CONTEXTUAL CLASSIFICATION BY STOCHASTIC RELAXATION

Mingchuan Zhang, Robert M. Haralick*, James B. Campbell #

Indiana Univ.-Purdure Univ. at Indianapolis, IN 46223.
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.
**Machine Vision International Ann Arbor, MI 48104.

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.

In this paper we first motivate a Bayesian context decision rule, and introduce a Markov-Gibbs model for the original LANDSAT MSS image, and then develop a new contextual classification algorithm, in which maximizing the a posterior probability (MAP) is based on the stochastic relaxation and annealing method. Finally, we present experimental results which are based on simulated and real multispectral remote sensing image to show how classification accuracy is greatly improved. 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 [4]; an estimation method of the context function was discussed by J.C.Tilton, S.B Vardenman and P.H Swain [5] [6] [8]; a recursive context classification using dynamic programming was presented by Haralick and M.C. Zhang[2].

In this paper we develop a new context classification approach, which is based on a stochastic relaxation algorithm and Gibbs distribution. These are not new concepts, as they have been used in statistical physics for many years. There, the problem of analysing the macroscopic properties of a physical system is translated into one of analysing the global properties of random fields with a given local structure. However, only Geman[3] has introduced these concepts into image restoration, and he has 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 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 introduce a Markov-Gibbs model for the original LANDSAT MSS image, and then develop a new contextual classification algorithm, in which maximizing the a posteriori probability (MAP) is based on stochastic relaxation and annealing method. Finally, we present experimental results with both simulated and real multispectral remote sensing data to show how classification accuracy is greatly improved. The algorithm is highly parallel and exploits the equivalence between Gibbs distributions and Markov Random Fields (MRF).

(II) Motivation and Proposed Approach:

The context information, which we would like to study, is a form of correlation existing among the succesive pattern classes in the two-dimensional 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 can be extracted by the specified spatial process. The most important quantity in the context Bayes' decision problem is the joint density function of all 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 practical reasons most of the previous studies deal with some specific cases, in which the 4-neighborhood assumption is invariably utilized in the context algorithms, and the contextual information is incorporated by considering the posterior class probabilities of pixel(i,j) given not only pixel(i,j) but also 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 to be introduced in this section improves

this by using a larger context.

Before presenting this rule, we must first give the notation, and assume that each pixel of the multi-band image considered in the paper has a N-tuple of finite gray tone values, and each component of it takes one value from the set D = { 0,...,255 } .

(II-1) Notation:

(1) M_r: designates row size of an image.

(2) M_c: designates column size of an image.

(3) I: designates the row index set of an image.
$I = < 1, ..., M_r >$

(4) J: designates the column index set of an image.
$J = < 1, ..., M_c >$

(5) (i,j) : designates an image position
$(i,j) \varepsilon I X J$

(6) d_{ij} : an N-tuple observed measurement vector from pixel (i,j).

(7) D_N : the collection of all measurement vectors in the neighborhood.

(8) C : assigned category labels in the neighorhood.

(9) C^o : assigned category labels in the neighborhood, excluding the central pixel of the neighborhood.

(10) $2N_r+1$: designates row size of an neighborhood.

(11) $2N_c+1$: designates column size of neighborhood.

(12) L: designates the local row index set of neighborhood.
$L = < -N_r, ..., N_r >$

(13) K: designates the local column index set of neighborhood.
$K = < -N_r, ..., N_c >$

(14) (l,k) : designates a local position in neighborhood.
$(l,k) \varepsilon L X K.$

(15) c_{lk}: assigned category labels of pixel (l,k) in neighborhood.

(16) $\Omega=\{\omega_1,\omega_2,....\omega_n\}$: the collection of all categories.

(17) θ : designate a pattern configuration of assigned labels in neighborhood.
$\theta = < \theta_{lk} \varepsilon\Omega, (l,k) \varepsilon L X K >$

(II-2) BAYESIAN CONTEXT CLASSIFICATION MODEL:

From the Bayesian Model[Haralick,1], the context classification problem can be stated as: to assign labels c to the pixels in the neighborhood of pixel (i,j) which minimizes the expected loss

$$\sum_{C^*} L(C,C^*)P(C^*|D,Q) \qquad (2.1)$$

where $P(C^*|D,Q)$ is the probability that the assigned labels are the true labels given i) the measurements of the pixels in the neighborhood of pixel (i,j) ii) the prior information Q we have about pixel dependencies. And where $L(C,C^*)$ is the loos incurred for the assignment of interpretation C to the pixels in the neighborhood of pixel (i,j), when the true interpretation are 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 assigned interpretations are correct. Thus, there is no distinction in loss between an incorrect assignment in which only one pixel is incorrectly assigned or an incorrect assignment in which all pixels are incorrectly assigned.

Such a loss function is defined by

$$L(C,C^*) = \begin{cases} 0 & \text{where } C = C^* \\ 1 & \text{otherwise} \end{cases} \qquad (2.2)$$

There are two assumptions about the would and pixel measurement process which can simplify the expected loss expression (2.1). 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,

$$P(D \mid C^*,Q) = \prod_{(i,j) \varepsilon L X K} P(d_{ij} \mid C^*,Q) \qquad (2.3)$$

The second assumption states that the n-tupe measurement of pixel (i,j) depends only upon the true interpretation, C^* associated with any other pixel. Hence

$$P(d_{ij} \mid C , Q) = P(d_{ij} \mid C_{ij}) \quad (i,j) \varepsilon \Omega$$

Under these assumptions, the optimal decision rule determines
interpretations C for the pixels in the neighborhood which minimize:

$$\sum_{C^*} L(C,C^*) \prod_{(i,j)} P(d_{ij} \mid c_{ij}) P(C^*)/P(D) \qquad (2.4)$$

With the loss function defined by (4.2), the best decision procedure chooses interpretation C which satisfy the maximally condition

$$\prod_{(i,j)} P(d_{ij} \mid C_{ij})P(C) >= \prod_{(i,j)} P(d_{ij} \mid Z_{ij})P(Z) \qquad (2.5)$$

for all $Z_{ij} \varepsilon \Omega$.

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

(III) Markov Random Fields and Gibbs states 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 and Gibbs model, which is pervasive in the image processing literature, constitute a promising natural way to capture context assumptions in classification.

Preston [17] or Spritser [21] give the basic theory of Markov Random Fields, and show that the Markov Random Field should enjoy the same wide variety of applications that Markov chains have. They make the material available to people outside of mathematics, as well as to discuss certain of its applications to other areas. On the use of the Markov Random Field for image processing and pattern recognition see Abend [15],Dobrushin[16], Wong , Kanal. More recent results are due to Martin [22], Laveen, Kanal [23], Chellappa [24].

Consistent with the two-dimensional(2-D) discrete Markov random field for multispectial image processing applications, we assume a random observation vector X_{ij}, whose component take one gray tone value from the set D={0,...,255}, and the pixel position (i,j) is defined on the two dimensional finite integer set of size I X J.

The Markov random field model may be defined as below:

Let $< X_{ij}, (i,j) \varepsilon I X J >$ be an observation from an image, and I, J and (i,j) are shown in the section I-2. It is postulated that this is generated by an appropriate 2-D (non causal) Markov Random Field model. The model characterizes the statistical dependency among pixels by requiring that

$$P(X_{ij} | X_{mn}: (m,n) \varepsilon I X J, (1,k) \neq (i,j)) = P(X_{ij} | X_{mn}: (m,n) \varepsilon N_{ij}) \quad (3.1)$$

Where N is the appropriate symmetric neighbor set. For instance N = {(0,1),(0,-1),(-1,0),(1,0) } corresponds to taking the simplest Markov model and by including more neighbors we can construct a higher order of Markov model.

It is 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 appropriate. 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.

(III-1) Three Markov Properties:

Before describing the relaxation algorithm, we first discuss three Markov properties for a random field studied in this paper [15-17]: i) the nearest neighbor property, ii) the Markov property, and iii) factorizability property characteristic to Gibbs states with nearest neighbor potentials.

.....

For information on title, please contact author.

The factorizability property is theoretical base of decomposing the potential functions by cliques. It is very useful for us to find the canonical potential form in our problem.

The facterlization property also can be stated as below:

Suppose that the graph G has several connected components G_{S1}, G_{S2}.... . If (X<P) has one of the properties given by nearest neighborhood, Markov Random Field or facterization function on the graph G. The each subgraph (X_{Si}, P_{Si}) has the same property on the subgraph G_{si}, i = 1,2,.... . Also if (X,P) has Markov property on G, then P is the product measure

$$P = P_{S1} * P_{S2} \cdots \quad (3.5)$$

Form above, we see that the facterlization property quarantine to decompose the complex potential function into summation of simple potential functions of each clique over the neighborhood. This is key point of our algorithm.

(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 interperetation C for the pixels in the neighorhood of pixel (i,j) which satisfy the maximality condition (2.2).

Recall, the joint probability P(D,C) is a Gibbs distribution over neighborhood, with corresponding energy function U and potentials Vc, and can be expressed as following:

$$P(D,C) = 1/Z \exp(-U(D,C)/KT) \quad (4.1)$$

$$U(D,C) = \sum_{L_i \varepsilon W} V_{Li}(D,C)$$

From (2.3), we know that with the zero-one loss function, the best decision procedure choses lable C which satisfy the maximality condition

$$\prod_{(i,j) \varepsilon L X K} P(d_{ij}|C_{ij})P(C) > \prod_{(i,j) \varepsilon L X K} P(d_{ij}|z_{ij})P(Z) \quad (4.2)$$

The Gibbs distribution, with corresponding energy function U(C) is formed as below

$$P(C) = 1/Z \exp(-U(C)/KT) \quad (4.3)$$

Z and K are constants and U is the energy function, which have the form

$$U(C) = \sum_{L_i \varepsilon W} V_{Li}(C) \quad (4.4)$$

W denote the set of cliques on the graph G. Each V_{Li} is function on G with the property that $V_{Li}(C)$ depends only on those coordinates of and assigned labels of pixel (i,j), for which locate in the clique. Such a family { V_{Li}, $L_i \varepsilon W$ } is called a potential . Z is the normalizing constant

$$Z = \sum_{L_i \varepsilon W} \exp (- U_{L_i}(C)/KT) \tag{4.5}$$

T, stands for "temperature"; for our purposes, T controls the degree of "peak' in the "density". Choosing T "small" exaggrates the models, making them easier to find by sampling; this is the principle of annealing, and will be applied to our late procedure.

From the previous section, we know the $B(C,D) = \Pi P(d_{ij}|C_{ij})P(C)$ also is a Gibbs distribution:

$$B(C,D)= 1/Z \exp (-U_A/KT) \tag{4.6}$$

Since $P(d_{lk}| c_{lk}) = \exp (\log P(d_{lk}| c_{lk}))$

$P(d_{lk}) = \exp(\log p(d_{lk}))$

and

$$P(C) = 1/Z \exp (- U_C(C)/KT) \tag{4.7}$$

So the U_A can be expressed as below.

$$U_A = - \Sigma \log P(d_{lk}|c_{lk}) + U_c(C) \tag{4.8}$$

The assign category, in the sense of Baysian inference, is determined by maximizing (4.6). The probability is maximized when energy is minimized-this analogous to the situation of thermal equilibrium in statistical physics, where the most probable molecular configurations occur at the lowest energies. For the case of Bayes context classification, the most probable label occurs when the negative exponent power is minimized. Using conventional gradient techniques, maximizing posterior probability is virtually impossible for all but the first order Markov Random Field models, because of the existence of many local extrma. However new multivariate or combinatorial optimization (finding the minimum of a given function depending on many parameters) - stochastic relaxation, which is developed by Kirkpatrick et al [19], offers a practical solution .

After creating the Gibbs models for Bayes' context classificati ', the problem now is to find U(C) or V_c . The V_c functions represent contributions to the total energy from external fields (singleton cliques), pair interactions (doubletons), and so forth.

Typically, several free parameters are involved in the specification of U, and Z is then a function of those parameters.

The most general form of U is that

$$U(\theta) = \sum V_{\{i,j\}}(X_{ij}) + \sum V_{\{(i,j),(i+1,j)\}}(X_{ij},X_{i+1,j})\} + \Sigma V_{\{(i,j),(i,j+1)\}}(X_{ij},X_{i,j+1}) \tag{4.9}$$

where the summation is over all (i,j) ε N, and N denotes the nearest-neighbor. The Ising model (1925), which is the earliest and best-known lattice system , can be thought of the special case of (4.9) in which X is binary (L=2), homogeneous (= strictly stationary) and isotopic (= rotational invariant); Its potential function is

$$V(\omega) = \alpha \Sigma Xij + \beta(\Sigma X_{ij}*X_{i+1,j} + \Sigma X_{ij}*X_{i,j+1}) \tag{4.10}$$

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

For our contextual classification case, the potential U(c) is a function of pattern configurations, the linear combination expression of coordinates as (4.10) is not suitable. But we still assume that the image is homogeneous and isotopic.

Before we derive the canonical potential for general case, we first give definitions of Gibbs ensemble as below[15-17]:

The random field is a Gibbs ensemble if there exists a potential V such that

$$P(\theta) = Z^{-1} \exp (\sum_{L_i \varepsilon W} V_{L_i}(\theta_k))$$

where the summation is over all subsets of the cliques L_i, and Z is chosen to make $\Sigma P(\theta) = 1$, for $\theta \varepsilon \Theta$

G.R. Grimmett[16] gave a useful theorem of potential function:

The random field is a Gibbs ensemble if and only if it is a Markov field. If the random field is a Markov field, then its potential function is given by

$$V(\theta) = \sum_{L_i \varepsilon W} (-1)^{|L_i(\theta)-A|} \log P(\theta_A) \tag{4.11}$$

where the summation is over all subsets of the clique L_i, and $\mu = (-1)^{|A-B|}$ is Mobius function, and |A-B| is the number of elements in A - B.

From above theorem we see that given a Markov random field on a finite graph, the local characteristics (see III-1) do uniquely determine this potential function and then the canonical potential can be determined from these local characteristics.

The problem in this paper can be stated ,that the assigned category is determined by minimizing the following expression:

$$V_A = \Sigma \log P(d_{ij}|C_{ij}) + \Sigma \log P(C_{ij}) + 2*\Sigma \log P\{C_{ij}|C_{i+1,j})\} + 2*\Sigma \log P\{C_{ij}|C_{i+1j})\} \tag{4.12}$$

These formulas can be motivated from the factorization property of Markov Random Field and neighboring clique assumption. We kept the the model as simple as possible, indeed, only cliques of size are involved in the above expression.

Because of the existence of many local extrema, the computation cost of maximizing posterior probability for Bayes classification is usually computational high. For example, a MSS image has N class categories and M X M lattice , the number of configurations is at least N^{M^2}. Hence, the identification of even near-optimal solutions is surprisingly difficult for such a relatively complex function. In the next section we present the implementation of stachastic relaxation context classification, which can remarkably overcome the computational difficulty.

(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 duo to Metropolis et al. [20]. The Metropolis procedure is that samples are randomly generated from a Gibbs distribution at constant temperature, thereby used to simulate the behavior of a physical system in thermal equilibrium. The algorithm is briefly distributed as below. For each state D_{ij} of a model D a random perturbation is made, and the change in energy, ΔU is computed. If $\Delta U<=0$, the perturbation is accepted, if ΔU is positive then the perturbation is accepted with probability

$$P(\Delta U) = \exp(-\Delta U/T) \qquad (5.1)$$

This conditional acceptance is easily implemented by choosing a random number R uniformly distributed between 0 and 1. If $R <= P(\Delta E)$ 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 a posterior probability. This fact was proved by Geman [3]. 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

$$T(k) > = C/\log(1+k) \qquad (5.2)$$

It is employed in executing the K^{th}, site replacement (i.e. the K^{th} classified labels in the iteration scheme). For every k, C is a constant independent of k. Then with probability converging (as $k \to \infty$) the configurations generated by the algorithm will be those of minimal energy.

(VI) summary of the stochastic relaxation context classification procedure:

In summary, the stochstic relaxation context classification procedure can be implemented as follows:

(i) Evaluate training statistics, this includes the mean vector and convariance requred for the Gaussian class conditional distribution.

(ii) Preclassify the image using a pixel independent or context free Bayes classification technique.

(iii) Evaluate the transition probability : $P(C_{ij}|C_{i,j-1},C_{i-1,j})$ from the preclassification results. See M. C. Zhang [18].

(iv) Using equation (4.12),(5.1),(5.2) perform the stochastic relaxation context classification.

(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 [3] proved the convergence properties of this algorithm, and showed how to markedly reduce the computational difficulty. As we mentioned before, for an MSS image which has N class categories and a M X M neighborhood, the number of configuration is at least N^{M^2}. In his 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 of context classification results 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 give 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 Landasat 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 are correspond uniform objects (categories) on the earth's surface. We believe that some smooth or homogenous pattern configurations are much more probable than others, and some irregular patterns have very low probabilities.

This fact gives us a stage the iterative procedure in that we may use these most probable homogenous patterns at the beginning of the iteration procedure. After that we randomly generate the pattern configuration, and skip irregular patterns which 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.

First we assume that uniform pattern configurations have higher occurring probabilities, and 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 the above testing we assume that some simple pattern configurations (Figure 1) also have higher occurring probabilities. Theses are assigned and tested again.

The pattern in Figure 1 a has upper and lower parts, and the assignment of these labels in each part is also based labels in each part is also based on the labels in that part assigned at the preclassification stage. Similarly, Figures 1-b,c and d show three other simple patterns.

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

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

$$IR = \frac{\text{number of pixels in the neighborhood}}{\text{number of classes in the neighborhood}}$$

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, this pattern is too irregular, and the procedure will skip testing and generate the next pattern instead.

Figure [6] shows the experiment results of proved schemes under 30 iterations. For same improvement of context classification results the proved scheme reduces number of iterations from about 600 to 30.

(VIII) Experimental results:

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

(VIII-1) Simulated data experiments:

Because classification accuracy can vary with different kinds of original input data set, use of some simple simulated data set to evaluate the effectiveness of the classification technique is very desirable. In this subsection we illustrate a simulated data experiment method, which is generated from the ground truth of a real remote sensing image.

The simulated data generating method proposed by P.H. Swain [6,8] as follows:

Using the ground truth (or classification map) and associated estimated mean vectors and contrivance matrices of the classes (developed in performing the no-context classification), new data vectors are produced by a Gaussian random number generator and composed into a new data set. Thus the new data had 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 are class-conditionally independent of adjacent pixels.

(4)There are no mixture pixels.

Data simulated in this manner consistent somewhat of 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.

(VIII-2) experiment results:

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 13 April 1976 MSS scene of Roanoke, VA, was selected as the first study area.It was classified by a conventional context free method in order to compare the results. The following ground cover classes are 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;
(6) Class 6: Wetland;
(7) Class 7: Barren land;
(8) Class 8: Tundra;

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 presented 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 conventional classification. The accuracy of context free classification with real remote sensing image, including Bayesian classifiers, 60%; AMOEBA (Bryznt, 1979), and ISODATA (Duda and Hart, 1973), are not unusual for scenes of this complexity.

The test image has four bands of digital, multispectral data. The mean vectors m_i and the covariance matrices Σ_i for each class i is 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 the expression (4.12)

The probability of each class $P(\omega_k)$ is directed calculated from preclassification result. The transition probability $P(C_{ij}|C_{ij-1},C_{i-1j})$ can get from maximum likelihood estimation or robust estimation (See Zhang 18). The $P(d_{ij}|C_{ij})$ is the class conditional distribution, which is estimated from the training sets and ground truth. In this experiment means and covarience matrices of each category were calculated from ground truth data. The class conditional probability $P(d_{ij}|C_{ij})$ are assumed multidimensional normal:

$$P(d_{ij} \mid C_{ij}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} EXP\left(-\frac{1}{2}(d_{ij}-m_k)^T \Sigma^{-1} (d_{ij}-m_k)\right) \quad (9.1)$$

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 Figure 2-14. The contingency tables for classification results are in table 1-2. The contingency table and classification images lead us to a conclusion that the stochastic relaxation context classification result, in most cases, provide results as good as the recursive context classification algorithm in [2]. 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 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 were eliminated, and each area was much more homogeneous in the contextual classification results. Boundaries remained accurately placed. The MSS four band image, ground truth map, which had been classified by professional analysts, are given in Figure 2 and Figure 3, respectively. The above comparison and Fig 6 indicated that a 5 to 10% improvement of accuracy was obtained by the context classification method. So in addition to the visual improvement, the context classification scheme improves the classification accuracy as well.

The second study area is California. Three MSS classification results where sizes of subsets are 130 X 90, 101 X 70, 130 X 60 respectively, are shown in Figure 6 to 12. These results show that the algorithm was effective in several different areas with varied categories and preclassification accuracies (these areas had about 90% preclassification accuracy).

The third study area is crop field at Clarke, Orgon. The Landsat MSS image is 12 band data set(Landsat MSS bands 4-7 from three dates). Thomas <1982> showed the accuracy of maximum likelihood classification and his canonical analysis method in the same study area is about 75%. Our context classification scheme raise the classification accuracy to 80.8%.

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

It can be see that the context free classification (Bayes classifer) is very sensitive to random noise, and context classification by stochastic relaxation is superior to any other classifier.

(IX) Algorithm in parallel:

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

Performance of the each neighbor operator is independent with other neighbors in the entire image. The amount of time required for each iterative of entire image is only direct proportion with number of pixels in the image.

This important property allows the algorithm to run in full parallel version, which will require extremely sophisticated new hardware.

A more modest degree of parallelism was noted by Geman [3]. Since the convergence theorems are independent of the details of the site replacement scheme n1,n2,... the graph associated with the MRF can be divided into collections of sites with each collection assigned to an independently running(asynchronous) processor. Each such processor would execute a raster scan updatary of its assigned sites. Communication requirements will be small if the division of the graph respects natural topology of the scene, provided of course that the neighborhood systems are reasonably local. Such an implementation with five or ten micro- or mini-computer, represents a straightforward application of available technology.

(X) Summary

We have developed a new multispectral image context classification with Markov Random Field, where remotely sensed data are more efficiently and 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 Markov-Gibbs Random Field. The implementation of the relaxation algorithm is on a form of optimization using annealing. In this paper we have first motivated a Bayesian context decision rule, and introduced a Markov-Gibbs model for the original LANDSAT MSS image, and then develop a new contextual classification algorithm, in which maximizing the a 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 iteration by using some special pattern configurations at the beginning of the iterative procedure. Finally, we have present experimental results which are based on simulated and real multispectral remote sensing image to show how classification accuracy is greatly improved. The algorithm is highly parallel and exploits the equivalence between Gibbs distributions and Markov Random Fields (MRF).

References

[1] R.M. Haralick, 1983. Decision Making in Context, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol, PAMI-5, No. 4, July.

[2] R. M. Haralick, M. C. Zhang and J. B. Campbell, 1984. Multispectral Image Context Classification Using the Markov Random Field.

[3] S. Geman and D. Geman, 1983. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.

[4] T.S. Yu and K.S. Fu, 1982. Recursive Contextual Classification using a spatial stochastic model, Pattern Recognition Vol., 16, No. 1, pp. 89-106.

[5] J.C. Tilton, S.B. Vardeman, and P.H. Swain, 1982. Estimation of Context for Statistical Classification of Multispectral Image Data. IEEE transaction on Geoscience and Remote sensing, Vol., GE-20, No. 4, October, 1982.

[6] P.H. Swain, S.B. Vardeman and J.C. Tilton, 1981. Contextual Classification of Multispectral Image Data, Pattern Recognition Vol., 13, No. 6, pp.

428-441.
[7] S.W. Wharton, 1982. A Contextual Classification Method For Recognizing Land use Patterns in High Resolution Remotely Sensed Data, Pattern Recognition Vol., 15, No. 4, pp. 317-324.
[8] P.H. Swain, H.J. Siegel and B.W. Smith, 1979. A method for Classifying Multispectral Remote Sensing Data using Context, 1979 Machine Processing of Remote Sensed Data Symposium. p. 343-353.
[9] J.E. Besay, 1974 Spatial Interaction and the Statistical Analysis of Lattice Systems, J/R. Statist. Soc. B36, 192-236.
[10] B. Y. Rubinstein, 1981. Simulation and the Monte Carlo Method.
[11] R. Chellappa. Two-Dimensional Discrete Gaussing Markov Random Field Models for Image Processing, Department of EE-systems and Image Processing Institute University of Southern California, Los Angeles, California.
[12] M. Hassner and J. Sklansky, 1980. The use of Markov Random Fields as Models of Texture, Computer Graphics and Image Processing Vol. 12, 357-370 (1980).
[13] K.S. Fu and T.T. Yu, 1980. Statistical Pattern Classification Using Contextual Information.
[14] C.M. Gurney, and Townshend 1983. The use of Contextual Information in the Classification of Remote Sensing Data, Photogrammetric Engineering and Remote Sensing, Vol., 49, No. 1, January 1983, pp. 55-64.
[15] P. Suomela, 1976 , Construction of Nearest Neighbour Systems.
[16] G.R. Grimmertt, 1973, A Theorem about random fields, Bull. London Math. Soc.,5,81-84
[17] C.J. Preston, 1974, Gibbs States on Countable sets, Cambridge Univ. Press.
[18] M.C. Zhang, 1984, Advanced Spatial Information Processing : Model and Application, Desertation proposal.
[19] S. Kirkpatrick, C.D. Gellatt, Jr., and M.P. Vecchi, "Optimization by simulated annealing", IBM Thomas J. Watson Research Center, Yorktown Heights, NY,1982.
[20] N. Metroplis, A.W. Rosenbluth, M.N. Rosenluth, A.H. Teller and E. Teller, "Eguations of state calculations by fast computing machines", J. Chem. Phys. , Vol. 21, pp, 1087-1091,1953 .
[21] F. Spitzer, 1971, Random Fields and Interacting Particle Systems , M.A.A. Summaer Seminar Nots.
[22] B.L.Martini, and R.M. May, 1970, Non-linear models for voting behavior, Unpublished manuscript.
[24] R.Chellappa and R.L. KAshyap, "Digital Image Restoration Using Spatial Interaction Models", IEEE trans. Account,Spech,Signal processing ,Vol. ASSP-30, PP.461-472.
[25] R.P. Kinderman, 1973, Random Fields : Theorems and Examples, J. of undergraduate Math. ,5,25-34.

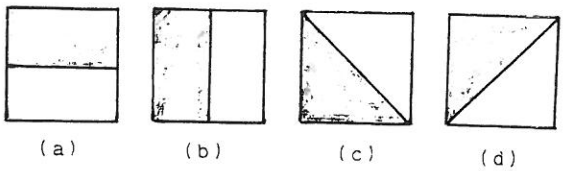

Figure 1. Foure simple pattern configurations for im proved scheme.

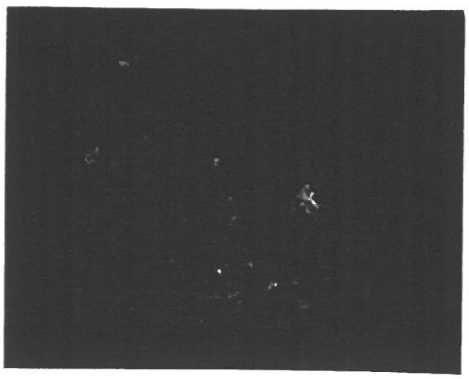

Figure 2. First band of MSS scence of Roanoke, VA,: Aprial 1976, image size 151 X 151.

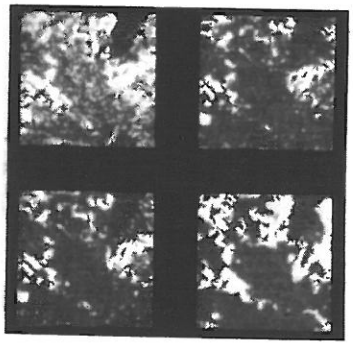

Figure 3-6: Classification results of MSS Roanoke.
Left-upp(fig 3): Bayes Preclassification result.
Right-upp(fig 4): Markov contextual classification.
Left-Bottom(fig 5): Stochastic relaxation contextual classifcation result.
Right-bottom(fig 6): Ground Truth image.

Figure 7-9: Classification result of MSS Scence of California. (H)
Left(fig 7): Bayes preclassification results.
Middle(fig 8): Markov contextual classification.
Right(fig 9): Stochastic relaxation context classification results.

Figure 10-12: Classification results of MSS Scence of California (I)
Left(fig 10): Bayes preclassification result.
Middle(fig 11): Markov contextual classification.
Right(fig 12): Stochastic relaxation context classification.

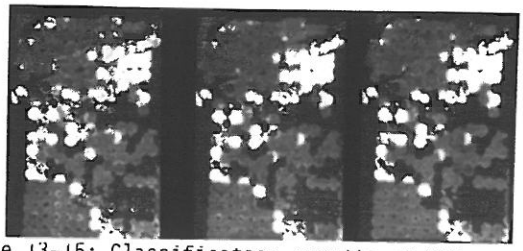

Figure 13-15: Classification results of MSS Scence of Cropfield at Clarke, Orgon (1982). Left(fig 13): Bayes preclassifcation result. middle(fig 14): Markov contextual classification. Right(fig 15): Stochastic relaxation context classification result.

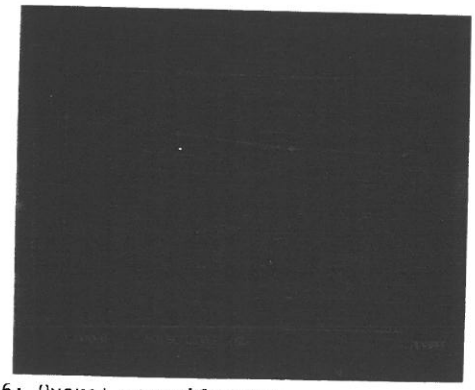

Figure 16: Overal classifcation accuracy curves VS noise level: Yellow line - context classification by stochastic relaxation, Green line - Dynamic Programming approch to context classification Red line - Bayes classification.

TABLE 1. Contingency tables for classification results of 13 April 1976 MSS scene of Roanke, VA. Scale factor of the number of pixels 10 ** 1 .

COL = assigned categories ROW = true categories

(A) Bayes classification result

CLASS	URB	AGR	RNG	FST	TOTAL	ACC(%)*
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%**

(B) Context classification result using a dynamic programming approach

CLASS	URB	AGR	RNG	FST	TOTAL	ACC(%)*
URB	999	418	0	20	1437	69.5%
AGR	183	371	0	24	578	64.2%
RNG	0	0	0	0	0	-
FST	58	28	0	167	253	66.0%
TOTAL	1240	817	0	211	2268	67.8%**

(C) Stochastic relaxation Context classification result

CLASS	URB	AGR	RNG	FST	TOTAL	ACC(%)*
URB	1108	317	0	12	1437	77.1%
AGR	179	385	0	14	578	66.6%
RNG	0	0	0	0	0	-
FST	50	35	0	168	253	64.4%
TOTAL	1337	737	0	194	2268	73.3%

TABLE 1. Contingency tables for classification results of 13 April 1976 MSS scene of Roanke, VA. Scale factor of the number of pixels 10 ** 1. (continue)

* Classification accuracy.
** Overall classification accuracy: ration of the number correctly classified pixels to the number of total classified pixels.

URB -- Urban or built-up Land
AGR -- Agricultural Land
RAN -- Rangeland
FSN -- Forest Land

Table 2. Contingency tables for classification results of test image 'Clark'. Scale factor of the number of pixels 10 ** 1 .

COL = assigned categories ROW = true categories

(A) Bayes classification result

CLASS	WHT	ALF	POT	CRN	BNS	APL	PAS	RNG	TOTAL	ACC(%)*
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	19	0	2	32	652	84.6%
CRN	1	5	1	65	2	0	0	4	78	83.3%
BNS	2	4	2	1	35	0	0	3	48	72.9%
APL	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%

(B) Context classification result using dynamic programming approach

CLASS	WHT	ALF	POT	CRN	BNS	APL	PAS	RNG	TOTAL	ACC(%)*
WHT	1073	26	26	1	11	0	10	601	1248	90.9%
ALF	89	390	150	3	1	0	1	34	668	58.4%
POT	58	29	534	2	6	0	0	23	652	81.9%
CRN	1	5	1	68	0	0	0	4	79	86.1%
BNS	1	6	2	1	36	0	0	3	49	73.5%
APL	0	2	0	0	0	0	0	0	2	0%
PAS	0	1	0	0	0	0	8	3	11	72.7%
RNG	19	16	15	1	3	0	1	339	394	86.1%
TOTAL	1681	605	777	82	58	0	23	1064	3103	80.5%

(C) Stochastic relaxation Context classification result

CLASS	WHT	ALF	POT	CRN	BNS	APL	PAS	RNG	TOTAL	ACC(%)*
WHT	1080	23	25	1	1	0	0	58	1188	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	72.9%
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%

* Classification accuracy.
** Overall classification accuracy: ration 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