextual Information* New York: Research Studies Press, John Wiley, 1980.

[2] P. H. Swain, S. B. Vardeman, and J. C. Tilton, "Contextual classification of multispectral image data," *Pattern Recognition*, vol, 13. no. 6, pp. 428–441, 1981.

[3] R. M. Haralick, M. C. Zhang and J. B. Campbell, "Multispectral image context classification using the Markov random field." *PECORA Proc.*, Oct. 1984, pp. 190–200.

[4] P. Whittle, "Some distribution and moment fornulae for the Markov chain."*J. Roy. Stat. Soc.*, vol. 17, series B, pp. 235–242, 1955.

[5] M. S. Bartlett, *An Introduction to Stochastic Processes*. Cambridge, MA: Cambridge Univ. Press, 1955.

[6] M. S. Bartlett, "Inference and stochastic processes," *J. Roy Stat. Soc.*, vol. A 130, pp. 457–477, 1967.

[7] M. S. Bartlett, "A further note on nearest neighbor models," *J. R. Stat. Soc.*, vol. A 131, pp. 579–580, 1968.

[8] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc.* vol. B36, pp. 192–236, 1977.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.* vol. 6, no. 6, pp. 721–724, 1984.

[10] R. M. Haralick, "Decision making in context," *IEEE Trans. Pattern Anal. Machine Intell.*, vol, PAMI-5, no. 4, pp. 417–428, July 1983.

[11] S. Kirkpatrick, C. D. Gellatt, Jr. and M. P. Vechi, "Optimization by simulated annealing," *Science*, no. 222, pp. 671–680, 1982.

[12] C. J. Preston, *Gibbs States on Countable Sets*. Cambridge, MA: Cambridge Univ. Press, 1974.

[13] R. P. Kinderman, "*Random field: Theorems and examples*," *J. Undergraduate Math*, vol. 5, pp. 25–34, 1973.

[14] G. R. Grimmertt, "A theorem about random fields," *Bull. Lon. Math. Soc.*, vol. 5, pp. 81–84, 1973.

[15] E. Ising, "Beitray sur theorie des ferromagnetismus," *Zeitschrift Physik*, vol. 31, pp. 253–258, 1925.

[16] N. Metroplis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines." *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.

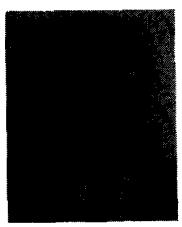
[17] S. S. Chen, "A data flow computer architecture for Markov image models," *IEEE Computer Society Workshop on Computer Architecture for Pattern Anal. Image Database Management*, 1985, pp. 75–79.

[18] P. H. Swain, H. J. Siegel, and B. W. Smith, "A method for classifying multispectral remote sensing data using context," Machine Processing of Remote Sensed Data Symposium, 1979, pp. 343–353.

[19] G. H. Ball and D. J. Hall, "ISODATA, a novel method of data analysis and pattern classification," Stanford Research Inst. Tech. Rep., 1973.

Ming Chuan Zhang received the B.Sc degree in electrical engineering from Beijing University, in 1968, the M.Sc degree in electrical engineering from the University of California, Santa Barbara in 1982, and the Ph.D degree in electrical engineering from Virginia Polytechnic Institute and State University Blacksburg, VA in 1986.

In 1985–1986 he was a visiting faculty of computer science in Purdue University at Indianapolis. Since 1986 he is an Assistant Professor of Computer Science in University of North Carolina at Charlotte. His current research interests include image processing, computer vision, pattern recognition, neural network and artificial intelligence.



Robert M. Haralick (S'62-M'69-SM'76-F'84) was born in Brooklyn, NY, on September 30, 1943. He received the B.A. degree in mathematics from the University of Kansas, Lawrence, in 1964, the B.S. degree in electrical engineering in 1966, the M.S. degree in electrical engineering in 1967, and the Ph.D. degree from the University of Kansas in 1969.

He has worked with Autonetics and IBM. In 1965 he worked for the Center for Research, University of Kansas, as a Research Engineer and in 1969 he joined the faculty of the Department of Electrical Engineering there where he last served as a Professor from 1975 to 1978. In 1979 he joined the faculty of the Department of Electrical Engineering at Virginia Polytechnic Institute and State University where he was a Professor and Director of the Spatial Data Analysis Laboratory. From 1984 to 1986 he served as Vice President of Research at Machine Vision International, Ann Arbor, MI. He now holds the Boeing Clairmont Egtvedt chaired professorship in the Department of Electrical Engineering, University of Washington, Seattle. He has done research in pattern recognition, multi-image processing, remote sensing, texture analysis, data compression, clustering, artificial intelligence, and general systems theory, and has published over 180 papers. He is responsible for the development of GIPSY (General Image Processing System), a multiimage processing package that runs on a minicomputer system.

Dr. Haralick is a member of the Association for Computing Machinery, Sigma Xi, the Pattern Recognition Society, and the Society for General Systems Research.

James B. Campbell received his B.A. in geography from Dartmouth College in 1966, the M.A. and the Ph.D. degrees in geography from the University of Kansas, Lawrence, in 1972 and 1976, respectively.

During military service he was a photo interpreter for the U. S. Army, and later he was employed by the Kansas Geological Survey. He is a Professor at Virginia Polytechnic Institute and State University, his teaching courses in remote sensing, cartography, and physical geography.

Dr. Campbell is the author of numerous scientific papers as well as reference books and texts addressing problems in land-cover mapping, remote sensing, and soil science.